International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 349-354, 2017, ISSN:-2319–8656

Effective Co-Extraction of Online Opinion Review Target-Word Pairs and Product Aspect Ranking

Chandana Oak
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Prof. Manisha Patil
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: The popularity and use of e-commerce are increasing day by day. Recent trends have shown that many people are now buying their products online through different e-commerce sites such as Flipkart, Amazon, Snapdeal, etc. Customers review and rate the products they have bought over multiple independent review sites such as gsmarena, social networking sites like Facebook, blogs, etc. Customers can also comment on the quality of service they have received after buying their product. These online reviews are of immense help to potential buyers for that product in decision making and also to manufacturers/sellers to get an immediate feedback about the product quality, product performance, after sales service, etc. As the number of reviews for a product is usually large, it is next to impossible to go through all the reviews and form an unbiased opinion about the product. Also, there are multiple sources of these online reviews. Hence, online review mining is gaining importance.

A product can have many product features, wherein some features are more important than others. Review usefulness can also be increased by ranking the product aspects as per their importance and popularity. Ranking product aspects manually is very difficult since a product may have hundreds of features. So, an automated method to do this is needed.

This paper presents a methodology for co-extracting opinion targets and corresponding opinion words from online opinion reviews as well as for product aspect ranking.

Keywords: Online review mining, Opinion mining, Opinion Co-extraction, Opinion Target, Opinion Word, Product Aspect Ranking, Sentiment Analysis

1. INTRODUCTION

The importance and popularity of e-commerce is increasing day by day as a lot many people have started buying their products online. The convenience of online shopping, a wide variety of different product ranges, huge discounts given by the retailers, etc. are some of the reasons which have contributed to enhancement in e-commerce popularity. While the trend of online shopping is increasing, logging of opinion about the product bought, has also been gaining importance. Customers who have bought a particular product, say mobile phone, log their opinion about the phone either on the ecommerce sites such as Flipkart, eBay, etc., independent review sites such as gsmarena, blogs or social networking sites such as Facebook or Twitter, etc. These online reviews are of immense help to potential buyers for that product in decision making whether to buy the product or not, and also to manufacturers/sellers to get an immediate feedback about the product quality, product performance, after sales service, etc. For any product, the number of online reviews is usually pretty large. Also, most of the reviews are verbose. Any potential customer or a manufacturer can find it very difficult to go through all those reviews and form an unbiased opinion about the product. Hence, opinion mining or sentiment analysis is proving to be extremely useful.

Additionally, a product has numerous aspects associated with it. For example, iPhone 3GS has more than three hundred aspects, such as “usability," “design," “application," "3G network". Some aspects are more important than the others, and have greater impact on the potential consumers’ decision making as well as on product development strategies. For example, some aspects of iPhone 3GS, like, “usability" and “battery," are more important than the others such as “usb" and “button”. Online reviews might focus on some specific products rather than on the overall product itself. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition. Potential customers can make wise purchasing decisions by focusing on some important aspects rather than on some less important ones. Similarly, manufacturers/sellers can also focus on improving the quality of these more important products by enhancing the overall product reputation. But, considering the large number of features a product has, it is impractical to manually identify these important aspects from a large number of online reviews. Hence, a method to identify these important products is much necessary.

1.1 Opinion Mining

Opinion mining or sentiment analysis is related with mining and analysis of natural language for tracking the mood or feedback of people about a particular product. It can be treated in short as a system to collect and classify different opinions about a particular product or service. [1] It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be overall positive about a digital camera, but can be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer.

[2] gives the generic architecture for opinion mining.

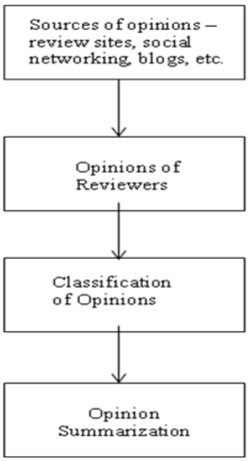

Sources of opinions – review sites, social networking, blogs, etc.
Opinions of Reviewers
Classification of Opinions
Opinion Summarization

Fig 1: Generic architecture for Opinion Mining

People log their opinions on different products on online ecommerce sites, social networking sites, blogs, etc. Thus, there can be multiple sources which a potential buyer can visit for obtaining first hand information of the product which he/she is interested in buying. Information retrieval techniques like web crawling, etc. can be used to extract opinions from all these sources and can be stored into a database. These extracted opinions are analyzed and classified as per their polarity i.e. if the opinion is positive or negative.

1.2 Product Aspect Ranking

A potential customer is mainly interested in the most important and popular features or aspects of a product. E.g. while buying a smart phone, he/she will look for screen resolution rather than usb. Thus, some product aspects hold more importance than others and they can also influence the overall rating for the product. Ranking product features would help increase the usefulness and effectiveness of online reviews. But, it is next to impossible to identify and rank these product features as per their importance manually. Hence, an automated method to do so is much needed.

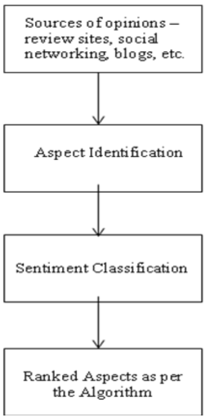

Sources of opinions – review sites, social networking, blogs, etc.
Aspect Identification
Sentiment Classification
Ranked Aspects as per the Algorithm

Fig 2: Generic architecture for Product Aspect Ranking

Fig 2 gives the generic architecture for product aspect ranking. Opinions can be extracted from different online ecommerce sites, social networking sites, blogs, etc. The next step is for aspect identification in which all the different product features are identified. The, these aspects are classified as either positive (i.e. good) or negative (i.e. bad) based on the comments that these features receive. This process is called as sentiment classification. Then these aspects are ranked based on some ranking algorithm.

2. MOTIVATION

Many online reviews are verbose, which reduces the readability and thereby the interest of any person reading the reviews. Also, there are many reviews for a product, thus making it difficult for a reader to go through all the reviews and derive meaningful information from them. Then the potential customer might go through just a few reviews making the opinion biased. Additionally, users can express their opinions either on e-commerce sites such as Snapdeal, Amazon or Flipkart, etc., or social networking sites, or blogs, etc. It makes it very difficult for a person to browse multiple sites to get all the reviews. Thus, opinion mining has become the need of the hour.

Product aspect ranking organizes the extracted features (such as processor speed, battery life, connectivity, screen resolution, etc.) of a product as per their importance to improve usefulness of reviews. But, identifying these features manually and then ranking them as per their importance is very difficult due to the large number of reviews. Thus, an automated method to extract and rank the important aspects of a product is much needed.

3. LITERATURE REVIEW

The importance and popularity of online review mining are increasing day by day. [2] gives an overview of different methods used in opinion mining and product aspect ranking. [3] proposes association mining technique to identify the most frequently occurring nouns and noun phrases in a review sentence. The ones which have a high occurrence frequency are then extracted as opinion targets. [4] proposes a Word based Translation Model (WTM) which is a graph based algorithm for extracting opinion targets. WTM is more effective than traditional opinion target extraction methods viz. adjacent methods and syntax based methods. This WTM method is independent of parsing performance, which plays a major role in syntax based methods and also of window size which is used in adjacent methods to find opinion relations with the corresponding opinion words. [5] proposes a method called Double Propagation which extracts opinion words or opinion targets iteratively from the words and targets already extracted during the previous iteration using syntactic relations. [6] highlights nearest neighbor rule which considers nearest adjective/verb to a noun as its opinion word. Thus, the span on which it works is limited and accuracy of results reduces.

[7] and [8] are based on sentence level extraction. In [7], Conditional Random Fields (CRFs) are used to jointly extract product aspects and positive/negative opinions from online review sentences. Phrase dependency parsing is used in [8] as many product features such as processor speed, screen resolution, wifi connectivity, etc. are noun phrases rather than just nouns.

An OPINE algorithm is used in [9] which proposes syntactic parsing of reviews to find opinion relations among words. But it is very much dependent on parsing performance which is affected in case of informal writing style of online reviews.

[10] uses a partially supervised word alignment model (PSWAM) which co-extracts opinion targets and opinion words based on a graph-based co-ranking algorithm. This algorithm extracts opinion words and targets more accurately than nearest neighbor rules or syntax based methods.

[11], [12], [13], [14] and [15] focus on product aspect ranking and its different applications. [11] proposes double

propagation for product feature extraction along with an algorithm for ranking of products based on feature importance. [12] and [13] use an aspect ranking algorithm which considers both the aspect frequency as well as their influence on overall opinion of the product.

Table 1: Summary of Related Work

Reference	Technique used for feature extraction	Is product feature ranking used?
Mining and summarizing customer reviews [3]	Association mining	No
Opinion target extraction using word based translation model [4]	Word Translation Model	No
Opinion word expansion and target extraction through double propagation [5]	Double Propagation	No
Mining opinion features in customer reviews [6]	Nearest neighbor rules	No
Structure-aware review mining and summarization [7]	CRF	No
Phrase dependency parsing for opinion mining [8]	Phrase dependency parsing	No
Extracting product features and opinions from reviews [9]	OPINE	No
Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model [10]	PSWAM	No
Extracting and ranking product features in opinion documents [11]	Double propagation with part-whole relationship	Yes
Aspect ranking: Identifying important product aspects from online consumer reviews [12]	Stanford parser and SVM classifier	Yes
Product Aspect Ranking and Its Applications [13]	Stanford parser and SVM classifier	Yes

The proposed system uses the following - The product reviews are partially parsed using Stanford parser. The “Opinion TW Co-extraction Algorithm” is proposed to co-extract opinion targets and corresponding opinion words from classified and semi-supervised product reviews. In addition to this, product aspect ranking is also included as part of proposed work.

4. SYSTEM OVERVIEW

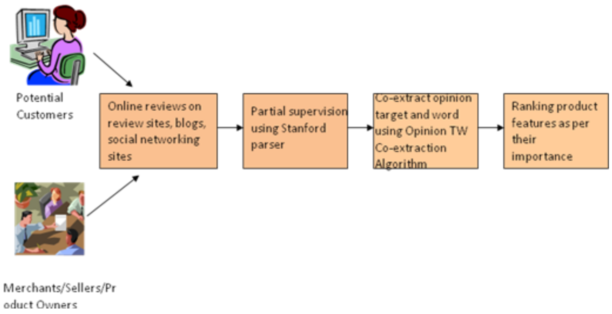

Potential Customers
Online reviews on review sites, blogs, social networking sites
Partial supervision using Stanford parser
Co-extract opinion target and word using Opinion TW Co-extraction Algorithm
Ranking product features as per their importance
Merchants/Sellers/Product Owners

Fig 3: Architecture for Proposed System

4.1 Modules

Fig. 3 gives the overall architecture for the proposed system.

Millions of reviews of multiple products are available online on various sites such as different e-commerce sites like eBay, Flipkart, Amazon, etc., or also on independent review websites like gsmarena, rottentomatoes, etc. Users of products can also express their opinion through social networking sites like twitter, Facebook, etc. as well as over blogs. These online reviews are referred to by potential consumers in order to get review of the product. Also, merchants/sellers refer to these reviews in order to get a first hand feedback about their product/service. All these sources will act as the source of online reviews for co-extraction algorithm.

4.1.1 Data Processing

The Data Processing module is partial supervision using Stanford parser. Here, the input review data is fed to Stanford parser module to extract opinion targets (i.e. nouns or noun phrases) and opinion words (i.e. adjectives). Here, partial supervision technique is being used which is more advantageous over completely unsupervised technique. A completely unsupervised technique shows poor results since it does not have any training data to start with. Partial supervision is also not heavily dependent on parsing performance like a completely supervised parsing technique does. Here, partial parsing of review data would be done using Stanford parses. Of the product review data, 50% of the review data would be considered for training the parser and remaining 50% data would be test data.

4.1.2 TW Co-extraction

The second module is for Co-extraction of opinion target and opinion word. Co-extraction of opinion target and opinion words would be done using “Opinion TW Co-extraction Algorithm”. This algorithm takes the partially supervised set of classified product reviews as input.

Algorithm 1: Opinion TW Co-extraction Algorithm

Input: Partially supervised set of classified product reviews I = {p1, p2, …, pn}
Output: The probability of co-alignment for sentences based on optimal association score

```
Initialization:
Initialize R as review data of products
Initialize review = i[data]
Initialize o_word [ ] = review [opinion word]
Initialize o_target [ ] = review [opinion target]
for each review[]
    for each o_target[]
        Nt := Count (opinion targets)
    end for
    for each o_word[]
        Nw := Count (opinion words)
    end for
    Ntw := Count (collocated opinion target and opinion word)
    P (o_target | o_word) = Ntw / Nw

    P (o_word | o_target) = Ntw / Nt
    Optimal Association Score OAS = [P (o_target | o_word) + P (o_word | o_target)] / 2
end for
Co-extract the opinion target and opinion word with higher OAS as being aligned with each other
```

The input to Opinion TW Co-extraction Algorithm is the set of product reviews which is partially parsed. Consider n to be the total number of product reviews. R denotes the review collection per product. For each category of product review, assign “review” collection with the partially parsed review data for that product. Fetch the opinion words from review into o_word based on training data. Similarly, fetch the opinion targets from review into o_target based on training data.

From the entire corpus, find Nt to be the total number of opinion targets and Nw to be the total number of opinion words. Also find the count of collocated opinion target and opinion word as Ntw. Find the estimated alignment probability for a potential opinion target and word pair as P (o_target | o_word) and P (o_word | o_target). To find the optimal association score (OAS), we take a mean of the 2 probability values fetched earlier. Co-extract the opinion target and opinion word with higher OAS as being aligned with each other.

To find one-many relation between opinion targets and opinion words (E.g. – The ambience and food in this restaurant are very good – here, good indicates both ambience and food), we find 2 opinion targets separated by a conjunction and aligned to one opinion word. Similarly, find 2 opinion words separated by a conjunction and aligned to one opinion target. Co-extract these as opinion word and target having one-many relation.

4.1.3 *Product Aspect Ranking*

The third module is Product Aspect Ranking. A product has many features. E.g. – if we consider mobile phone as a product, it has features like processor speed, screen resolution, battery life, wifi connectivity, etc. Similarly, if we consider a camera, it has aspects like shutter speed, lenses, picture quality, etc. These product features can be treated as opinion targets. Of all the features that a product has, some can be more important than others. A potential customer gives more importance to these features rather than giving much focus on the less important ones while buying any product. Also, manufacturers can focus mainly on the more important features while deciding on their product development strategies. Ranking these features as per some parameters, like frequency at which the product features were commented on, etc. would help increase the usefulness of online reviews. But, since the number of product reviews and also the number of features a product has is large, it is next to impossible to rank these features manually. The proposed work build a ranking framework on top of co-extraction framework which will organize all the product features as per their popularity i.e. as per the number of reviews a product feature receives, and also as per the influence a feature has on the overall opinion of the product. This will improve the usability of the review summarization. We will use the same set of opinion targets extracted by the Opinion TW Co-extraction Algorithm as product aspects or features and Naïve Bayes classifier for identifying the polarity (i.e. whether the reviews are positive or negative) of the feature expressed in the review based on opinion word extracted for the feature.

Algorithm 2: Product Aspect Ranking Algorithm

```
Input: Product name and opinion target-word pairs
Output: Ranked product aspects
Initialization:
Initialize set of product reviews as R
Initialize Negative sentimental word set N[]
Initialize Positive sentimental word set P[]
Initialize set of product aspects as aspect[]
Initialize set of overall ranking of aspect Or[]
for each aspect[]
    Initialize array for sentiments as S[]
    Int i = 0
    For each i
        If S[i] is negative then
            N[i] = S[i]
        Else if S[i] is positive then
            P[i] = S[i]
        End if
    end for
end for
Overall rating of product as o[]
Calculate Importance weight for aspect wt[] as frequency of the aspect commented on
Or[] = o[] * wt[]
Rank the product aspects as per Or[] value
```

4.2 Mathematical Model

1. Co-extraction of opinion target and opinion word –

$$P\ (\text{o-target} \mid \text{o-word}) = Ntw / Nt \tag{1}$$

$$P\ (\text{o-word} \mid \text{o-target}) = Ntw / Nw \tag{2}$$

$$OAS = [P\ (\text{o-target} \mid \text{o-word}) + P\ (\text{o-word} \mid \text{o-target})] / 2 \tag{3}$$

Where,

Nt := Count (opinion targets)

Nw := Count (opinion words)

Ntw := Count (collocated opinion targets and words)

2. Product Aspect Ranking –

Consider the opinion of various aspects as o[]

Consider the importance weight for aspect as wt[]

Then, product rank can be found as –

$$Or[] = o[] * wt[] \tag{4}$$

Rank the product aspects as per Or[] value.

4.3 User Interface of the Application

The graphical user interface of the application is simple and self-explanatory. The user has to provide the user id and password to the application as the system allows only authorized users to enter. After authentication, the user is asked to select the product which he/she wishes to get opinions on. The display window will have all the opinion target-word pairs displayed which are extracted from a set of online reviews along with their OAS, using which the user can form an opinion about the product. Then, on the Product Aspect Ranking screen, the user would see a list of product features ranked as per their importance along with their polarities. This summarization would assist the user in decision making.

5. RESULTS AND DISCUSSION

Place Co-extraction of opinion target and corresponding opinion word as well as product aspect ranking enhance the usability of online reviews and assist users in quick and unbiased decision making about the product.

The first dataset selected is extracted from Amazon. This dataset has more than 3800 review sentences (i.e. review ids) for phone and its accessories. The second dataset selected is from CRD database which depicts phone records. This dataset has 546 review sentences. In the application, the dataset is first loaded and all the review sentences are selected. The next step extracts all the opinion targets and opinion words list. Then we execute the co-extraction algorithm i.e. Opinion TW Co-extraction algorithm to give the count of each opinion target, count of each opinion word, count of co-located opinion target and word and also calculates the Opinion Association Score (OAS) for each pair.

The results for proposed Opinion TW Co-extraction algorithm are compared against the base system. The graphs shown below depict the results for the base and proposed systems.

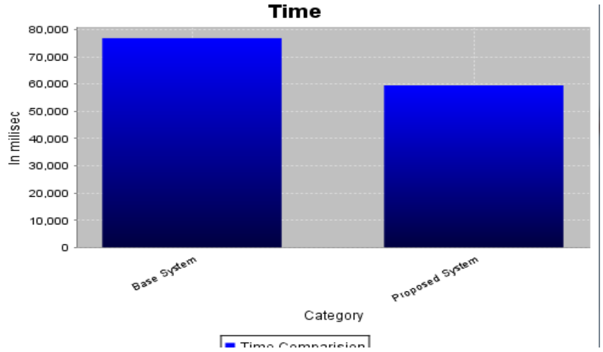


Fig 4: Execution time for Amazon dataset

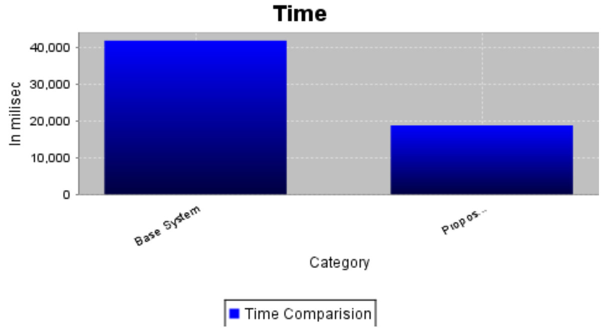


Fig 5: Execution time for CRD dataset

Fig 4 and Fig 5 show the execution time needed to execute the Amazon dataset and CRD dataset respectively. The execution times for base algorithm and proposed Opinion TW Co-extraction Algorithm are compared and the time needed to execute proposed algorithm is much lesser as compared to base algorithm. Thus, the proposed algorithm shows a better time complexity as compared to base algorithm.

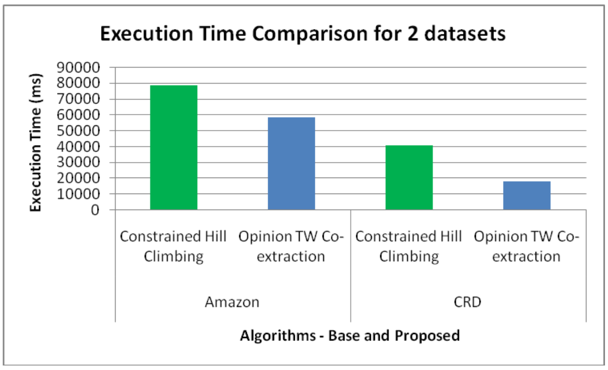


Fig 6: Execution time comparison for two datasets

As can be seen from fig. 6, since the number of records (or number of review sentences) for Amazon dataset is larger as compared to CRD dataset, the execution time for Amazon dataset is also higher than CRD dataset.

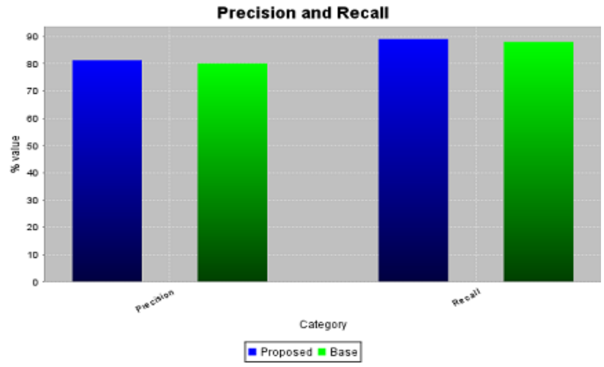


Fig 7: Precision and Recall for Amazon dataset

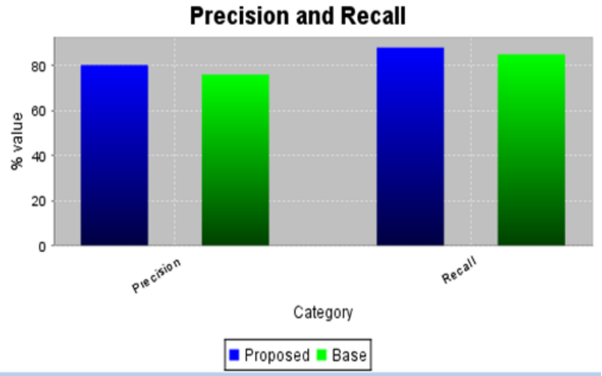


Fig 8: Precision and Recall for CRD dataset

Fig. 7 and fig 8 show the comparison between precision and recall parameters for Amazon and CRD datasets respectively. The values for precision and recall for base system and proposed system are comparable, with those for proposed system being slightly higher.

6. CONCLUSION AND FUTURE WORK

This paper presents an overview of the proposed system along with the mathematical model and results, which will be used for opinion target and corresponding opinion word coextraction from online reviews. As can be seen from the results, the time taken to execute the proposed algorithm is lesser as compared to that for base algorithm. Thus, proposed algorithm shows better time complexity as compared to base algorithm. Also, precision and recall for base system and proposed system are comparable with each other. It also gives an overview of the Product Aspect Ranking feature that would be built on top of the co-extraction framework.

As future work, this system can be integrated with recommendation system and product recommendations can be made based on the product category selected and the product aspect rankings seen.

7. ACKNOWLEDGEMENT

This work would not have been possible without the support and help of many individuals. I would like to extend my sincere thanks to all of them. I am indebted to Prof. Manisha Patil, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant supervision as well as all the other staff members for providing important information regarding the work.

8. REFERENCES

[1] http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining

[2] C. Oak, M. Patil, “Opinion Mining and Product Features Ranking: A Survey”, International Journal for Modern Trends in Science and Technology, Volume: 02, Issue No: 10, October 2016

[3] A M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.

[4] K. Liu, L. Xu, and J. Zhao, “Opinion target extraction using word based translation model,” in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.

[5] G. Qiu, L. Bing, J. Bu, and C. Chen, “Opinion word expansion and target extraction through double propagation,” Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011

[6] M. Hu and B. Liu, “Mining opinion features in customer reviews,” in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755–760.

[7] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, “Structure-aware review mining and summarization.” in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 653–661.

[8] Y. Wu, Q. Zhang, X. Huang, and L. Wu, “Phrase dependency parsing for opinion mining,” in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.

[9] A.-M. Popescu and O. Etzioni, “Extracting product features and opinions from reviews,” in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 339–346.

[10] Kang Liu, Liheng Xu, and Jun Zhao, “Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model”, IEEE transactions on Knowledge and Data Engineering, Vol. 27, No.3, March 2015

[11] L. Zhang, B. Liu, S. H. Lim, and E. O’Brien-Strain, “Extracting and ranking product features in opinion documents,” in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.

[12] J. Yu, Z.-J. Zha, M. Wang, and T. S. Chua, “Aspect ranking: Identifying important product aspects from online consumer reviews,” in Proc. ACL, Portland, OR, USA, 2011, pp. 1496–1505.

[13] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua, “Product Aspect Ranking and Its Applications”, IEEE transactions on Knowledge and Data Engineering, Vol. 26, No.5, May 2014.

[14] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. Artificial Intelligence, 165(1):91–134.

[15] K. Zhang, R. Narayanan, and A. Choudhary. Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN, 2010.

[16] K. Zhang, R. Narayanan, and A. Choudhary, Mining Online Customer Reviews for Ranking Products, Technical Report, EECS department, Northwestern University, 2009.

[17] http://ijiset.com/vol2/v2s6/IJISET_V2_I6_106.pdf

[18] Rajesh Sharadanand Prasad, U. V. Kulkarni, “Implementation and Evaluation of Evolutionary Connectionist Approaches to Automated Text Summarization “, Journal of Computer Science 6 (11), @ 2010 Science Publication, ISSN: 1549-3636, Page - 366-76, Sep-2010.

[19] Rajesh Sharadanand Prasad, Dr. U. V. Kulkarni, “An Automated Approach to Text Summarization Using Fuzzy Logic”, International Journal on Computer Engineering &amp; Information Technology, IJCEIT, Volume 23, Issue – 01 ISSN: 0974-2034, Page - 07-21, May -2010.

[20] Rajesh Prasad, Dr. U.V. Kulkarni, “Two Approaches to Automatic Text Summarization: Extractive Methods and Evaluation”, International Journal of Computer Engineering and Computer Applications, IJCECA, ISSN: 0974-4983, Ma -2010

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 355-361, 2017, ISSN:-2319–8656

Enhanced Cross Domain Recommender System using Contextual parameters in Temporal Domain

Swapna Joshi
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Prof. Manisha Patil
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: Cross-domain collaborative filtering (CDCF) is an evolving research topic in the modern recommender systems. Its main objective is to alleviate the data sparsity problem in individual domains by mixing and transferring the knowledge among the related domains. But there is also an issue of user interest drift over time because user’s taste keeps on changing over time. We should consider various temporal domains to overcome user interest drift over time problem to predict more accurately as per the user’s current interest. This paper discusses how to achieve effective ratings recommendations using contextual parameters in temporal domain in this research line. It calculates the contextual parameters as per user’s current timestamp. This will enhance the recommendations more in line with the current temporal domain. It also deals with cross domain recommendations for both movies and novels based on their categories and similarities.

Keywords: Collaborative filtering; temporal domain; cross-domain; recommender systems

1. INTRODUCTION

The main problem of existing Collaborative Filtering (CF) methods is to find similar users or items and to measure similarities between them. Unit now, most of the existing CF methods are single-domain based, which generates predictions based on single rating matrix. These methods can only find similar users or items in a single domain. However, in practical recommendation scenarios, multiple related CF domains might be presented at the same time and then finding similar users or items across domains becomes possible, such that common rating knowledge might be shared among various related domains.

Cross-domain collaborative filtering (CDCF) intends to share the common rating knowledge across multiple related CF domains to enhance the prediction performance. CDCF methods extract knowledge from various related input domains containing additional user’s preference data to improve recommendations in the target domain. CDCF can also benefit multiple data owners by improving quality of service in different related domains.

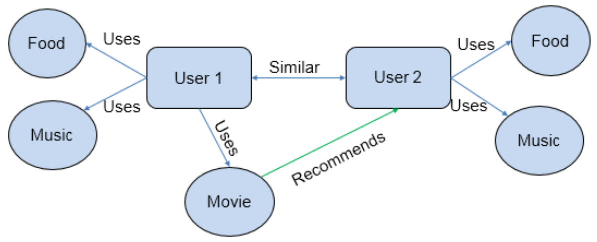


Fig 1: Generic Cross Domain CF Architecture

Major Issues in current CDCF system:

a. User Interest Drift Over Time:

User’s taste keeps on changing over the changing time because it continuously gets affected by multiple factors such as moods, contexts, cultures, festivals, seasons etc. For example, a person who does not like animation based movies, might like to watch them in future due to superior 3D animation technologies with very good visual and sound effects or if the person having baby might like to watch it now which he has never watched before. Another example would be a person buying outfit items of a specific type may like to buy the outfits of different style if he/she is relocating to an altogether different geographical area.

b. Data sparsity:

Many commercial recommender systems are used to evaluate large item sets (e.g. Amazon.com recommends books and CDnow.com recommends music albums). In these systems, even active users may have purchased and recommended for under 1 percent of the items (very negligible quantity out of total 2 million books). Accordingly, a recommender system based on nearest neighbor algorithms might not be able to suggest any item recommendations for a user. Thus, the accuracy of recommendations would be poor.

2. MOTIVATION

In most of the CDCF approaches, more stress is given on referring the historical ratings from multiple sites of the related domains. But there is a problem in recommender systems that the user interests keep drifting over a period. Hence the temporal domains should also be considered in the cross-domain recommender systems. As users’ interests keep changing over time and a user is more likely to be interested in different item categories at different point of time, we cannot simply refer the previous rating patterns of the same user in the historical temporal domains. We should also apply the current time’s context to enhance the recommendations which will be more suitable with current time. For example, a person who has watched the movies of a particular category, might not like the recommendations on the movies of similar category again and again. His interest might change over the time context. Hence the recommender system should not just calculate the historical temporal ratings but should also apply the contextual parameters for the current time domain's context variance.

3. LITERATURE REVIEW

The Cross domain collaborative filtering is an evolving research topic in recommender systems. The survey was cried out to study two major issues in current CDCF systems viz. user interest drift over time and rating sparsity.

[1] was studied to understand the basic Cross-domain recommender systems. In cross-domain, we refer multiple auxiliary domains for user ratings/inputs from related domains and then use these ratings to transfer the collective rating matrix to the target domain for recommendations. For example, to recommend a movie of 3D animation genre to the end user, the target domain is 3D animation genre movie. So we consider multiple input auxiliary domains as books, music, and other movie genre similar to the movie subject etc. It used a derived method based on a Bayesian latent factor model which can be inferred using Gibbs sampling. This addresses the challenge of modeling user-interest drift over time by considering the historical time slice patterns of the user.

[2] and [3] were studied which deal with data sparsity reduction problem. In [2], Latent factor model is used based on the ONMTF framework to cluster the users and items in Rth domain simultaneously and latent space model is used to construct the cluster-level rating pattern of user-item co-clusters. But It considered only one dimension for rating the cross domains.

In [3], principled matrix-based transfer learning framework is used that considers the data heterogeneity. The principle coordinates of both users and items in the auxiliary data matrices are extracted and transferred to the target domain to reduce the effect of data sparsity.

[4], [5] and [6] deal with different methods of cross domain collaborative filtering with temporal domain.

In [4], the temporal domains are considered. The ratings provided by the same user at different time may reflect different interests as those ratings are provided by different users. The proposed algorithms are a variant of standard neighborhood-based (either user based or item based) CF. The main idea behind the proposed approaches is to enhance inter-domains edges by both discovering new edges and strengthening existing ones. In [5], Factorization model and item-item neighborhood model are used. In both factorization and neighborhood models, the inclusion of temporal dynamics proved very useful in improving quality of predictions, more than various algorithmic enhancements. [6] uses derived method based on a Bayesian latent factor model which can be inferred using Gibbs sampling. This deals with user interest drift over time in single domain.

While working in CDCF recommender system, one critical aspect is how to collect and transfer the knowledge from different input auxiliary domain to the target domain for giving the recommendations.

The survey was done on [7], [8] and [9] to understand the methods of transferring the knowledge across the domain. These Transfer learning methods are used to first individually collect the ratings matrix for each auxiliary input domain and then transferring the collective ratings to the target domain.

[8] uses Principled matrix-based transfer learning framework that considers the data heterogeneity. The principle coordinates of both users and items in the auxiliary data matrices are extracted and transferred to the target domain to reduce the effect of data sparsity.

Table 1: Overall Evaluation of Related work

Reference Number	Paper Name	Data Sparsity	User Interest drift
1	Cross-domain recommender systems	No	Yes
2	Can movies and books collaborate? Cross domain collaborative filtering for sparsity reduction	Yes	No
3	Transfer learning in collaborative filtering for sparsity reduction	Yes	No
4	Cross-domain collaborative filtering over time	No	Yes
5	Collaborative filtering with temporal dynamics	No	Yes
6	A spatio-temporal approach to collaborative filtering	Yes	Yes
8	Transfer learning for collaborative filtering via a rating-matrix generative model	Yes	No
10	Rating Knowledge Sharing in Cross-Domain Collaborative Filtering	Yes	Yes

[10] was studied rigorously which deals with the cross domains over different sites (domains), transferring the rating knowledge from these sites to recommend in target domain by using knowledge transfer method. Along with using multiple sites/domain to collect the user ratings, it also extends the model of [4] for using temporal domain to deal with the issue of user interest drift over time. It uses the principle that user has multiple counterparts across temporal domains and the counterparts in successive temporal domains are different but closely related. Series of time slices are meant as related domains and a user at the current time slice depends on the previous historical time slices. Model is built on a cross-domain CF framework by viewing the counterparts of the same user in successive temporal domains are different but related users. If we can find the unchanged rating patterns (static components) shared across temporal domains, the drifting factors of users (changing components) in each temporal domain can be easily captured. Bayesian generative model to generate and predict ratings for multiple related CF domains on the site-time coordinate system, is used as the basic model for the cross-domain CF framework which is extended for modeling user-interest drifting over time.

Table 2: Specific Evaluation of Related work (for user interest drift over time issue)

Paper Name	Using historical time slice data	Using Current time context
Cross-domain recommender systems [1]	Yes	No
Cross-domain collaborative filtering over time [4]	Yes	No
Collaborative filtering with temporal dynamics [5]	Yes	No
A spatio-temporal approach to collaborative filtering [6]	Yes	No
Rating Knowledge Sharing in Cross-Domain Collaborative Filtering [10]	Yes	No

The above evaluation of related work shows that the current work is mainly dealing with the temporal domain ratings from the historical data. The proposed work will deal with the historical ratings along with the current time frame’s contextual parameters like season, holiday, locations, category etc. This will enhance the ratings and will be more current in time.

4. SYSTEM OVERVIEW

The aim of proposed work is to implement the concept of context based parameters in the cross-domain recommender system for historical and current time slices of the user over cross domain techniques. The ratings calculation is done to obtain the average recommendations for the movie/book selected by the user. The project aims at developing an efficient system architecture that enables the calculation of ratings based on various contextual parameters for both historical and current time slices.

For a recommender system, various types of contexts can be used as per the applications’ requirements. Physical context represents the time, position, and activity of the user, but also the weather, light, and temperature when the recommendation is supposed to be used. Social context represents the presence and role of other people (either using or not using the application) around the user, and whether the user is alone or in a group when using the application. Interaction media context describes the device used to access the system (for example, a mobile phone or a kiosk) as well as the type of media that are browsed and personalized. The latter can be ordinary text, music, images, movies, or queries made to the recommender system. Modal context represents the current state of mind of the user, the user’s goals, mood, experience, and cognitive capabilities.

Proposed idea calculates the recommendation ratings for the movie/book in which user is interested in using contextual parameters. Contextual parameters such as season, holidays, location and category are applied to the historical ratings of the users and to the current user’s time context. Then to calculate the final recommendations, both the historical and current contextual parameters will be mapped together. It will also apply mixing of cross domain rating recommendations to handle the data sparsity in movies and books domains.

Database containing movies and books ratings given by various users is the input to the recommendation system along with the contextual parameters applied for each time slice. The output is final recommendation for the movies and books for which user is interested in along with other related movies and books in similar categories/context.

The proposed work considers the current time domain context along with historical data. This can be viewed in multiple contexts/aspects. Contexts can be added based on the season changes, location changes, cultural changes, festival periods etc. By using these current contextual parameters along with the existing work of historical time domain, the user recommendations will be more effective and more current addressing the current context of the user.

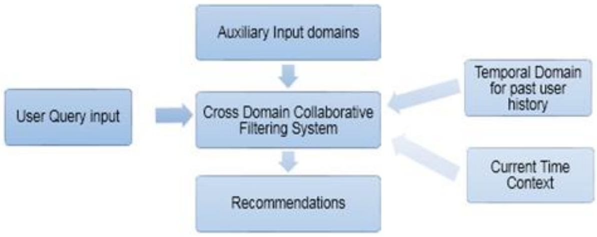

Auxiliary Input domains
User Query input
Cross Domain Collaborative Filtering System
Temporal Domain for past user history
Current Time Context
Recommendations

Fig 2: Architecture for Proposed System

4.1 Algorithm

Input: Movie and Novel ratings dataset

Output: The recommended movie and novel ratings as per user query

```
String event = {season, holiday, location, movie category}; // Initialize the contextual parameters
movie_mat=[user][movie]; // Initialize movies rating matrix.
book_mat=[user][book]; // Initialize books rating matrix.
current_user ← {season, holiday, location, movie category}; // Calculate user state
sliceM_mat ← movie_mat/12 ;
sliceB_mat ← book_mat/12 ;
For (int i=0;i<12;i++) {
//Compare user context with event context
sum_rating += rating; }
avg_rating ← sum_rating / total_ratings_cnt
```

4.1.1 Dataset pre-processing module

This is used to pre-process the historical movie and books ratings data. The dataset pre-processing will divide the input dataset of each year into 12 different slices having one slice per month as opposed to the 4 slices per year in the base system. By this more granular level rating matrix can be generated which is more aligned with the ever-changing user interests. This can provide better ratings recommendations. Also, the contextual parameters will be applied on these slices on granular levels so that the season, holidays, locations will be applied more closely to the current matching context.

Step1: Collect ratings data from 3 different movie rating sites.

Step2: Apply data enrichments using various parameters such as of location, season, holidays and movie category.

Step3: Divide the user ratings in 12 slices with 1 slice per month so that we can get fine grained ratings

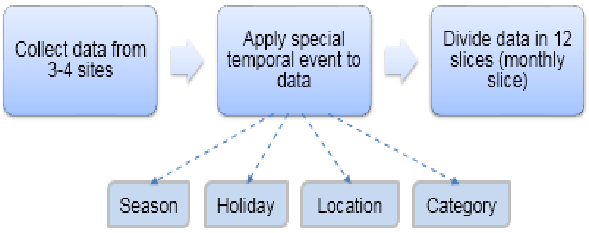

Collect data from 3-4 sites
Apply special temporal event to data
Divide data in 12 slices (monthly slice)
Season
Holiday
Location
Category

Fig 3: Flow diagram of the Data Preprocessing module

4.1.2 *Main module*

The pre-processed dataset will be used to calculate the ratings recommendations. The historical time slices along with contextual parameters will be stored in the database for past 3 years with monthly time frame per slice.

Datasets of two different domains for movies and books are used for cross-domain functionality. Bayesian network word analyzer is used to calculate the similarity index of the 2 domains. If the similarity index is 80% or above, then those two domains are similar and could provide better recommendation results in cross-domain framework.

When the current user enters a movie/book name to see the ratings, the current user's state will be assigned with the same contextual parameters of location, season, holidays and movie category. Then the mapping of current user's state will be done with the historical ratings to calculate the actual rating recommendations.
From the mapping calculations for all the parameters, the average movie ratings will be calculated and the average from all the time slices will be finally calculated for presenting the recommendations result to the end user.

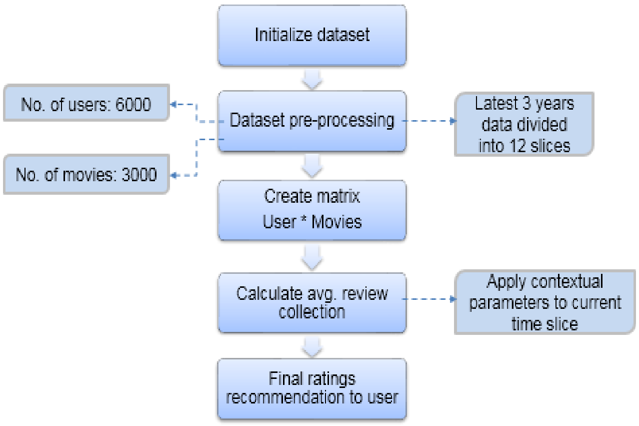

Initialize dataset
No. of users: 6000
Dataset pre-processing
Latest 3 years data divided into 12 slices
No. of movies: 3000
Create matrix User * Movies
Calculate avg. review collection
Apply contextual parameters to current time slice
Final ratings recommendation to user

Fig 4: Flow diagram of the main system

Also, the rating recommendations for some other related movies and books in the similar movie category will be presented to the end user. It can be customized as per user's interest and then filtering as per the contextual parameters by applying post-filtering contextual techniques. This will help the user to choose from other similar movies and similar books which are available in the database.

Contextual postfiltering (or contextualization of recommendation output) is used in the proposed system. In this paradigm, contextual information is initially ignored, and the ratings are predicted using any traditional 2D recommender system on the entire data. Then, the resulting set of recommendations is adjusted (contextualized) for each user using the contextual information.

System Components:

1. **Data Pre-processor component:**
   - Load ratings data from multiple sites of movies and books
   - Apply special contextual parameters of different events such as location, season and category to the data
   - Divide data in 12 monthly slices

2. **Rating Analyzer:**
   - Get user query for the movie or novel search and customization details if applied by the user
   - Create user's contextual event parameters
   - Compare user's query contexts and customizations with existing data set by mapping with the historical ratings to calculate the rating recommendations.
   - Provide ratings for the queried movie or novel.
   - Also, provide the ratings and recommendations for other related movies and novels in the similar context. It can be customized as per user's interest and then filtered as per the contextual parameters.

4.2 Mathematical Model

Step 1:
Select the highly available movies from the Netflix.com dataset for past 3 years. Initialize User * Movie matrix for the users and movies.
Let c be the number of movies and r be the number of users.
The ratings will be denoted as Ui Mj = Rmij

$$\begin{array}{c|cccc|}
 & M_1 & M_2 & M_3 ..... & M_c \\
U_1 & R_{11} & R_{12} & R_{13} ..... & R_{1c} \\
U_2 & R_{21} & R_{22} & R_{23} ..... & R_{1c} \\
U_3 & R_{31} & R_{32} & R_{33} ..... & R_{3c} \\
. & & & & \\
. & & & & \\
U_r & .... & .... & .... & R_{rc}
\end{array}$$

Similarly prepare the rating matrix for books ratings.
The ratings will be denoted as: Ui Bj = Rbij

Step 2:
Divide the ratings matrix data in 12 different slices with one slice per month and save in the Slice Matrix set table.
Slice_matrix = User_matrix / 12

Assign season, holiday, location and category as the contextual parameter values for each movie/book rating in the Slice matrix set.

Um = [Um_season, Um_holiday, Um_location, Um_category]

Ub = [Ub_season, Ub_holiday, Ub_location, Ub_category]

Step 3:

Calculate the similarities between movies and the books in the matrix based on categories using Bayesian theory. Similarity index should be 80% or more to have the cross-domain functionality.

Step 4:

Get user input query for the interested movie/book as Uq. Assign season, holiday, location, category contextual parameter values for Uq as

Uq = [Uq_season, Uq_holiday, Uq_location, Uq_category]

Step 5:

Compare event contexts array in different historical time slices with current user's context array to calculate the review ratings:

sum_review = ∑ Rum | Rum => Ruq

Calculate average review ratings from all time slices:

avg_review = sum_review / review_count

Step 6:

Provide the cross-domain recommendations of both movies and books for context based filter criteria if applied.

Step 7:

The evaluation matrix of the system is calculated by Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters.

RMSE = $\sqrt{\sum_{i\in S}(r_i - \hat{r}_i)^2/|S|}$.

MAE = $(\sum_{i\in S}|r_i - \hat{r}_i|)/|S|$.

4.3 User Interface of the Application

The graphical user interface of the application is simple and self-explanatory. The user should provide the user id and password to the application as the system allows only authorized users to enter. Admin and user are the two roles provided. After authentication, the Admin is asked to select the datasets of different domains and then click to calculate the similarity index. Also, the contextual events should be applied and results are to be saved. With the user login, any user can provide his query for interested movies or novels. User can select the movie/novel name from the provided drop down box if available and click on submit. The results are displayed in 3 different frames. First frame shows the rating recommendations of the selected movie/novel. Second frame shows the other related movies and their ratings. Third frame shows the other related novels and their ratings.

The user can also search as per the context as in category, writer name, season and location. Then the recommendation results are displayed based on those selected contextual parameters.

5. RESULTS AND DISCUSSION

The contextual parameters based temporal domain is used for recommendation system. We have used dataset from MovieLens.com. For the historical movies and books ratings dataset, the contextual parameters of season, holiday, location, and category are applied in the database. Consider a user submits the query for the recommendations of a particular movie or book. The system calculates the contextual parameters for the user's current state before calculating the ratings recommendations. These contextual parameters are compared with the parameters of historical movies and books ratings and those with matching parameters will be picked up to calculate the rating averages. This fetches the ratings recommendations which are in close match with the current user's likings and the current temporal domain scenario.

Also, the other related movies and books with similar category and closely matching contextual parameters will be presented to the user. These recommendations will be more in line with the user's current time domain and will be based on the customizations as per his own taste if he has rated any movies or books in the past. By this, the system takes care of the same user's historical selection patterns and the similar likings of other categories which he might not have seen/rated before but those movies or books which might closely match with the user's current temporal domain context such as season, holiday and location.

For example, the user is searching for a movie in comedy category then the system will provide the ratings of that movie by matching the user's contextual parameters with historical ratings' contextual parameters. System will also provide the recommendations of some other movies in let's say children category movies if the current season is of summer vacations and user has not customized input for movie category. Also by applying cross domain parameters, the books based on the queried movie and its category, will also be recommended to the user.

Below graphs show the expected results of the qualitative parameters of the base system with the proposed system. By using the contextual parameter based temporal domain system, the quality of the recommendations is more granular and closer with the user's interest in current time context scenario.

The performance of the proposed system is calculated based on the qualitative parameters of the rating accuracy. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters of the proposed system are lower than the base system due to mixing cross domains. They are calculated for 300 samples of the dataset. The performance of the proposed system on the qualitative parameters of the rating accuracy is better with more granular recommendations and mixing of sites. Hence the recommendations are closer with user's interest in current time contextual scenario.

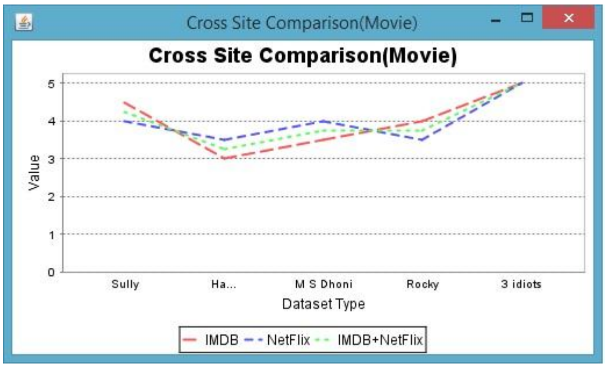


Fig 5: Cross Site Comparison for Movies

Below table shows the comparative values of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) parameters of the base and proposed system

Table 9.1: Comparison Results of RMSE and MAE

Parameter name	Base System	Proposed System
RMSE Movie	0.860	0.846
RMSE Novel	0.916	0.887
MAE Movie	0.671	0.632
MAE Novel	0.806	0.796

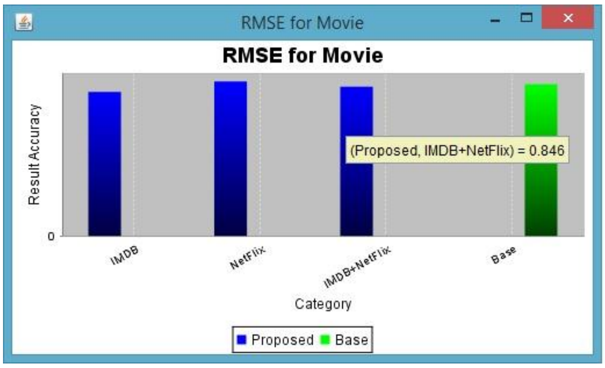


Fig 6: RMSE for Movies

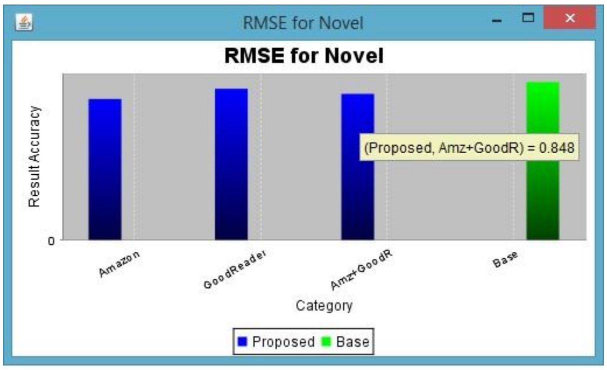


Fig 7: RMSE for Novels

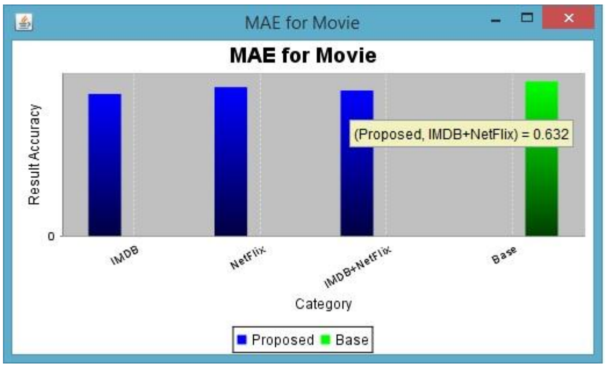


Fig. 8 MAE for Movies

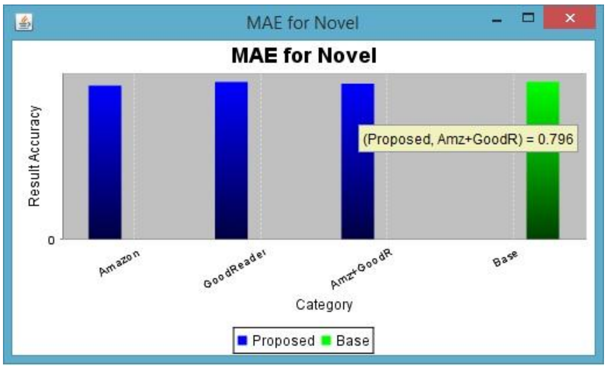


Fig. 9 MAE for Novels

6. CONCLUSION AND FUTURE WORK

The proposed work deals with the addition of temporal domain and contextual parameters to the historical movie and book ratings and slicing them on more granular level. The current user's state is calculated based on these contextual parameters. As per user's input query, the enhanced ratings recommendations for similar movies and books are presented by matching the user's contextual parameters in cross-domain matrix.

The future work may involve addition of more related domains to deal with data sparsity. It can also be integrated with social networking sites to provide rating recommendations on related domains. We can also implement machine learning system for the ratings datasets for analytics purposes.

7. ACKNOWLEDGEMENT

This work would not have been possible without the support and help of many individuals. I would like to extend my sincere thanks to all of them. I am indebted to Prof. Manisha Patil, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant supervision as well as all the other staff members for providing important information regarding the work.

8. REFERENCES

[1] P. Cremonesi, A. Tripodi, and R. Turrin, "Cross-domain recommender systems," in Proc. IEEE 11th Int. Conf. Data Mining Workshops (ICDMW), Vancouver, BC, Canada, 2011, pp. 496–503.

[2] B. Li, Q. Yang, and X. Xue, "Can movies and books collaborate? Cross domain collaborative filtering for sparsity reduction," in Proc. 21st Int. Joint Conf. Artif. Intell. (IJCAI), 2009, pp. 2052–2057.

[3] W. Pan, E. W. Xiang, N. N. Liu, and Q. Yang, "Transfer learning in collaborative filtering for sparsity reduction," in Proc. 24th Conf. Artif.Intell. (AAAI), 2010, pp. 230–235.

[4] B. Li et al., "Cross-domain collaborative filtering over time," in Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI), 2011, pp. 2293–2298.

[5] Y. Koren, "Collaborative filtering with temporal dynamics," in Proc. Int. Conf. Knowl. Discov. Data Mining (KDD), Paris, France, 2009, pp. 447–456.

[6] Z. Lu, D. Agarwal, and I. S. Dhillon, "A spatio-temporal approach to collaborative filtering," in Proc. 3rd ACM Conf. Recommender Syst., New York, NY, USA, Oct. 2009, pp. 13–20.

[7] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in Proc. 25th Int. Conf. Mach. Learn. (ICML), Helsinki, Finland, 2008, pp. 880–887.

[8] B. Li, Q. Yang, and X. Xue, "Transfer learning for collaborative filtering via a rating-matrix generative model," in Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML), 2009, pp. 617–624.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[10] Bin Li, Xingquan Zhu, Ruijiang Li, and Chengqi Zhang "Rating Knowledge Sharing in Cross-Domain Collaborative Filtering" in IEEE TRANSACTIONS ON CYBERNETICS, VOL. 45, NO. 5, MAY 2015

[11] B. Li, "Cross-domain collaborative filtering: A brief survey," in Proc. 23rd IEEE Int. Conf. Tools Artif. Intell. (ICTAI), Boca Raton, FL, USA, Nov. 2011, pp. 1085–1086.

[12] Swapna Joshi and Prof. Manisha Patil "Effective Cross-Domain Collaborative Filtering using Temporal Domain – A Brief Survey" International Journal for Modern Trends in Science and Technology Volume: 02, Issue No: 10, October 2016 ISSN: 2455-3778.Pg.88-92 Oct 2016.

[13] Swapna Joshi and Prof. Manisha Patil "Enhanced Rating Recommendations using CDCF in Temporal Domain" at, Smt. Kashibai Navale College of Engineering, in International conference on Internet of Things, Next Generation Networks and Cloud Computing 2017 (ICINC-2017).

[14] https://www.amazon.com

[15] Ms. Sanjeevani Dhaneshwar , Mrs. Manisha Patil "Context Awareness in Data Mining Applications – A Survey" International Journal of Science and Research (IJSR) Volume 5 Issue 1 ISSN (Online): 2319-7064 Pg no 253-255 January 2016.

[16] Ms Manisha Kumbhar, Dr. Rajesh S. Prasad "Genres based Collaborative filtering: An approach to improved Quality of Recommendation". Proceeding of International Conference on Internet of Things, Next Generation Network and Cloud Computing 2016 ISSN: 0975 – 8887.Pg. no 423-427 March 2016.

[17] Ms. Jyoti Pandey , Mrs. Manisha Patil "Recommender System Using Clustering Based On Collaborative Filtering Approach" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (4) , , 3419-3423 ISSN-0975-9646 Pg.3419-3423 2015.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 362-368, 2017, ISSN:-2319–8656

An SOA Framework for Web-based E-learning Systems – A case of Adult Learners

Eric Araka
School of Computing and Informatics
University of Nairobi
Nairobi, Kenya

Lawrence Muchemi
School of Computing and Informatics,
University of Nairobi
Nairobi, Kenya

Abstract: The existing e-learning design models for adult learners are mostly built on pedagogical principles and are more appropriate for younger learners. They may not support learning activities that may be suitable for adult learners. To make use of adult learners' principles, a dedicated learning model based on SOA, and intended to make learning process collaborative while allowing learners to use their experiences and self-directedness was developed. The SOA framework was able to deliver a remote service into a web-based LMS; in this case Moodle-based LMS so that it can be internally utilized by the adult learners in accessing learning materials from other systems as services. Analysis from the evaluation process was adequate enough to give a clear effect of using the SOA framework when compared to the initial survey carried out on the current e-learning systems which the learners used in their learning.

Keywords: eLearning, SOA, LMS, Andragogy, Adult learners

1. INTRODUCTION

The e-learning environments that are used to instruct adult learners are the same ones that are used to instruct children and do not obviously support diverse learning activities that may be suitable for adult learners. Pedagogy is not as useful in an adult learning environment because it does not utilize the learner's capabilities, experiences and adult learning characteristics [4] [5].

The wealth of real-life experiences that adult learners possess is a great resource that can be utilized for their learning. These experiences cannot be utilized in the current e-learning environments, which are only used to deliver information to the learners. Adult learners need an environment that can enable them to be involved in the construction of their own course content, giving then the freedom to make use of the experiences that they possess [10][11]. Bichelmeyer [3] argues that to this end e-learning for adult learners has been presented as though it only involves only one type of educational experience as it only provides learners with information they need with less emphasis on the learning process itself.

This research focuses on a design model of e-learning systems based on the theory of andragogy [14] (as an alternative to pedagogical model of instruction) to enhance learning for adult learners.

2. LITERATURE REVIEW

2.1 Theory of Adult Learners

Adult learners are those who perform roles associated with adults by one's culture e.g., workers, spouses, parents and perceive themselves to be responsible for their own lives [12]. Adults are experienced, self-directed and are independent thinkers in their learning and they seek to learn from and about their social and work environments and the roles they play there as opposed to young leaners. Learning by being guided by others is therefore unfitting in adult education environments [14].

According to Knowles [13] [14], andragogy rests on four crucial assumptions about adult learners and how they differ from child learners (1) their self-concept moves from dependence to self-direction, (2) their growing reservoir of experience begins to serve as a resource for learning, (3) their readiness to learn becomes oriented increasingly toward the developmental tasks of their social roles, and (4) they begin to want to apply what they have learned right away to life's real challenges. Adults' orientation towards learning shifts from one of subject centeredness to one of problem centeredness.

The review of andragogy theory indicates that adult learners are characterized by a great experience and self-directed learning [6]. E-learning systems for adult learners, therefore, should enable learners go through instructional materials delivered via the Web at their own pace with no or minimal interaction with an instructor and be involved in content creation. This theory also implies that adult learning should therefore emphasize on knowledge construction by learner actively exploring and discovering for.

Adult learners possess more life and domain specific knowledge, different motivations to learn, and more available resources than young learners [3]. Basing on this assumption, and adult learners' characteristic of self-directedness and experiences, the web-based system for e-learning described in this study allows adult learners to use this reservoir on experience and to be involved in the creation of content for learning.

2.2 Service Oriented Architecture & Web Services

Service Oriented Architecture (SOA) is a description of how different parts of the system interact and communicate to achieve a desired result. It is "an interconnected set of services which in its basic form is a message-based interaction between software agents, each accessible through standard interfaces and messaging protocol" [16]. SOA is implemented through the use of web services [17].

A number of SOA e-learning frameworks have been developed. These frameworks include the NSDL [21], OKI-Model [21], IMS DRI [22] and as summarized by [15]. Scott (2003) also describes an SOA framework that integrates different systems of an institution that occupy vertical positions e.g. VLE, Library Management System and Student Records (MIS). The framework integrates the components of different systems in an institution in order to reduce replication/overlapping of functions.

"*A web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format called WSDL. Other systems interact with the web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards*."[1]

The importance of utilizing SOA in e-learning systems includes integration, interoperability, scalability, and reusability [17]. Interaction, knowledge building and collaboration features of e-Learning systems cannot be ignored and it is necessity that they be considered in the adoption of SOA in adult e-Learning systems. The attributes of web services that include reusability, composability, discoverability and loose coupling [16][17] extends the importance of employing SOA in e-learning systems to allow integration of different systems to communicate and utilize the services [18] they contain from each other.

There are a number of disparate systems that expose learning resources as web services and can be utilized within LMSs through the SOA framework. These web services include;

1. **Repository of learning objects** based on web services which in the context of this study refer to reusable content components for education and training.
2. **Application Programming Interfaces**: There are a number of programs written and stored on various applications as learning objects and may be accessed through APIs that are exposed as set of related web services that can be accessed through different protocols. This when brought close to students through the framework discussed in this study allows students to access written programs that are executed within the LMSs and enables students to solve complex programming exercises without having to install the API on their machines and even allow collaboration and grading during and after the programming exercises [15]
3. **Evaluation engines.** There are a number of evaluation engines that provide electronic quizzes, computer-scored homework assignments, and practice exams.

The Prototype based on the SOA framework developed through this study is embedded in a Moodle-based LMS to allow for the access of the above services. The communication between the Service Oriented LMS (service consumer) and the other web based systems (service providers) is through message passing through the OAuth protocol which uses HTTP POST requests in its transport layer.

The SOA framework enhances adult learning while taking consideration of the adult learning principles. The framework helps adult learners employ the great reservoirs of experiences gained through professional work skills and knowledge in material construction as well as provision of services that will allow for opportunities for rehearsal, feedback, application, and transfer. The interaction and collaboration services are oriented in the current e-learning systems so as to be fit for adult learning.

3. METHODOLOGY

This study used an approach that involved the following procedure. First, an existing pedagogical-based eLearning system based at one of the leading public University in Kenya was investigated on the extent to which it implemented the adult learners' characteristics. Second, a survey (on a sample set of 228) to establish the motivational level of the adult students who were carrying their studies through e-learning was done. The results from the investigation and survey led to the development of a framework describing the principles of adult learners that need to be incorporated in an SOA framework. An andragogical-based e-learning framework that runs on an SOA model was developed from the adult learners' principles framework (Figure 1).

A prototype based on the andragogy based eLearning framework was then developed and was used in a simulated Moodle-based LMS. This prototype was used to validate the proposed SOA framework. A sample comprising of 58 students was used in the validation process.

4. RESULTS, ANALYSIS & DESIGN

The findings from the approach described in the methodology are as follows.

4.1Findings from the investigation of the current e-learning environment

The investigation of the eLearning system was important as it helped in the description of the design and the suitability of the e-learning environment to the adult learners who were engaged in this study. The main objective of the investigation was mainly to understand the design principles of the existing eLearning system and then provide the context for the design of an SOA based model for adult learners.

It was established that the e-learning environment used to instruct study participants who are adult learners is the same one used to instruct the undergraduate students. Grounded on the pedagogy principle, the e-learning system is used especially to deliver content or learning materials to the learners. The learning environment does not utilize the learner's capabilities, experiences and adult learning characteristics [4] [5].

4.2Findings from the study Survey

The survey was carried out to establish the level of motivation and satisfaction on the adult learners who used the current eLearning system put in place for their learning. Part I of the survey collected background information regarding the survey participants. The respondents who participated in the survey are learners who are pursuing their postgraduate studies. Majority of the participants belong to the age group between 25 – 34 and 35 – 44 which represents 54% and 36% respectively.

From the analysis, 94.5% of the number who were involved in the research displayed the characteristics of an adult learner i.e. 79% were married, playing a role of a parent/spouse and had accomplished the undergraduate student period of four years of traditional college/university learning, 42% of whom are employed as working professionals. When asked for the reason for enrolling for the e-learning course only 4% indicated that their main reason was to receive a certificate, while 96% indicated that they enrolled in the e-learning course in order to develop and enhance their skills. This analysis agrees with what Knowles defined who an adult learner is [13].

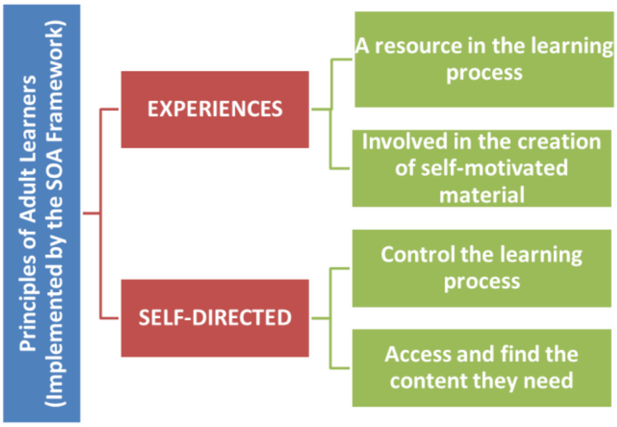
Part II of the study explored the motivation/satisfaction of their learning process through e-learning. When asked about the frequency of their interaction with an instructor, a majority of those surveyed responded that they occasionally had interactions while they took the e-learning course. In particular, 12% of the respondents indicated that they never had such interactions and 76% of those surveyed responded that they occasionally or seldom interacted with an instructor.

As Holton [6] argues, 96% of the respondents indicate that they like learning on their own pace. This reflects that any learning environment should give the adult learners intellectual freedom, experimentation and creativity. Sixty four (64) % of those involved in the survey indicated that their achievements and previous experiences were not utilized in the learning process. Cook asserts to this as she argues that researchers with expertise in education and information communications technologies have not applied their findings to the adult learners. She continues to argue that this has resulted in teaching methods and strategies that are ineffective in teaching and instructing the adult learners. Of the adults who participated in this study 81% responded that they prefer interaction and collaboration with other students, as noted in the following comment by a participant who was a full-time working professional: "*Free interaction with other students and instructors and sufficient e-learning material provided"*

Adult learners possess characteristics that should be incorporated in their training. Experience, a great resource that can be tapped in the learning process and self-directedness, through which adult learners can participate in the learning process. These are special components in adult learning process which allow the learners pick and choose what they are only interested with.

From the survey conducted in this study it is clear that the current design models are not suited for adult learners. The same environment is used for both andragogy and pedagogy classes.

Figure 1: Principles of adult learners implemented by the SOA Framework

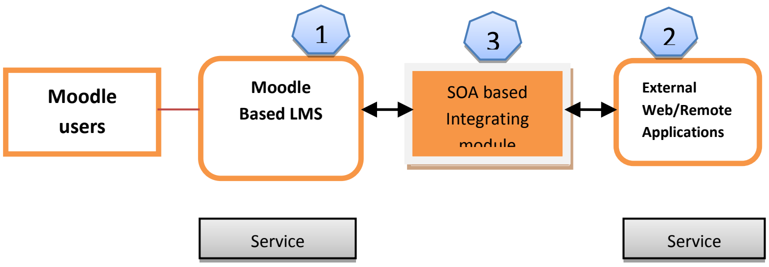

Principles of Adult Learners
(Implemented by the SOA Framework)
EXPERIENCES
A resource in the learning process
Involved in the creation of self-motivated material
SELF-DIRECTED
Control the learning process
Access and find the content they need

4.3 The SOA Framework Design

4.3.1 Context diagram of the SOA based integrating framework

The SOA framework enables the Moodle Course students while logged into the LMS to connect over to other remote tool, blog or another LMS and be automatically authenticated and allowing them proceed to use their experience in selecting the content they want in the Moodle course(s).

Figure 2: Context diagram of the SOA based integrating framework


1
3
2
Moodle users
Moodle Based LMS
SOA based Integrating module
External Web/Remote Applications
Service
Service

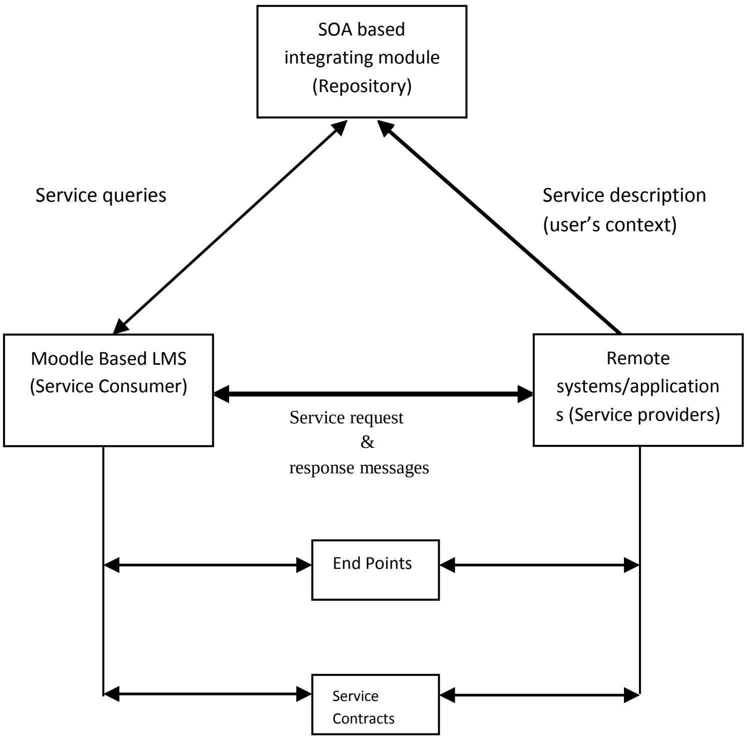
The architecture comprises of three parts: **Service Consumer - Moodle LMS,** acting as the service consumer, and containing the users of the e-learning management systems; **External Web/Remote Application**, acting as the service provider, it stores the remote content to be accessed by users of the LMS system upon clinking the Launch URL of the LTI integrator; **SOA based integrating module**, is the application at the core of this architecture which provides an API between the service consumer (Moodle LMS) and service provider (remote application). The **SOA based integrating module** is implemented through the use of IMS specification standards and allows embedding of the remote system in the LMS by assigning information in the launch process such as launch URL, shared key and the secret.

4.3.2 The SOA Framework Architecture

The SOA Framework provides a way for the service consumer (Moodle LMS) to send a user to another system i.e. service provider (this is the service that integrates with the LMS). It allows the user to be authenticated and allow access to a specific course when the service provider renders the content.

The service provider and service consumer communicates through the use of a consumer key and a shared secret which allow any message to be passed between the two systems. The messages are signed using the OAuth (soap uses this protocol over the transport layer) protocol for secure API authorization.

Figure 3: The SOA Framework Architecture



The **endpoints** provides communication to the service providers and the context for the connection, that is, what service has been requested, where to locate it and its final presentation on the service consumers browser. This is achieved through a graphical user interface which provides for configuration module where details about point-to-point connection between LMS and remote applications are entered. Here, the service consumer which is the LMS knows the endpoint (as configured) and send request is launched when the user clicks on the link(s) provided within the course.

The **service contracts** in this framework define the relationship between the provider and the consumer, that is, what the service provider will give to the consumer. The contract actually defines what functionality the provider provides to the LMS, what data it will return and in what form. The agreement between the LMS and service providers are entered into when the administrator fills out a form that typically provides the details governing the two interacting parties detailing what happens after the connection established after authentication.

4.3.3 Messages and Communication between the SOA components

The SOA Framework provides a way for the service consumer (Moodle LMS) to send a user to another system i.e. service provider (this is the service that integrates with the LMS). It allows the user to be authenticated and allow access to a specific course when the service provider renders the content. The service provider and service consumer communicates through the use of a consumer key and a shared secret which allow any message to be passed between the two systems. The messages are signed using the OAuth (soap uses this protocol over the transport layer) protocol for secure API authorization.

4.4 Principles of Adult Learners implemented by the SOA framework

This framework addresses the principles of adult learners as cited in Figure 1 as follows:

4.4.1 Experience

a. Resource in the learning process & Creation of self-motivated learning material

The experience that adult learners possess is a great resource in the learning process. The design model, based on SOA, is intended to make eLearning process collaborative while allowing the learners to use their experiences in publishing articles within an e-learning system and to allow them share what they learn. The published articles are subjected to thoughtful contributions from other faculty groups; informal-competency groups i.e. those who possess similar levels of knowledge and experience. As identified by Rodrigues [19] blogging, the method adopted by this research, leads to transformative learning which is based on learners' experiences and encourages reflection and free thinking leading improvement of knowledge. The combination of the contributions from the learners leads to material that one can use to refer to a specific topic of learning as contributed by different adult learners in their professional fields.

4.4.2 Self-Directedness

a. Control the learning process

The SOA framework when implemented on eLearning systems for adult learners, enables learners go through instructional materials delivered via the Web at their own pace with no or minimal interaction with an instructor.

b. Access and find the content they need

This is reflected in the SOA framework by providing to learners an e-learning environment that gives the adult learners intellectual freedom, experimentation and creativity. The different sources of learning materials exposed to the learners through the design model enables them access and find the content they need. This is in contrary to what the current pedagogical models provide.

5. VALIDATION OF THE SOA FRAMEWORK

The validation process was carried out on the prototype that was developed and simulated on the Moodle-based LMS. The purpose of the validation was used to establish whether the adult learners' characteristics identified in this study were incorporated in SOA framework. These included;

a) Use of **adult learners experience** in the enhancement of learning process and **creation of self-motivated learning materials**.

b) **Self-directedness of adult learners so that;**

i. They can control the learning process.
ii. They can access and find the content they need.

5.1 Procedure used for evaluation

The validation process involved the following procedure. First, the prototype was made available to the learners for inspection using Abstract Tasks (AT) approach. Abstract Tasks (AT) inspection method describes activities to be performed during the inspection and captures usability experience while identifying the application features on which it is important to focus inspection and describes the actions the inspector should perform during the evaluation [2]. AT inspection aims at allowing inspectors who may not have a wide experience in evaluating e-learning systems to perform accurate evaluations by performing specific tasks. A survey (on a sample size of 58) to establish increment on the motivational level compared to the initial survey (carried on the pedagogical based e-learning system) was then carried out. Secondly, system logs/views were collected from the prototype to measure;

i. The extent which learners employed the use of their experiences in the creation of self-motivated learning materials through collaboration amongst learners. The self-motivated learning materials were created through the archived ideas exchanged by the learners and;
ii. Whether the students were able to access the sources of learning materials on their own while within the Moodle-based LMS. This was done by analyzing the number of objects (external learning sources or documents) accessed by the learners in the simulated e-learning system.

5.2 Evaluation Results

a. Results from the survey

The survey (on a sample size of 58) was carried out in comparison with the initial survey which informed the research the need for an SOA framework. The survey participants were also part of those who were involved in the survey that informed the research on the need for the SOA framework. The inspectors were first required to carry out the Abstract Task activities before responding to the questions raised in the survey. The analysis of collected data (using SPSS tool) indicated that there was an increment on the level of motivation on the learners as compared to the level of motivation as indicated on the survey carried out on the use of the pedagogical based eLearning system which the learners used initially.

Participants in the evaluation process could access a pre-set course in the Moodle course module. From the course the learners could then access the Wordpress Site within the course. The wordpress site provided a platform through which learners could collaborate and interact. There were other links that were provided within the course module which when clicked led to access of interactive course content/resources from remote websites. These sites exposed their content as web services which were then fetched while within the pre-set course in Moodle LMS. The SOA framework delivered a design model that incorporated the adult learners' characteristics, it allowed the adult learners to feel that the experience and intellectual ability were respected and appreciated.

It was observed that 76% of learners were in agreement that the SOA-based eLearning system provides more sources of learning materials while58% indicated that they felt their experiences and previous achievements were acknowledged and respected. This is in concord with other previous studies [6] [8] [12] which established that when left to explore on their own, they feel respected, more satisfied and motivated to engage in e-learning and that adult learners like taking their own responsibility in their learning as they are independent and autonomous in their thinking [8][12]. From the survey it can be concluded that the SOA framework delivers an adult learning environment that utilizes the learner's capabilities, experiences and adult learning characteristics.

b. Results from the system views/logs

There were two types of system logs/views that were collected to establish the implementation of adult learners' experiences and self-directedness in the SOA framework respectively.

i. Measure for incorporation of adult learners' experiences in the creation of self-motivated learning materials

The system logs collected from the system were used to establish whether the students exchanges and access of remote objects/content increased positively to a point of demonstrating whether the students were able to collaborate and interact with one another within the wordpress blogging platform. The logs captured by the system indicated the number of participants who accessed the platform and either viewed other students' posts or created their own posts. There were a total of 22 students whose logs were recorded into the system and analysed as follows showing the students who accessed and viewed other students' posts. There was an average of 2.6 views for each student which indicates that approximately each student was able to view or interact with others at least 3 times.

The average number of exchanges shows that for each post created, there were 3 other posts/comments on the original posts created. This research study was able to observe that through the SOA framework, collaboration among instructor and learners was achieved as compared to the current e-learning systems where there is no collaboration at all.

ii. Measure for provision of self-directed learning

The course that was created on the SOA framework for evaluation purpose provided access to 3 external education applications that delivered remote content in the Moodle based LMS as services. Since the total number of students who accessed the remote systems was 22, the average number of access for each student to the external systems is 3 indicating that for each of the 3 educational applications each student had an access.

The system logs/views analysis shows that the collaborative and interactive learning environment which was made available for students allows the students to make use of their experiences to collaborate and interact among themselves.

6. CONCLUSION

Adults are independent, experienced and self-directed when it comes to learning. If they are treated the same way as children, they feel that their independence and experience in the knowledge gained is never appreciated or acknowledged [14].

This study established that there is a possibility of using an andragogy based eLearning framework that runs on an SOA platform. The SOA framework employs the Wordpress blogging tool in a Moodle LMS to facilitate collaboration as an important principle to adult learning process. The framework also allows access of external content which demonstrates collaboration at the institutional level and provides an environment for collaboration, creation of self-motivated learning materials amongst adult learners. When adult learners, engaging in web-based e-learning are provided with a learning environment which allows them to collaborate interactively, they feel motivated to learn and this leads to their satisfaction [8] [12]. This way, adult learners' experiences are exploited and brought to the center of the learning process.

The SOA framework integrates the principles of adult learners into the learning process hence maintaining the motivation hence satisfaction for adult learners who carries their studies via e-learning environments. This is true when the adult learners' experiences are exploited and incorporated in the learning process and that provides them with opportunities to interact with themselves and instructors [6].

The SOA framework contributes to the improvement of the motivation of adult learners who sometimes may drop out of their learning because of lack of motivation as majority of the eLearning systems currently in use are more appropriate for younger learners. Knowles [13] [14] also argues that adults should not be treated the same way as children when it comes to learning.

However the SOA framework has a perceived weakness. The blogging feature, implemented by the framework, though widely adopted by many has never been used in a formal environment to provide learning process. Therefore its adoption into the LMS through the SOA framework will only be able to provide creation of self-motivated materials informally. As its adoption into the learning process now takes place, one feature lacks; how do we make the exchanges of ideas within the LMS formal and thereafter accept the self-motivated material for use by others in future. The big question is; is it possible to make the Wordpress blogging functionality formal within an LMS and hence part of learning? If so, how will this be achieved especially through the use of the Service Oriented framework developed and implemented in this study?

7. ACKNOWLEDGMENT

The author(s) wish to acknowledge that this research study was supported by a Scholarship award from German Academic Exchange Service (DAAD).

8. REFERENCES

[1] Alonso, G. & Casati, F., 2005. Web services and service-oriented architectures. Data Engineering, 2005. ICDE 2005.

[2] Ardito, C., Costabile, M.F., Angeli, A., & Lanzilotti, R. 2006. Systematic Evaluation of e-Learning Systems : an Experimental Validation. Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, pp.195–202.

[3] Bichelmeyer, B., 2005. Best Practices in Adult Education and E-Learning: Leverage Points for Quality and Impact of CLE. Val. UL Rev., 40(2), pp.509–520.

[4] Billington, D., 2000. Seven Characteristics of Highly Effective Adult learning Programs. The Adult Learner in Higher Education and the Workplace, pp.1–4.

[5] Cercone, K., 2008. Characteristics of Adult Learners with Implications for Online Learning Design. AACE Journal, 16, pp.137–159.

[6] Holton, D.L., 2010. Using Moodle to teach constructivist learning design skills to adult learners. In T. Kidd & J. Keengwe, eds. Adult Learning in the Digital Age: Perspectives on Online Technologies and Outcomes. Information Science Reference, pp. 40–51.

[7] Holton, D.L., 2010. Using Moodle to Teach Constructivist Learning Design Skills to Adult Learners. , pp.1–12.

[8] Ismail, I., Gunasegaran, T. & Idrus, R.M., 2010. Does E-learning Portal Add Value to Adult Learners ? , 2(5), pp.276–281.

[9] Johnson, D., 2010. Teaching with author's blogs: Connection, Collaboration & Creativity. Journal of Adolescent & Adult Literacy, 53(3), pp.172–180.

[10] Jun, J., 2005. Understanding dropout of adult learners in e-learning. , (1). Available at: https://getd.libs.uga.edu/pdfs/jun_jusung_200505_phd.pdf [Accessed August 27, 2014].

[11] Jun, J., 2004. Understanding the factors of adult learners dropping out of e-learning courses. The 45th annual Adult Education Research Conference, pp.279–284.

[12] Kim, K. J., 2009. Motivational challenges of adult learners in self-directed e-learning. Journal of Interactive Learning Research, 20(3), pp.317–335.

[13] Knowles, M., 1996. The ASTD training & development handbook: A guide to human resource development 4th ed. R. Craig, ed., New York: McGraw-Hill.

[14] Knowles, M., 1980. The modern practice of adult education, New York: Association Press, & Cambridge Book Publishers.

[15] Leal, J. & Queirós, R., 2011. Using the Learning Tools Interoperability Framework for LMS Integration in Service Oriented Architectures. Technology Enhanced Learning.

[16] Luthria, H. & Rabhi, F., 2009. Service Oriented Computing in Practice: An Agenda for Research into the Factors Influencing the Organizational Adoption of Service Oriented Architectures. Journal of theoretical and applied electronic commerce research, 4(1), pp.39–56.

[17] MacLennan, E. & Van Belle, J.P., 2014. Factors affecting the organizational adoption of service-oriented architecture (SOA). Information Systems and e-Business Management, 12, pp.71–100.

[18] Papazoglou, M.P. et al., 2008. Service-Oriented Computing: A Research Roadmap. International Journal of Cooperative Information Systems, 17, pp.223–255.

[19] Rodrigues, A.A., 2012. Empowering Adult Learners through Blogging with iPads and iPods.

[20] Eduardo M. D. Marques and Paulo N. M. Sampaio, 2012. "NSDL: An Integration Framework for the Network Modeling and Simulation," International Journal of Modeling and Optimization vol. 2, no. 3, pp. 304-308

[21] Open Knowledge Initiative (O.K.I.) Website. http://www.okiproject.org.

[22] IMS Digital Repositories Interoperability - Core Functions Information Model Revision: 13 January 2003

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 369-373, 2017, ISSN:-2319–8656

Geographical Analysis of General Land Use Change in Latur and Nilanga Tahsil (1993-96 to 2010-13)

O.M. Jadhav
Assistant Professor
Shivneri Mahavidyalaya,
Shirur Anatpal,MH, India.

Abstract: Men fulfill their basic need from land. They have being used land for cultivation and settlement. Hence, the relationship of people with land is as old as man. The analysis of land use change is essentially to the analysis of changing relationship between people and land. Dynamic movement of people can be detected through this study. Day by day population pressure is increased on land. Hence proper use of land is very important for sustainable development of any region. Every change in land use should be noticed for development planning and management. In this study Technical Committee on Co-ordination of Agricultural Statistics (T.C.C.A.S.) recommendations of standard classification of land class has been considered. Latur and Nilanga Tahsil of Marathwada region are chosen as study area. The regional variations in spatial pattern of land uses are examined from 1993-96 to 2010-2013. Categorywise more change has been observed in land put to non agricultural use and net sown area. Negative change has observed in land put to non agricultural use, barren lands, permanent pastures and other grazing land, area under miscellaneous tree crops etc., cultural waste land, current fallows and other fallow land. Positive change has observed net sown and area under forest. It is good sign for study region development.

Keywords: Lund Use Study, Land Use Change Study

1. INTRODUCTION

Land use study is the base of any regional planning. Use of land defines relationship between human activity and land. Land use is not new in geography. Land use means the surface utilization of all developed and vacant land on a specific point at a given time and space. This change may be due to two most probable reasons. Firstly, the requirements of the society may be the cause for bringing change in the land use. Secondly, the technological impact also promotes changes so that an individual as well as the society is able to maximize the advantages. The demand for new uses of land may be stimulated by a technological change or by a change in size, compositions and requirements of a concerning community. Some changes are short lived while others represent a more constant demand (J.N. Jakson 1963). The study of land use is of pivotal importance in the point of view of planning and development of an area. "It is natural that different types of living which are represented by social values and certain industrial controls will create different patterns of land use within the limits imposed by different agro-physical controls" (Jasbir Singh, 1974)

2 STUDY AREA

Study region is part of Latur district. Latur district is included ten tahsils. This study area consist current Latur tahsil and area of Nilanga tahsil before 23 June 1999. These are important tahsils of Latur district. Latur tahsil is divided into following five revenue circles. These are Kasarkheda, Latur, Gategaon, Tandulja and Murud. Nilanga tahsil is divided into following eight revenue circles. These are Nilanga, Shirur Anantpal, Hisamabad, Ambulga, Kssarshirshi, Kasar Balkunda, Madansuri and Aurad Shahajani. Latur tahsil is located in the north western part of Latur district. Nilanga tahsil is located in the southern part of latur district. Study area North side is bounded by Renapur and Chakur tahsil. East side is bounded by Udgir and Deoni tahsil. South and West side is bounded by Ausa tahsil and Osmanabad district. Study area lies between 17^0 52′ north to 18^0 32′ north latitudes and 76^0 12′ east to 76^0 41′ east longitudes. The total area of study is 2577.35 sq. km.

The height of study region is in-between 510 to 700 meters from sea level. The main river is the Manjra flowing in the northern and eastern part of study area. Other important rivers are the Terna and Tawarja. Both rivers flow west to east direction through the study region. Study region is covered by deep black soil and medium black soil. The average normal rainfall of study region is 714 millimeters. There is lot of variation in temporal and spatial distribution of rainfall in study area.

3 OBJECTIVES

1) To examine the general land use pattern under nine categories of land use.
2) To identify change in general land use pattern from 1993 to 2013.

4 METHODOLOGIES

In India as per the Technical Committee on Co-ordination of Agricultural Statistics (T.C.C.A.S.) recommendations of standard classification, the total geographical area of study area is divided into nine land use categories. The following are the main land use categories. 1) Area under forest 2) Areas under non agricultural uses 3) Barren lands 4) Permanent pastures and other grazing land 5) Area under miscellaneous tree crops etc 6) Cultural waste land 7) Current fallows 8) Other fallow land 9) Net sown area. Area under all land use categories has been collected circlewise form Socio- economic review and district statistical abstract of Latur district, District census & hand book, Gazetteer, Agricultural epitomes, season and crops reports published by the department of agriculture. Volume of Change is calculated using in following formula.

Volume of Change = *(Last Year Land use Area in %) - (Base Year Land use Area in %)*

5 RESULT AND DISCUSSIONS:

5.1 Circlewise changes in general land use from 1993-1996 to 2010-13: There are differences in physical factors in different circles of study region. The general land use is shown by Map No. 1 and 2 . Table No. 1 indicates that the general land use of different circles of this region from 1993-1996 to 2010-13.

5.1.1 Transformations in Area under forest:

The area under forest was 921 hectares in 1993-96 and it was 1702 hectares in 2010-13 in study region. The area under forest was increased by 0.30 % in study region. Positive changes were recorded in Latur (0.26%), Kasarkheda (0.30%), Tandulja (0.19), Nilanga (0.11), Ambulga (0.20), Kasarshirsi (0.57), Kasarbalkunda (0.14), Madansuri (0.10) circles and nigativ changes were recorded in Aurad shajani 1.72) Murud (0.13), Gategoan (0.48). The highest positive change has been observed in Kasarshirsi circle (0.57%) and the lowest positive change was recorded in Aurad shajani (1.72) circle during the period of 1993-96 to 2010-13.

5.1.2 Transformations in Land Put to Non Agricultural Uses: In this land use category negative transformations have been observed Latur (0.25%), Kasarkheda (0.52%), Murud (0.13%), Gategon (1.7%), Tandulja (1.55%), Nilanga (0.98%), Shirur Anatpal (0.63%), Hismabad (0.82%), Ambulga (3.15%), Kasarshirsi (1.29%), Kasarbalkunda (6.39%), Madansuri (3.57%) and Aurad shajani (1.77%) circles during the period of 1993-96 to 2010-13. The highest nigative change was recorded in Kasarbalkunda (6.39%) circle. Very significant change is observed in this category in study area. The area under non agricultural uses was 4.95 % in 1993-96 and it was 2.95 % in 2010-13. Overall 2 % negative change has found in this category.

5.1.3 Transformation in Barren and Uncultivable Land:

The overall 1.19 % negative change of this category is recorded during the period under study time. In this land use category negative transformations have been observed Latur (0.18%), Kasarkheda (0.73%), Murud (1.04%), Gategon (0.3%), Tandulja (1.0%), Nilanga (1.65 %), Shirur Anatpal (0.2%), Hismabad (1.32%), Ambulga (2.07%), Kasarshirsi (1.34%), Kasarbalkunda (4.36%), Madansuri (1.41%) and Aurad shajani (0.6 %) circles during the period of 1993-96 to 2010-13. The highest negative change in Barren and uncultivable land was recorded in kasarkheda (4.36 %) circle and the lowest negative change was noticed in Latur (0.18%) circle during the period under study.

5.1.4 Transformation in Cultivable Waste Land:

The negative change has been recorded in area under cultivable waste land during the period under study. The highest negative change in under cultivable waste land was recorded in Kasarbalkunda (3.3 %) circle and the lowest negative change was noticed in Shirur Anatpal (0.38%) circle during the period under study. The 1.16% negative change has been noticed in study region during the period 1993-96 to 2010-13. The negative change in cultivable waste land was observed in Latur (0.57%), Kasarkheda (0.89%), Murud (1.1%), Gategon (0.85%), Tandulja (1.41%), Nilanga (1.08 %), Shirur Anatpal (0.38%), Hismabad (2.19%), Ambulga (1.06%), Kasarshirsi (1.52%), Kasarbalkunda (3.3%), Madansuri (0.83%) and Aurad shajani (1.07 %) circles.

5.1.5 Transformation in Permanent Pastures and Other Grazing Land:

Negative changes have been observed in the area of this category. Nearly 1.12% was decreased in study area during the period 1993-96 to 2010-13. The highest negative change in area under permanent pastures and other grazing land was recorded in Kasarbalkunda (1.93%) circle and the lowest negative change area under permanent pastures and other grazing land was observed in Latur (0.23%) circle during the period under study. The negative change in permanent pastures and other grazing land was observed in Latur (0.23%), Kasarkheda (0.49%), Murud (0.38%), Gategon (0.69%), Tandulja (1.89%), Nilanga (1.37%), Shirur Anatpal (1.25%), Hismabad (1.42%), Ambulga (0.66%), Kasarshirsi (0.95%), Kasarbalkunda (1.93%), Madansuri (1.47%) and Aurad shajani (0.34 %) circles.

5.1.6 Transformation in Land Under Miscellaneous Tree Crops and Groves Not Included in Net Sown Area: Area under this category negative change was observed in all the circles of study area during the period 1993-96 to 2010-13. The area under this category was decreased by 0.86 % during the period 1993-96 to 2010-13.

The highest negative change was recorded in Tandulja (1.9%) circle and the lowest negative change was observed in Latur and Madansuri (0.23%) circle during the period 1993-96 to 2010-13. The negative change in this category in Latur (0.23%), Kasarkheda (0.76%), Murud (0.93%), Gategon (0.26%), Tandulja (1.9%), Nilanga (1.61%), Shirur Anatpal (0.75%), Hismabad (1.02%), permanent pastures and other grazing land was observed in

Latur (0.23%), Kasarkheda (0.49%), Murud (0.38%), Gategon (0.69%), Tandulja (1.89%), Nilanga (1.37%), Shirur Anatpal (1.25%), Hismabad (1.42%), Ambulga (0.66%), Kasarshirsi (0.95%), Kasarbalkunda (1.93%), Madansuri (1.47%) and Aurad shajani (0.34 %) circles.

5.1.7 Transformations in Current Fallows:

Due to uneven monsoon rainfall, small size of holding, low per hectare yield of agriculture, the marginal land holding farmers put their land as a current fallow land. The average study area negative transformation of this category land is 0.72%. The highest negative change in current fallows are observed in Shirur Anatpal (1.68%) circle and the lowest negative transformations took place in Gategon (0.09%) circle during the

The negative change in this category in Latur (0.11%), Kasarkheda (0.73%), Murud (0.74%), Gategon (0.09%), Tandulja (1.37%), Nilanga (-0.57%), Shirur Anatpal (1.68%), Hismabad (0.72%), Ambulga (0.88%), Kasarshirsi (0.9%), Kasarbalkunda (0.95%), Madansuri (0.68%) and Aurad shajani (0.2%) circles. Due to population pressure increased in study area the area under current fallows has been decreased.

5.1.8 Transformations in Other Fallows:

Nearly 0.68% area under other fallow land is also decreased in Latur and Nilanga tahsil. The highest negative change in area under other fallows has been recorded in Kasarbalkunda (0.93%) circle and the lowest negative change was observed in Latur (0.17%) circle during the period under study. The negative change in area under other fallows was recorded in Latur (0.17%), Kasarkheda (0.69%), Murud (0.61%), Gategon (0.55%), Tandulja (1.31%), Nilanga (0.68%), Shirur Anatpal (0.57%), Hismabad (0.24%), Ambulga (1.23%), Kasarshirsi (0.74%), Kasarbalkunda (0.93%), Madansuri (0.64%) and Aurad shajani (0.31%) circles.

5.1.9 Transformations in Net Sown Area:

Due to utilization of fallow land the net sown area has been increased in study area. Only positive change was recorded in the area of net sown during the period under study. The positive change in net sown area was recorded by

5.2 Overall Volume of Change in General Land from 1993-1996 to 2010-13:

After the consideration of all the land use categories, it is necessary to measure the overall volume of change of nine general land use from 1993-1996 to 2010-13. In overall volume of change includes the land actually involved in the transform from one category to the other category. If the index o

Table 1

Land Utilization of Different Circles of St

ame Of Circles	Year	al Geographical Area in Hectares)	a Under Forest	nd Put To Non ricultural Use	& Uncultivable Land

Source: Computed by Author

dynamic conditions exist there. Table no. 2 indicates that an index of different circles of the study region. The index of volume of change was more than 19% were observed for Kasar Balkunda (19.26%) means the dynamic conditions of land use

Kasar Shirsi	1993-96	14907	65	622	367	321	325	122	211	159	12715
	%	100	0.44	4.17	2.46	2.15	2.18	0.82	1.42	1.07	85.30
	2010-13	14907	151	355	167	94	184	98	77	49	13732
	%	100	1.01	2.38	1.12	0.63	1.23	0.66	0.52	0.33	92.18
	Vol. of ch.	00	0.57	-1.29	-1.34	-1.52	-0.95	-0.16	-0.9	-0.74	6.88
Kasar Balkunda	1993-96	21166	201	2078	1227	998	537	432	315	400	14978
	%	100	0.95	9.82	5.80	4.72	2.53	2.04	1.49	1.89	70.76
	2010-13	21166	230	727	304	300	128	135	114	204	19024
	%	100	1.09	3.43	1.44	1.42	0.60	0.64	0.54	0.96	89.88
	Vol. of ch.	00	0.14	-6.39	-4.36	-3.3	-1.93	-1.4	-0.95	-0.93	19.12
Madansuri	1993-96	15170	20	877	398	295	435	221	186	190	12548
	%	100	0.13	5.78	2.62	1.94	2.87	1.46	1.23	1.25	82.72
	2010-13	15170	35	335	184	168	212	186	84	93	13873
	%	100	0.23	2.21	1.21	1.11	1.40	1.23	0.55	0.61	91.45
	Vol. of ch.	00	0.10	-3.57	-1.41	-0.83	-1.47	-0.23	-0.68	-0.64	8.13
Auradsha	1993-96	16214	271	418	242	258	124	128	230	198	14345
	%	100	1.67	2.58	1.49	1.59	0.76	0.79	1.42	1.22	88.47
	2010-13	16214	89	132	144	85	68	87	198	148	15263
	%	100	0.55	0.81	0.89	0.52	0.42	0.54	1.22	0.19	93.97
	Vol. of ch.	00	-1.72	-1.77	-0.6	-1.07	-0.34	-025	-0.2	-0.31	5.5
Total Latur & Nilanga	1993-96	257735	921	12751	5960	5778	4974	3764	3092	2818	217800
	%	100	0.36	4.95	2.31	2.24	1.93	1.46	1.20	1.09	84.51
	2010-13	257735	1702	76.09	2879	2796	2082	1552	1245	1069	237409
	%	100	0.66	2.95	1.12	1.08	0.81	0.60	0.48	0.41	92.11
	Vol. of ch.	00	0.3	-2	-1.19	-1.16	-1.12	-0.86	-0.72	-0.68	7.2

is observed in this circle. A semi-dynamic land use condition has been recorded in Tandulja (9.45%), Ambulga (9.59), Murud (7.09%), Nilanga (7.44%), Shirur Anatpal (5.44%), Hismabad (6.73), Kasarshirsi (6.88) and Aurad Shajani (5.5%) and madansuri (8.83) circles. The static conditions din land has been observed in Latur (1.79%), Kasarkheda (4.72%) and Gategaon (4.90%) circles. (Table No. 2 and Map No. 3)

Categorywise more change has observed in land put to non agricultural use and net sown area. Negative change has observed in land put to non agricultural use, barren lands, permanent pastures and other grazing land, area under miscellaneous tree crops etc. , cultural waste land, current fallows and other fallow land positive change has observed Net sown and area under forest. It is good sign for study area development. (Figure 1)

6. ACKNOWLEDGMENTS

I thanks to my guide principal Dr. B. G. Sonwane for valuable guidance for this research work. Also I would like to express thank to cartographer Dr. Parmeshwar Poul.

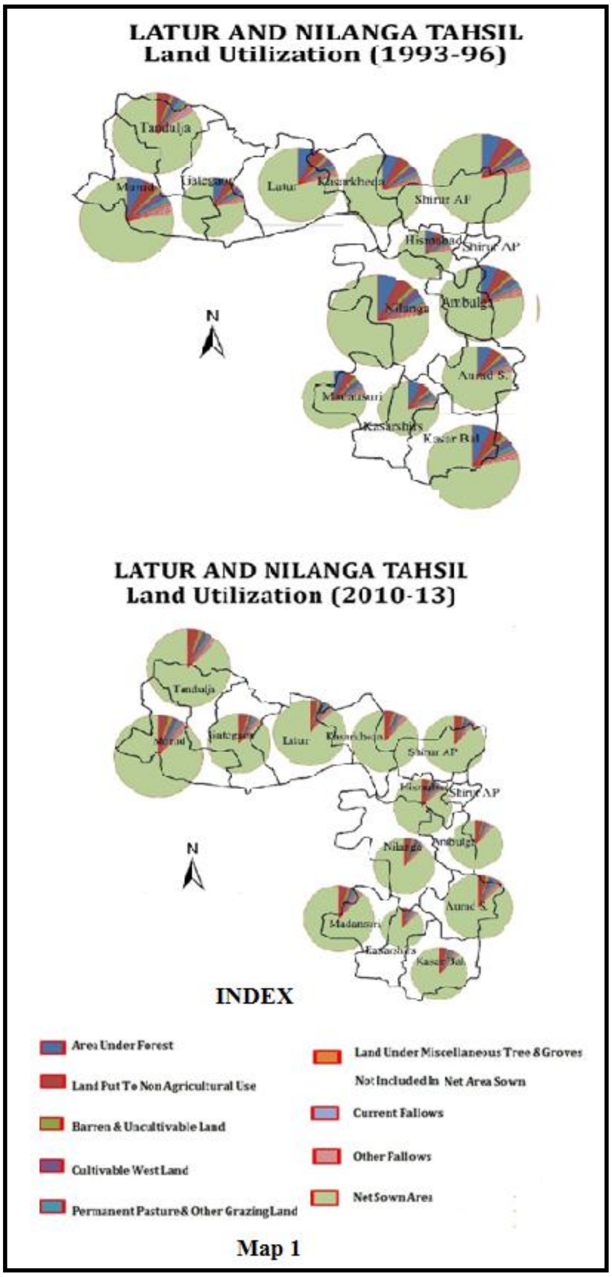
Table 2

Volume of change in General Land Use

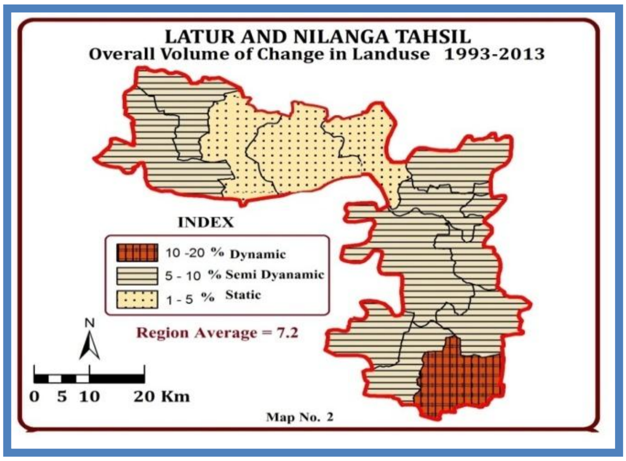
(1993-1996 to 2010-13)

Sr. No.	Name of The Circle	Volume of Change Index in %
1	Latur	1.79
2	Kasarkheda	4.72
3	Murud	7.09
4	Gategoan	4.90
5	Tandulja	9.45
6	Nilanga	7.44
7	Shirur Anantpal	5.44
8	Hisamabad	6.73
9	Ambulga	9.59
10	Kasarshirsi	6.88
11	Kasar Balkunda	19.26
12	Madansuri	8.83
13	Aurad shajani	5.5
14	Total Latur & Nilanga	7.2

Source: Computed by Author



Map 1



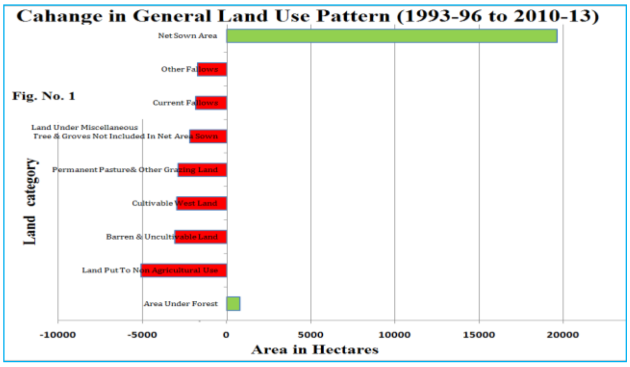


Fig. No. 1

7. REFERENCES:

[1] Das, M. M. (1981): “Land Use Pattern in Assam”, Geographical Review of India, Vol.43, No.3, pub by Calcutta, Geographical society of India, 1981, 243-244.

[2] Jasbir Sing & S.S. Dhillon (1984): “Agricultural Geography”, Tata Mc Graw Hill Publishing comp. ltd; New Delhi 1984, 112-113.

[3] Jainendra Kumar (1935) : Land use. Analysis Intra India publication New Delhi. 65

[4] Majid Husain (1979): “Agricultural Geography” Intra-India publication, Delhi 1979,114 - 155

[5] Socio – Economic review of Latur District (1993-2013): Govt. of Maharashtra.

[6] Statistical review reports (1993-2013): Panchyat Samiti of Latur and Nilanga

[7] Socio-Economic reviews & district stastical abstract, Latur district, Latur (1993-96 to 2010-13):

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 374-377, 2017, ISSN:-2319–8656

Fuzzy traditional EPQ model allows backorders with planned shortages by yager's ranking method

B.Bharani
Assistant Professor,
Dept Of Mathematics,
Guru Nanak College,Chennai

Dr. A. Praveen Prakash
HOD, Dept Of Mathematics,
Hindustan University,
Kelambakkam, Chennai

Abstract: The nature of this paper reports of an investigation of a set of continuous time ,constant demand inventory models under the condition of yield uncertainty. We consider a classical EPQ model, that models are lot sizing models whose objective is the determination of fuzzy optimal production order and shipment quantity.In this paper, we discuss the inventory problem with fuzzy backorder. Yager's ranking method for fuzzy numbers is utilized to find the inventory policy in terms of the fuzzy total cost. Finally a numerical example is given to illustrate the model.

Keywords: Fuzzy number, inventory, backorder, yager's method

1. INTRODUCTION

Within the context of traditional inventory models, the pattern of demands is either deterministic or uncertain. In practice, the latter corresponds more to the real-world environment. To solve these inventory problems with uncertain demands, the classical inventory models usually describe the demands as certain probability distributions and then solve them. However, some times, demands may be fuzzy, and more suitably described by linguistic term rather than probability distributions. If the traditional inventory theories can be extended to fuzzy senses, the traditional inventory models would have wider applications. Usually, inventory systems are characterized by several parameters such as cost coefficients, demands etc. Accordingly, most of the inventory problems under fuzzy environment can be addressed by fuzzying these parameters. For instance, Park [8] discussed the EOQ model with fuzzy cost coefficients. Ishii and Konno [3], Petrovic et. al. [9], and Kao and Hsu [4] investigated the Newsboy inventory model with fuzzy cost coefficients and demands respectively. Roy and Maiti [10] developed a fuzzy EOQ model with a constraint of fuzzy storage capacity. Chang [1] construct a fuzzy EOQ model with fuzzy defective rate and fuzzy demand. Yao and Chiang [13] compare the EOQ model with fuzzy demand and fuzzy holding cost in different solution methods. Kao and Hsu [5] find the lot size-reorder point model with fuzzy demand. Besides, there is another kind of studies which fuzzes the decision variables of inventory models. For example: Yao and Lee [15] developed the EOQ model with fuzzy ordering quantities; Chang and Yao [2] investigated the EOQ model with fuzzy order point; Wen-Kai K. Hsu and Jun-Wen Chen [11] studied Fuzzy EOQ model with stock out. Madhu & Deepa [7] developed an EOQ model for deteriorating items having exponential declining rate of demand under inflation & shortage. Kun-Jen Chung, Leopoldo Eduardo Cárdenas-Barrón [6] compare the complete solution procedure for the EOQ and EPQ inventory models with linear and fixed backorder costs. Recently W. Ritha etal. [14] fuzzified EOQ Model with one time discount offer allowing back.

In this paper, we discuss the inventory problem with fuzzy back order. The decision variables are the ordering quantity Q and the back order quantity S. The approach of this paper is to find the optimal order quantity Q* with the minimum cost determined from Yager's ranking method.

2.PRELIMINARIES:

Definition :

A fuzzy set $\hat{A}$ is defined by

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)); x \in X, \mu_{\tilde{A}} \in [0,1]\}.$$

In the pair $(x, \mu_{\tilde{A}}(x))$, the first element x belong to the classical set A, the second element $\mu_{\tilde{A}}(x)$ belong to the interval $[0,1]$, called membership function or grade membership. The membership function is also a degree of compatibility or a degree of truth of x in $\tilde{A}$.

Definition : α**-cut**

An α-cut of a fuzzy set $\hat{A}$ is a crisp set A_{α},that contains all the elements of universal set X having a membership grade in A greater than (or) equal to the specific value of α.

i.e., $A_{\alpha} = \{x \in X \mid \mu_{\tilde{A}}(x) \geq \alpha\}$

Generalised Fuzzy number

Any fuzzy subset of the real line R, whose membership function satisfies the following conditions is a generalized fuzzy number.

(i) $\mu_{\tilde{A}}(x)$ is a continuous mapping from R to the closed interval $[0,1]$.
(ii) $\mu_{\tilde{A}}(x)$=0, $-\propto \leq x \leq a,$
(iii) $\mu_{\tilde{A}}(x) = L_{(x)}$ is strictly increasing on $[a_1, a_2]$
(iv) $\mu_{\tilde{A}}(x) = 1, a_1 \leq x \leq a_2.$
(v) $\mu_{\tilde{A}}(x)$ = R(x) is strictly decreasing on $[a_1, a_2]$
(vi) $\mu_{\tilde{A}}(x) = 0, , a_4 \leq x \leq \propto.$

Where a_1, a_2, a_3 and a_4 are real numbers.

3. YAGER'S RANKING METHOD:

If the α-cut of any fuzzy number $\tilde{A}$ is $\frac{1}{2}\int_0^1 [A_L(\alpha) + A_U(\alpha)]\ d\alpha$.

4. MODEL DEVELOPMENT

Notations Used:

D - Annual Demand in units

K - Ordering cost per order

c_h - Holing cost per unit

c_b - backordering cost per unit per year

Q - lot size per order

S - size of the back order quantity

Q* - optimal value of Q

Mathematical Model

Consider the traditional Economic Order Quantity (EOQ) model that allows backordering with the following total annual cost function

$$C_{Trad}(Q,S) = \frac{D}{Q}k + \frac{(Q-S)^2}{2Q}C_h + \frac{S^2}{2Q}C_b \qquad (1)$$

The objective is to find the optimal order quantity which minimize the total cost

The necessary conditions for minimum

$$\frac{\partial C_{Trad}(Q,S)}{\partial Q} = -\frac{DK}{Q^2} + \frac{C_h}{2} - \frac{S^2C_h}{2Q^2} + \frac{S^2C_b}{2Q^2} = 0$$

Differentiate (1) partially w.r.to S, we obtain

$$\frac{\partial C_{Trad}(Q,S)}{\partial S} = \frac{SC_h}{Q} + C_h + \frac{SC_b}{Q} = 0$$

$$S = \frac{C_hQ}{C_h + C_b} \qquad (2)$$

Substitute S in (1), Hence the optimal order quantity is

$$Q^* = \sqrt{\frac{2DK(C_h + C_b)}{C_hC_b}}$$

The EOQ model with back order and fuzzy demands

Let$\tilde{D}, \tilde{K}, \widetilde{C_b}, \widetilde{C_h}$ be the trapezoidal numbers and they are defined as follows i.e, they are described by the α-cuts.

$$K(\alpha_K) = [L_K^{-1}(\alpha_K), R_K^{-1}(\alpha_K)]$$

$$D(\alpha_D) = [L_D^{-1}(\alpha_D), R_D^{-1}(\alpha_D)]$$

$$C_h(\alpha_{C_h}) = [L_{C_h}^{-1}(\alpha_{C_h}), R_{C_h}^{-1}(\alpha_{C_h})]$$

$$C_b(\alpha_{C_b}) = [L_{C_b}^{-1}(\alpha_{C_b}), R_{C_b}^{-1}(\alpha_{C_b})]$$

Now

$C_{Trad}(Q,S) = \frac{D}{Q}k + \frac{(Q-S)^2}{2Q}C_h + \frac{S^2}{2Q}C_b$ can be rewritten as

$$\tilde{C}_{Trad}(Q,S) = \frac{\tilde{D}}{Q}\tilde{k} + \frac{(Q-S)^2}{2Q}\widetilde{C_h} + \frac{S^2}{2Q}C_b$$

Yager's ranking index can be derived as

$$\tilde{C}_{Trad}(Q,S) = \frac{K_1(\alpha_D, \alpha_K)}{Q} + \frac{(Q-S)^2}{2Q}K_2(\alpha_{C_h}) + \frac{S^2}{2Q}K_3(\alpha_{C_b})$$

Where,

$$K_1(\alpha_D, \alpha_K) = \frac{1}{4}\left\{\int_0^1 L_D^{-1}(\alpha_D)d\alpha_D.\int_0^1 L_K^{-1}(\alpha_K)d\alpha_K + \int_0^1 R_D^{-1}(\alpha_D)d\alpha_D.\int_0^1 R_K^{-1}(\alpha_K)d\alpha_K\right\}$$

$$K_2(\alpha_{C_h}) = \frac{1}{2}\left\{\int_0^1 L_{C_h}^{-1}(\alpha_{C_h})d\alpha_{C_h} + \int_0^1 R_{C_h}^{-1}(\alpha_{C_h})d\alpha_{C_h}\right\}$$

$$K_3(\alpha_{C_b}) = \frac{1}{2}\left\{\int_0^1 L_{C_b}^{-1}(\alpha_{C_b})d\alpha_{C_b} + \int_0^1 R_{C_b}^{-1}(\alpha_{C_b})d\alpha_{C_b}\right\}$$

Hence the optimal order quantity is

$$Q^* = \frac{2K_1(K_2+K_3)}{K_2K_3}$$

NUMERICAL VALIDATION:

To validate the proposed model, consider the data, $D = 1200, K = 100, C_h = 25, C_b = 50$ and hence by graded mean method

$\tilde{D}$ = (1000,1100,1300,1400)

$\tilde{K}$ = (98,99,101,102)

$\tilde{C_h}$ = (15,20,30,35)

$\tilde{C_b}$ = (40,45,55,60)

Thus

$D(\alpha_D)$= (1000+ $100\alpha, 1400 - 100\alpha)$

$K(\alpha_K) = (98 + 5\alpha, 102 - 5\alpha)$

$C_h(\alpha_{C_h}) = (15 + 5\alpha, 35 - 5\alpha)$

$C_b(\alpha_{C_b}) = (40 + 5\alpha, 60 - 5\alpha)$

$$K_1(\alpha_D, \alpha_K) = \frac{1}{4}\left\{\int_0^1 (1000 + 100\alpha)d\alpha . \int_0^1 (98 + 5\alpha)d\alpha + \int_0^1 (1400 - 100\alpha)d\alpha . \int_0^1 (102 - 5\alpha)d\alpha\right\}$$

$K_1(\alpha_D, \alpha_K)$ = 60112.5

$K_2(\alpha_{C_h}) = \frac{1}{2}\left\{\int_0^1 (15 + 5\alpha)d\alpha + \int_0^1 (35 - 5\alpha)\, d\alpha\right\}$

$K_2(\alpha_h)$ = 25

$K_3(\alpha_{C_b}) = \frac{1}{2}\left\{\int_0^1 (40 + 5\alpha)d\alpha + \int_0^1 (60 - 5\alpha)\, d\alpha\right\}$

$K_3(\alpha_{C_b}) = 50$

Optimal Order quantity $Q = \sqrt{\frac{2DK(C_h+C_b)}{C_hC_b}}$=120 units

Optimal total cost $C_{Trad}(Q, S) = 2000$

Fuzzy Optimal Order quantity $Q^* = \sqrt{\frac{2K_1(K_2+K_3)}{K_2K_3}}$ = 85 units

Fuzzy Optimal total cost $C_{Trad}(Q^*, S^*)$ = 1475.588

CONCLUSION:

The purpose of this paper is to study the inventory models under fuzzy environment. This fuzzy model assists in determining the optimal expected total cost per cycle amidst the existing fluctuations. In this paper the Yager's ranking method of optimization is employed.

REFERENCES:

[1] Chang, H. C. "An application of fuzzy theory to the EOQ model imperfect quality items," Computers and Operations Research, 31, 2079-2092 (2004).

[2] Chang, S. C. and J. S. Yao, "Economic reorder point for fuzzy backorder quantity," European Journal of Operational Research, 109,183-202 (1998).

[3] Ishii, H and T. Konno, "A stochastic inventory problem with shortage cost," European Journal of Operational Research, 106, 90-94(1998).

[4] Kao. C. and W.K. Hsu, "A Single-period inventory model with fuzzy demand," Computer and Mathematic with Application,43,841-848(2002).

[5] Kao. C. and W.K. Hsu, "Lot size-reorder point inventory model with fuzzy demand," Computer and Mathematic with Application, 43, 1291-1302(2002).

[6] Kun-Jen Chung, Leopoldo Eduardo Cárdenas-Barrón, "The complete solution procedure for the EOQ and EPQ inventory models with linear and fixed backorder costs" Mathematical and Computer Modeling, 55, 2151-2156 (2012).

[7] Madhu Jain and Deepa Chauhan "Inventory Model with Deterioration, Inflation and permissible delay in payment." Aryabhatta J. of Maths & Info. Vol. 2 (2) pp 164-174 [2010]

[8] Park, K. S, "Fuzzy-set theoretic interpretation of economic order quantity," IEEE Trans. System, Man, Cybernetics, SMC-17, 1082-1084(1987).

[9] Petrovic D., R. Petrovic and M. Vujosevic,"Fuzzy model for the newsboy problem," International Journal of Production Economics,45, 435-441(1996).

[10] Roy, T. K., and M. Maiti. "A fuzzy EOQ model with demand-dependent unit cost under limited storage capacity," European Journal of Operational Research, 99, 425-432(1997).

[11] Wen-Kai K. Hsu and Jun-Wen Chen, "Fuzzy EOQ model with stock out," Department of Shipping Transportation & Management, National Kaohsiung Marine University, Taiwan, R.O.C.

[12] Yager, R.R., "A procedure for ordering fuzzy subsets of the unit interval", Information Sciences, 24, 143-161(1981).

[13] Yao, J. S. and J. Chiang, "Inventory without backorder with fuzzy total cost and fuzzy storing cost defuzzified by centroid and signed distance, "European Journal of Operational Research, 148,401-09(2003).

[14] W. Ritha and M. Sumanthi " Fuzzy EOQ Model with one time discount offer and Allowed Back order ." Aryabhatta J. of Mathematics & Informatics vol. 4 (1) pp. 13-22 [2012].

[15] Yao, J. S. and H. M. Lee, "Fuzzy inventory with or without backorder for fuzzy order quantity with trapezoid fuzzy number," Fuzzy Sets and Systems, 105, 311-337 (1999).

Supply Chain Enhancement Using Improved Chaid Algorithm for Classifying the Customer Groups

C.P.Balasubramaniam,
Ph.D Research Scholar ,
Karpagam Academy Of Higher Education,
Karpagam University,
Coimbatore,Tamilnadu
India

Dr.R.Gunasundari
Associate Professor,
Department of Information Technology,
Karpagam Academy Of Higher Education,
karpagam university
Coimbatore,Tamilnadu
India

Abstract: Lot of Companies is involving in the activities of the Supply Chain that is will used for different functions like company process management and relation between the suppliers and customers agents. The decision-tree based approach is used to make learn and recognize the logical methods of a tree structure. A state-of-the-art supply chain management [8] gives the coding rules as well as the Logical rules features needed by the system. Each attribute is classified and tested during the layout analysis and Logical features are collected and compared to the company's synthetic data set. The Agent-Based Modeling method is employed in the study is the Improved Chi -squared Automatic Interaction Detection" (I-CHAID method). All the user needs are tested around 50 dataset attribute contents belonging to the SCM with I-CHAID method; this will be representing the lower error rate for determining the logical labels are less than 5%. And also the efficiency like precision, recall, error rate will calculate and explains in detail functioning of I-CHAID with respect to the company's supply chain management.

Keywords: Decision-Tree Based Approach; Agent Based Modeling, Improved Chi-Squared Automatic Interaction Detection, Error Rate.

1. INTRODUCTION

Supply Chain Management is mainly based on the two core ideas. The first idea is that practically every product that reaches an end user represents the cumulative effort of multiple organizations. These organizations are mainly focused to collectively as the supply chain.

The second idea is that while the supply chains have existed for a long time, most of the organizations have only paid attention to what was happening within their "four walls." Very less businesses method, much less managed, the last movement chain of activities that ultimately delivered the products to the final customer. The result is not proper method and ineffective supply chain method.

Supply chain management, is an active management of supply chain activities to enlarge the customer value and achieve a sustainable competitive advantage. It represents a conscious effort by the supply chain firms to develop and run the supply chains in the most effective & efficient ways possible. Supply chain activities manage everything from the product development, sourcing, production, and logistics at the same time information of systems needed to coordinate these activities.

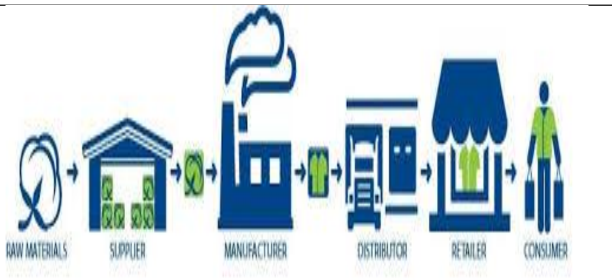


Figure1.Supply Chain Structure

The organizations that make up the supply chain are "linked" together through the physical flows and information flows. Physical flows develop the transformation, storage of goods, movement, and materials. These are the most visible piece of supply chain. Information deals allow the various method of supply chain partners to coordinate their long-term plans, and to control the day-to-day flow of goods and material up and down in the supply chain

Supply Chain Management (SCM) is the oversight of materials, information, and finances as they move in a process from the supplier, wholesaler, manufacturer, retailer to consumer.

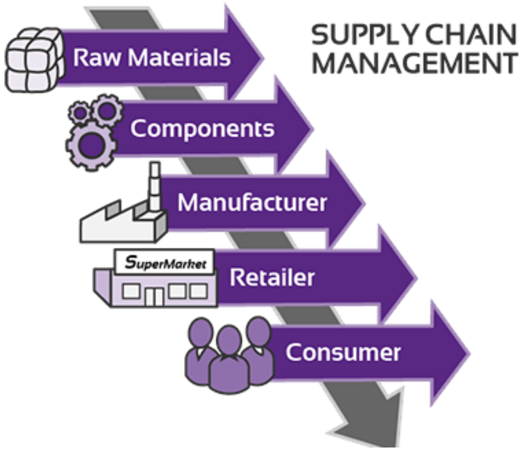


Figure2.Supply Chain Process

Supply chain management develops coordinating and integrating the flows on both within and among companies. The main goal of supply chain management system is to reduce the inventory. As a solution for the successful supply chain management, sophisticated software systems with the Web interfaces are competing with the Web-based Application Service Providers (ASP) who promises to provide a part or all of the SCM service for companies who rent their service.

2. LITERATURE REVIEW

Bolstorff and Rosenbaum (2003), understanding the SCM depends on the motivation and interest of those involved with this concept. A technology provider may associate the SCM with software, like Enterprise Resource Planning and Advanced Planning and Scheduling systems, third-party logistics providers align SCM with the distribution practices. Consulting companies may align it with their intellectual property, for example, along in the same line, it is possible to say that the same phenomenon is seen in the academia.

Halldorsson et al.(2007) discuss that SCM is unified theory and there is no one replace this theory. Depending on each situation, one can choose a theory as the dominant explanatory theory, and then complement it with one or several of the other theoretical perspectives". To establish a work of reference that allows us to mitigate the gap between the present SCM research and Practice and the theoretical explanations and to understand the SCM in practice such as economic perspective; socio-economic perspective; and strategic perspective. Also explaining the two research questions is something mentioned below like, how to structure a supply chain and how to achieve a goal of certain structure of supply chain. Finally their proposed system gives the final framework for looking at two different problem areas within the SCM such as third-party logistics and new product development.

Davis (1989) and Leblanc (1992) algorithms are based on the definition of a within-node homogeneity measure, unlike Segal's algorithm which tried to maximize between-node separation. In addition to that the decision tree is classified as the survival trees and regression tree. Survival tree [12] based analysis is a powerful non-parametric method of clustering the survival data [13] for prognostication to determine the importance and effect of various covariates.

Su and Fan (2004) extended the CART formally known as the classification tree and regression trees was designed by Brieman et al. (1984) for addressing the tree size selection and other issues related to the formation of the related attribute tree structure.

CART algorithm described as the three different procedural categories like growing a large tree, pruning the sequence of nested sub-trees, and finally selecting a best-sized tree. In addition to that there are two close approximations are available in the formation of analysis of correlated failure times when forming the CART tree. Marginal method is the marginal distribution approach of correlated failure times is formulated for classification method.

The approach is the infirmity model to the regression setting. The CART algorithm to multivariate survival [11] data by introducing a gamma distributed frailty to account for the dependence among the survival times [13] based on the likelihood ratio test as the splitting function.

Gordon and Olshen (1985) gave the first adaption of CART algorithm and it concise the recursive partitioning scheme of tree-structured [2] for classification, regression and probability class estimation are adapted to cover the censored survival analysis [3].

The assumptions only required are those which guarantee the identify ability of the conditional distributions of lifetime covariates. Thus, the techniques are applicable to more general situations than the famous semi-parametric model of Cox. This subject is used to the censorship of data. Wasserstein metrics is measure the distances between Kaplan-Meier [1958] curves and certain point masses.

A variable selection approach is useful to inference method, the Generalized Likelihood Ratio (GLR) test, is employed to address the hypothesis testing problems for existing the models.

Allen Zaklad et al. discussed that the Sustainable Supply chain improvement is the business process development, enabling technology, and social system transformation. They presented a model of supply chain intervention that will enable you to address the hidden side of supply chain operations in the context of business processes and technology.

Osman.A.H and Naomie developed the Semantic Plagiarism Detection Scheme mainly focused on Chi-squared Automatic Interaction Detection. It discussed based on the semantic text plagiarism detection technique based on the Chi-squared Automatic Interaction Detection.

It was very useful the analyses and compares the each text in Semantic Plagiarism Detection Scheme and each argument generated by the CHAID algorithm scheme [17] in order to select the important arguments were the another features are added of the method.

3. AGENT DECISION NETWORK FOR SELECTING THE SUPPLIER

Agent decision network is selecting the particular supplier that maximizes the usage of associated with the supply chain. This system has been developed by the agents, and improve need to buy two different types of materials to manufacture the product.

Each material is associated with a list of possible supplier agents with it. The reputation node and the offer node are the maintained on the choice of the supplier where the offer node is modeled by a deterministic node characterize the cost of material offer proposed by the supplier.

This is determined by comparing the received offer to the market price and/or evaluating the quality of product and characteristics of the offered products. The supplier agent influences the commercial transaction method to acquire the specific material, represented by the node Transaction Status. The reputation and the offer most jointly determine the utility associated with the transaction between the customer agent and the selected supplier agent.

According to the probability distribution of reputation, the value of the expected utility is associated with the offers. To determine the utility of the entire sub-chain, the transaction utility nodes for all the materials are used together.

The Transaction probabilistic nodes consist of influence the probabilistic node Supply chain, raw materials, instead, which expresses the probability that a supply chain can be successfully developed.

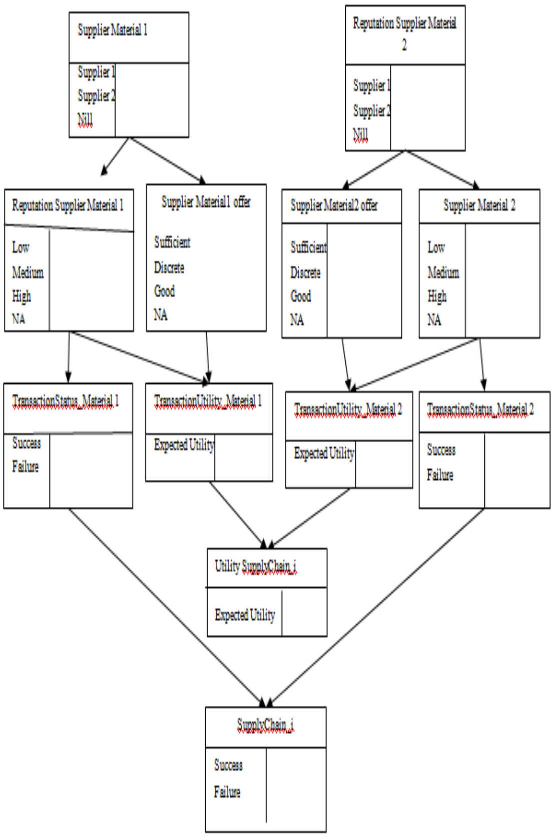


Figure 3.Agent Decision Network for Selecting the Supplier

4. IMPROVED CHAID ALGORITHM

Chi -squared Automatic Interaction Detection is the process of test, the mining data from the decision tree structured framework. The new skeleton are proposed in the CHAID named as Improved CHAID which has two processing schemas which are provided for the CHAID like insertion of logical rules and knowledge to support the decision tree.

In the Improved CHAID , there are two setups have been tested such as, each logical element have a own decision tree structured as well as single decision tree structured [2] for all the logical elements. Both trees are recognizing by the Improved CHAID with respect to the kernel of the system. The Improved CHAID are the trust with confidence on a decision tree.

The Improved CHAID algorithm is mainly focused on the data process whereby two, three or more tree elements are distinguished by the decision tree structure. Decision tree is one common method used in the data mining [16] to extract the predicted information. Morgan and Sonquist uses the regression trees in terms of decision tree for prediction and explanation process with the help of AID (Automatic Interaction Detection).

Generally Chi-squared Automatic Interaction Detection is discrimination and classification methods developed, based on the same representation paradigm by trees. Many methods are intended to increase the probability of solving some problem proposed in the past few era. Specially to improve the quality of the Quinlan's system, leading to the famous C4.5 method. This mobility emerged the concept of lattice graphs which was popularized by the induction graphs of the SIPINA method.

Rokotamalalaetal explains the concept of the decision tree construction for classification and discrimination problems based on the SIPINA software. The main problem is to predict the output value of an object from a set of variables, discrete or continuous. This prediction process is the problem solving in order to develop and finding a partitioning of the individuals elements on the decision tree.

The objective is to improve individual groups, the most homogeneous as possible from the point of view of the variable to be predicted. The main idea is to represent the empirical distribution of the attribute to be predicted by each and every node of the decision tree. Thus, the tree build favors the more discriminating attributes. Here, the difficulty is to choose among N attributes characterizing the structure elements that made it possible to have the best discrimination rate.

Depending upon the prediction of variable candidate and the variable, the characterizing has the two different conditions of things, generally the segmentation and statistical criteria available in the supply chain management. Both are used to entropy of Shannon and its alternatives. The segmentation define a contingency table crossing the variable to be predicted and the descriptor candidate.

All those actions happened for the process comprehension. Consider the notations to describe the numbers resulting from the crossing of the attribute class with V modalities and a descriptor with U methods.

Table 1. Number table during the crossing of two variables

$Q = P$	p_1	p_u	p_U	$\sum_{u=1}^{U} p$
q_1	w_{11}	w_{1u}	w_{1U}	w_u
q_2	w_{21}	w_{2u}	w_{2U}	
q_v	w_{v1}	w_{vu}	w_{vU}	
q_{V-1}	$w_{((V-1),u)}$	$w_{((V-1),u)}$	$w_{((V-1),U)}$	
q_V	w_{V1}	w_{Vu}	w_{VU}	
$\sum_{v=1}^{V} q$	w_v			w

To calculate the relevance of a variable in the segmentation, CHAID develops the independence deviation χ^2 defined by the following equation.

$$\chi^2 = \sum_{v=1}^{V} \sum_{u=1}^{U} \frac{\left(w_{vu} - \frac{w_v \times w_u}{w}\right)^2}{\frac{w_v \times w_u}{w}}$$

Equation (1)

The values of $\chi^2$are not bounded, they are in the range$[0, +\infty]$. The main disadvantage is the high emphasis of the descriptors having a high number of modalities. To decrease this negative impact, it is much more manageable to normalize by the number of freedom degrees. The formula T of T$_{\text{schuprow}}$ has values now in a range [0; 1]. This new concept of equation gives the Improved CHAID algorithm.

$$T_S = \frac{\chi^2}{\sqrt[w]{(V-1) \times (U-1)}}$$

Equation (2)

There are three lemma were used to find the significance of the decision tree which is based on the method of crossing of two variables.

Lemma 1:

If splitting of χ^2_{split} and $T_{S_{split}}$ which is given as the two different conditions for performing the decision tree

Condition 1:χ^2>T_S decision tree were not performed.

Condition 2: if the threshold is increased of T_S to the χ^2>T_S automatically the shortest decision tree was performed.

Lemma 2:

If merging of χ^2_{merge} and $T_{S_{merge}}$ which is given as the two different conditions for performing the decision tree as large and compact one. Before merging the different variable, it is need to verify the merging variable with the class distribution whether it has a statistical significance of threshold value and threshold level. This assumption gives the two conditions.

*Condition 1:*from lemma 1, when T_S values are decreased with the respect to the statistical significance the largest decision tree was formed.

Condition 2: simultaneously T_S values are increased with the respect to the statistical significance; the compact decision tree was formed.

Lemma 3:

Let us consider the level of process as "φ" and their independent significance tests as "τ" it gives the significance probability of ρ that will be got no significant differences in all these tests, which is described as following described condition.

the product of the individual probability $(1-\varphi)^{\tau}$, from this considerations of test, the study apply this in the following example consideration like level of 0.03, and τ is 10 however, it is got the individual probabilities as 0.74 from this the study now has a 26% chance that one of these 10 tests will turn of out & significant, despite each individual test performance only being at the 3% level. In order to guarantee that the overall significance test is still at the "φ" level, we have to adapt the significance level α′ of the individual test. This following results are relation between the overall and the individual significance level were compared with the following comparison results

$$(1-\varphi')^{\tau} = (1-\varphi)$$

Equation (3)

This equation can easily be solved for φ':

$$\varphi' = 1-(1-\varphi)^{\frac{1}{\tau}}$$

Equation (4)

Which for small φ reduces to:

$$\varphi' = \frac{\varphi}{\tau}$$

Equation (5)

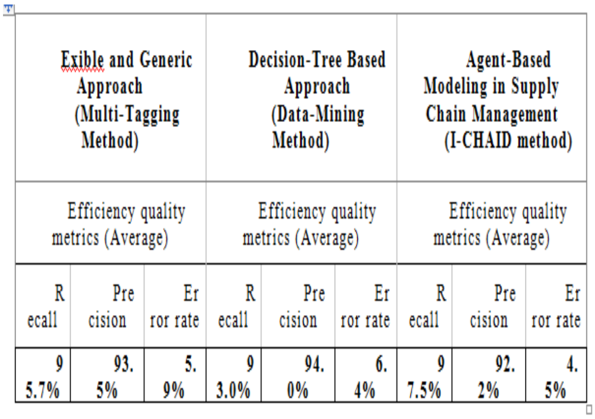
To get an overall significance level φ and perform τ individual tests, simply obtain the significance level for the individual tests by φ'.

5. EXPERIMENTAL RESULTS

The proposed experimental approaches are around five data set attributes were listed named as the Supplier Agent, Production Manager Agent, Dealer Agent, Client Agent, Inventory Agent that containing the 1000 data set and that is splinted in to two block each. In this era two approaches [19] are compared using the measures: recall, precision, and error rate with the proposed system.

The following table represents the three different approach are compared and shows the quality metrics like the recall, precision, insertion rate and error rate.

Table 1: Average Efficiency Quality Metrics



Exible and Generic Approach (Multi-Tagging Method)			Decision-Tree Based Approach (Data-Mining Method)			Agent-Based Modeling in Supply Chain Management (I-CHAID method)		
Efficiency quality metrics (Average)			Efficiency quality metrics (Average)			Efficiency quality metrics (Average)		
Recall	Precision	Error rate	Recall	Precision	Error rate	Recall	Precision	Error rate
95.7%	93.5%	5.9%	93.0%	94.0%	6.4%	97.5%	92.2%	4.5%

The Agent-Based Modeling in Supply Chain Management (I-CHAID method) is more efficient to obtain better results Supply Chain Management. However the results of both methods are close to the proposed system but give the efficient results.

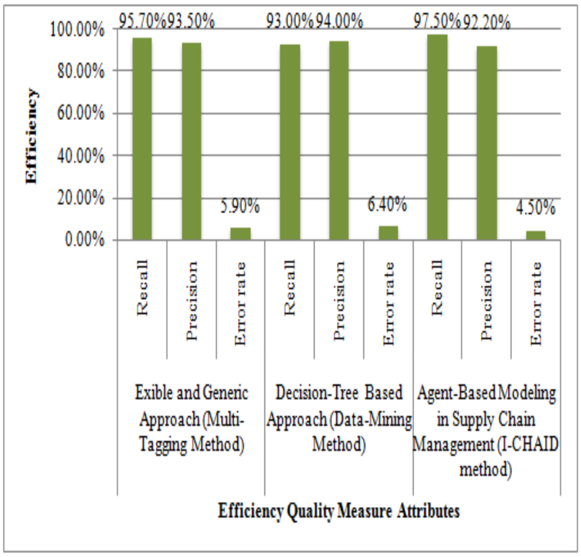


Figure 4. Efficiency Quality Measure

The Agent-Based Modeling in Supply Chain Management (I-CHAID method) is more accuracy to obtain better results in Supply Chain Management. However the results of both methods are close to the proposed system but the proposer system gives the accuracy.

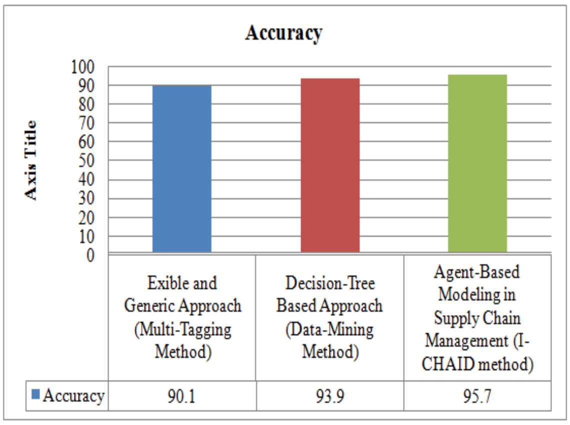


Figure 5. Various SCM Accuracy Comparisons

5. CONCLUSION

To determine the nature and performance of the proposed work for SCM with presence of CHAID improved and Bounded in extent of the Synthetic dataset is used in terms of Supplier, Production Manager, Dealer, Client and Inventory attribute set. Which contains the Client city, Client Name, Client Phone number, Client State, Cost, Dealer Name, Dealer id, Item Name, Product Name, Product Price, Product Quantity, Product id, Product name, Supplier Name, Supply Quantity. The above Synthetic dataset are managed and classified by the proposed improved CHAID framework. The obtained results of accuracy and its performance measures of recall, precision and error rate are measured and compared with the existing technique for the input supply chain [10] management dataset. The classification of this algorithm is viewed in tree structure where the decision tree is classified. Figure 4 gives the accuracy comparison for the proposed framework and existing technique like Multi-Tagging Method and *Data*-Mining Method [16].The accuracy here is measured by sensitivity and specificity. From the figure 3, the Efficiency quality metrics obtained by the proposed technique Improved CHAID for SCM in recall is 97.5% and precision is 92.2 % and error rate 4.50% which is better when compared with the existing technique of Exible and Generic Approach and Decision-Tree Based Approach. The comparison graph for Efficiency quality measure attributes are calculated by sensitivity and specificity is shown in the figure 3. From the figure 4 proves that the accuracy of proposed improved CHAID with SCM gives maximum while comparing with the existing technique for SCM.

6. REFERENCES

1. Gordon and Olshen, "Tree-structured survival analysis", Cancer Treatment Reports, Journal Article, Research Support, U.S. Gov't, Non-P.H.S., Research Support, U.S. Gov't, P.H.S. [1985, 69(10):1065-1069].

2. Xiaogang Su, Juanjuan Fan "Multivariate Survival Trees: A Maximum Likelihood ApproachBased on Frailty Models",Biometrics journal of the international Biometrics society, Volume 60, Issue 1, pages 93–99, March 2004.

3. A. Bela, T. Moinel, Y. Rangoni "Improved CHAID Algorithm for Document Structure Modeling"

4. Camarinha-Matos, LM. andAfsarmanesh, H, Collaborative networked organizations: a research agenda for emerging business models. Massachusetts, Kluwer Academic, 2004.

5. Lee, J.-H., and Kim C.-O. (2008). "Multi-agent systems applications in manufacturing systems and supply chain management: a review paper." International Journal of Production Research 46(1): 233-265.

6. Moyaux, T., Chaib-Draa, B., and D'Amours, S. (2007). "Information sharing as a coordination mechanism for reducing the bullwhip effect in a supply chain." Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews 37(3): 396-409.

7. Frayret, J.-M., D'Amours, and Montreuil, B. (2004a). "Coordination and control in distributed and agent-based manufacturing systems." Production Planning & Control 15(1): 1-13.

8. Forget, P., D'Amours, S., Frayret, J.-M., and Gaudreault, J. (2008b). Design of multi-behavior agents for supply chain planning: an application to the lumber industry. Supply Chains: Theory and Application. V. Kordic, I-TECH Education and Publishing: 551-568.

9. Bolstorff, R. R., and Rosenbaum, R. (2003). Supply chain excellence. New York, AMACOM.

10. Halldorsson, A., Kotzab, H., Mikkola, J.H., and Skjott-Larsen, T. (2007). "Complementary theories to supply chain management." Supply Chain Management: An International Journal 12(4): 284-296.

11. Davis, R. and Anderson, J. (1989): Exponential survival trees, Statistics in Medicine 8, pp 947-962.

12. Lebalanc, M.; Crowlry, L. (1992): Relative risk trees for censored survival data, Biometrics. v48. 411-425.

13. Su, X. G.; Fan, J. J. (2004): Multivariate survival trees: a maximum likelihood approach based on frailty models, Biometrics 60, pp. 93-99.

14. Kaplan, E.L.; Meier, Paul. (1958): Nonparametric estimation from incomplete observations, J. Am. Stat. Assoc. 53, 457-481.

15. Kass G. (1980), "an exploratory technique for investigating large quantities of categorical data, Applied Statistics" 29(2), 119-127, 1980.

16. Anita Prinzie, Dirk Van den Poel, "WITHDRAWN: Constrained optimization of data-mining problems to improve model performance: A direct-marketing application", http://www.researchgate.net/publication/222555006.

17. Gilbert Ritschard "CHAID and Earlier Supervised Tree Methods", Publications récentes du Départementd'économétrie, http//www.unige. ch/ses/metri /cahiers.

18. HongwenZheng , Yanxia Zhang "Feature selection for high dimensional data in astronomy" Accepted for publication in Advances of Space Research, arXiv:0709.0138v1 [astro-ph] 3 Sep 2007

19. Jose Manuel Serra, Laurent A. Baumes, "Zeolite synthesis modelling with support vector machines: A combinatorial approach", publication at: http:// www.researchgate.net /publication/6537094

20. C.P. Balasubramaniam, V. Thigarasu "Agent-based Modeling in Supply Chain Management using Improved C4-5", Research Journal of Applied Sciences, Engineering and Technology 9(2): 91-97, 2015, Maxwell Scientific Organization, 2015.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 384-390, 2017, ISSN:-2319–8656

Comparative Analysis of Hybrid K-Mean Algorithms on Data Clustering

Navreet Kaur
Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Shruti Aggarwal
Department of Computer Science and Engineering
Sri Guru Granth Sahib World University
Fatehgarh Sahib, India

Abstract: Data clustering is a process of organizing data into certain groups such that the objects in the one cluster are highly similar but dissimilar to the data objects in other clusters. K-means algorithm is one of the popular algorithms used for clustering but k-means algorithm have limitations like it is sensitive to noise ,outliers and also it does not provides global optimum results. To overcome its limitations various hybrid k-means optimization algorithms are presented till now. In hybrid k-means algorithms the optimization techniques are combined with k-means algorithm to get global optimum results. The paper analyses various hybrid k-means algorithms i.e. Firefly, Bat with k-means algorithm, ABCGA etc. The Comparative analysis is performed using different data sets obtained from UCI machine learning repository. The performance of these hybrid k-mean algorithms is compared on the basis of output parameters like CPU time, purity etc. The result of Comparison shows that which k-means hybrid algorithm is better in obtaining cluster with less CPU time and also with high accuracy.

Keywords: Data mining, Clustering, Hybrid K-means Algorithm, ABCGA, CPU time

1. INTRODUCTION

Data mining [1] is a powerful concept for data analysis and defined as the discovery of hidden pattern from data sets. It is also defined as the Mining of knowledge from huge amount of data. Data Mining was developed to make useful discoveries from the data independently, without depending on the statistics. It is an important subfield of the computer science. The goal of the data mining process is to discover the interesting patterns from the data sets and then transforming them into a structure which is understandable for further use. Data mining is the analysis step of the knowledge discovery in databases process or KDD. Data mining is the extraction of patterns and knowledge from large amounts of data, not the extraction of data itself. It is easily applied to any form of large-scale data or information processing as well as any application of computer decision support system, including artificial intelligence, machine learning and business intelligence.

1.1 Clustering

Clustering [2] is the process of divide the population or data points into different groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. Clustering can be considered the most important unsupervised learning problem. A loose definition of clustering could be "task of organizing objects into groups whose members are similar in some way". Clustering can be said as identification of similar classes of objects. The various types of clustering methods [3] are given below:

1.1.1 Partitioning Methods

The most fundamental version of cluster analysis is partitioning, which organizes the object of dataset into groups or clusters. Some commonly used Partitioning Based clustering techniques are k-means, k-medoids.

1.1.2 Hierarchical methods

Hierarchical based clustering method group the data objects into a hierarchy or tree of clusters. Representing data objects in the form of a hierarchy is useful for data summarization and visualization. These methods are used to find spherical-shaped clusters. For example: Divisive and alggormative.

1.1.3 Density based methods

To find the clusters of arbitrary shape, this technique can model clusters as dense regions in the data space separated by sparse regions this is basic technique for density based clustering. For Example: DBSCAN (Density –based clustering based on connected regions with high Density)

1.1.4 Grid-based methods

The above clustering techniques are data-driven but the grid base clustering method takes space-driven approach by partitioning the embedding space into cells independent of distribution of input objects. For Example: STING (Statistical Information Grid).

1.2 K-means clustering

K-means clustering [4] is popular for cluster analysis in data mining. K-means clustering aims to partition m data elements into k clusters in which each data element belongs to the cluster with the minimum distance between them and which results in a partitioning of the data elements into clusters.

K-means clustering is a type of unsupervised learning when you have unlabeled data. The algorithm works iteratively to assign each data element into one of K groups based on the features that are provided. Data elements are clustered based on feature similarity. The results of the K-means clustering algorithm are:

1. The centroids of the K clusters, they are used to label new data
2. The Labels for the training data (each data point is assigned to a single cluster)

1.2.1 Steps for k-means algorithm

Input: Number of clusters to be formed, k and a database Y= {y1, y2, y3 ...yn} containing n data elements.
Output: A set of k clusters

Method:

1. *The numbers of clusters k to be formed are chosen.*
2. ***T**he centroids are selected randomly.*
3. *The distance between each data element and cluster centroid is calculated.*
4. *The data element is assigned to the cluster centroid where distance between cluster centroid and data element is minimum than other cluster centroids.*
5. *Calculate the new cluster centroid of the data element for each cluster and update the cluster centroid.*
6. *Repeat from third step if data element was reassigned otherwise stop.*

In k-means algorithm user need to specify the number of cluster to be formed in advance and also k-means algorithm converges to local minima rather than a global optimum result. Due to these limitations various hybrid k-means optimization clustering algorithms are designed.

2. LITERATURE SURVEY

In literature survey the various hybrid k-mean algorithms are discussed like KFFA, KABC, K-Krill herd, KACO, PSO-ACO-K etc.

2.1 Hybrid K-mean Algorithms using Swarm Based Optimization Techniques

Hybrid k-means algorithms are those algorithms that combines k-means algorithm with optimization algorithms. The hybrid k-means algorithms are used to reduce the limitations of k-means algorithm and these are given below.

In KFFA algorithm [5] the k-means clustering algorithm is optimized using firefly algorithm [6]. To find the centroids for specified number of clusters the firefly is used and then to refine the centroids and clusters k-means is applied. In KACO algorithm [7] firstly ACO is applied because cluster quality is based on it. PSO-ACO-K algorithm [8] combines the particle swarm optimization, ant colony optimization and k-means. In KCUCKOO algorithm [9] Cuckoo search leads too much iteration because it randomly selects initial centroids and to overcome this problem they are selected using k-means. In KBAT algorithm [10] the optimization BAT algorithm helps to reduce the local optimal problem of k-means clustering algorithm. In KABC algorithm [11] the k-means is combined with artificial bee colony algorithm for optimization and clusters formed are better than k-means algorithm. In K-Krill herd algorithm [12] the krill herd is used to initialize the centroids for clusters in k-means. Krill herd is used to provide local optimal results. BAT k-medoids [13] clustering algorithm is combined with BAT algorithm to solve the optimization problems of k-medoids algorithm. In KPSO algorithm [14] the results of PSO algorithm is used as the initial seed of the k-means algorithm and k-means algorithm will be applied for refining. The Tabu-KHM algorithm [15] combines the optimization property of tabu search and the local search capability of k-harmonic means algorithm.

2.2 Hybrid K-means algorithms using Bio-inspired Optimization Techniques

In KGA [16] genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. In KFP algorithm [17] the flower pollination algorithm is used to reduce the disadvantages of k-means local optima and its results are used to select the centroids of clusters in k-means. In KFSS [18] the bio inspired fish school search optimization algorithm is used along with k-means algorithm. K-Means and K-Harmonic with Fish School Search Algorithm [19] provides more optimized results than KFSS. IGSA-KHM algorithm [20] not only helps the KHM clustering escape from local optima but also overcomes the slow convergence speed of the IGSA. CGA [21] is clustering based Genetic Algorithm with polygamy selection and dynamic population control technique. According to CGA the fitness values obtained from chromosomes in each generation were clustered into two non-overlapping clusters. ABCGA means Adaptive Biogeography Clustering based genetic algorithm. In hybrid technique the CGA which means Clustering based Genetic Algorithm is used along with ABPPO which stands for Adaptive biogeography based predator-prey optimization. In the proposed technique the clustering process is similar to k-means algorithm.

3. COMPARATIVE ANALYSIS

In Comparative analysis the various hybrid k-means algorithms are compared based on some output parameters and these comparisons are described below.

3.1 Data set

The data sets used are data sets from the UCI Machine Learning Repository are Wine, Iris, Seed, Breast Cancer and Liver Disorders data set. Also number of attributes and number of instances in these data sets are described. These data sets are shown in Table 1.

Table 1. Data Sets

S.No	Name	Number of instances	Number of Attributes
1	Wine	179	14
2	Iris	150	4
3	Seed	210	5
4	Breast Cancer	699	10
5	Liver Disorders	345	7

This table shows that the total number of attributes in Wine data set is 14 and number of instances is 179, total number of attributes in Iris data sets is 4 and total number of instances is 150. The total number of attributes in Seed data set is 5 and total number of instances is 210.The total number of attributes in breast cancer data set is 10 and total number of instances is 699. The total number of attributes in Liver disorders data set is 7 and total number of instances is 345.

3.2 Output Parameters

The output parameters are those parameters based upon which the performance of existing clustering algorithm is compared with the new hybrid k-means optimization algorithm. Some output parameters are described below:

3.2.1 *TP: True Positive*

It is defined as the proportion of positives that are identified correctly. It is also called as Sensitivity. Example: Sick people who are correctly identified as having the condition.

3.2.2 *TN: True Negative*

It is defined as the proportion of negatives that are identified correctly. It is also called as Specificity. Example: Sick people who are correctly identified as not having the condition.

3.2.3 *FP: False Positive*

They are those which are identified incorrectly. Example: Healthy people incorrectly identified as sick.

3.2.4 FN: False Negative

They are those which are incorrectly rejected. Example: Sick people incorrectly identified as healthy.

Some major parameters based upon which the performance of proposed algorithm is evaluated and compared with existing algorithm are described below.

3.2.5 Accuracy

It is defined as only the proportion of true results.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \text{…………….eq. (i)}$$

Here TP = True positive, FP= False positive, TN= True negative, FN= False negative.

3.2.6 *Purity*

Purity is defined as the percent of the total number of objects that were classified correctly. To compute purity each cluster is assigned to the class which is most frequent in a cluster. It describes the cluster quality.

$$\text{Purity} = \frac{1}{N}\sum_k \max j\, |w_k \cap c_j| \text{.......eq. (ii)}$$

Where N = number of objects, k = number of clusters, cj is set of classes and wk is the set of clusters.

3.2.7 CPU Time

CPU time means the time taken required by the computer to perform a given set of computations. If CPU time is less than the clustering algorithm is better than algorithms having more Computation time.

3.3 Comparison of K-means, CGA and ABCGA algorithm using Purity

The comparison of CGA algorithm and k-means algorithm with the ABCGA algorithm using Purity when number of clusters is 8 for five data sets that are Wine, Iris, Seed, Breast cancer and Liver Disorders taken from UCI Machine Learning Repository.

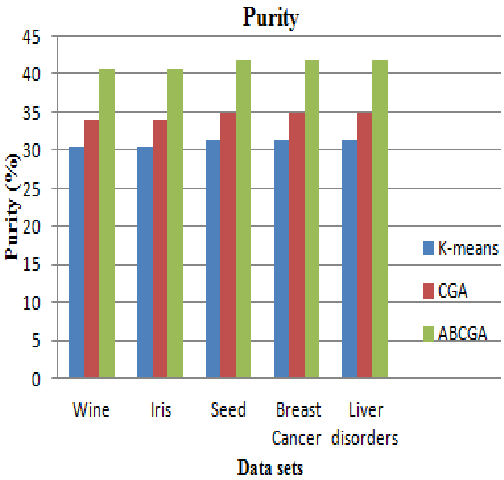


Figure 1: Comparison using Purity

In the above Figure 1 comparison it shows that the ABCGA algorithm purity is high for the all five data sets than CGA and k-means Algorithm for Clustering.

3.4 Comparison of KFFA, KBAT and KFPA algorithm using CPU Time [17]

The comparison of these three KFFA, KBAT and KFPA algorithms are based on the CPU time using two data sets Iris and wine from UCI Machine Learning Repository.

Table 2. Comparison using CPU time

Data set	KFFA	KBAT	KFPA
Iris	8.7	3.2	3.2
Wine	19	3.88	3.87

In the above comparison it shows that the KBAT and KFPA algorithm require less CPU time than KFFA Algorithm for Clustering.

3.5 Comparison of KFSS and KPSO algorithm using Accuracy[19]

The comparison of these two KFSS and KPSO algorithms are based on the Accuracy using two data sets Iris and wine from UCI Machine Learning Repository.

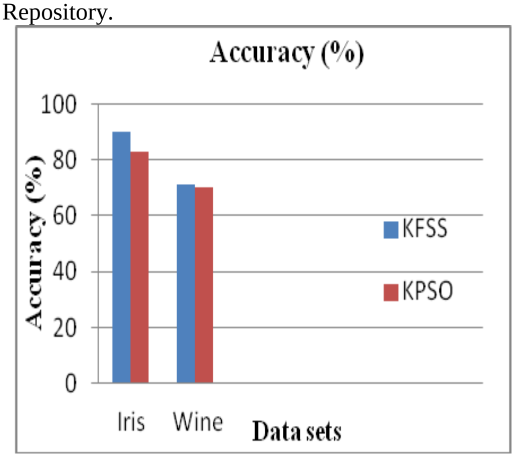


Figure 2: Comparison using Accuracy

In the above comparison it shows that the KFSS algorithm accuracy is high than KPSO Algorithm for Clustering.

The comparative Analysis of various hybrid k-means algorithm is done in the paper using various output parameters. The performance is compared for different data sets that are Iris, Wine, Seed, Breast Cancer and Liver Disorders. The comparison of k-means, ABCGA and CGA is done by using purity output parameter which shows that ABCGA has high purity than other two algorithms for clustering. The KFFA, KBAT, KFPA is compared based on the CPU time whose results shows that KFPA and KBAT requires low CPU time. Another comparison is done using accuracy output parameter for KFSS and KPSO algorithm which shows that the accuracy for KFSS algorithm is high for Iris and wine data sets.

4. CONCLUSION

In this paper, the data clustering, clustering techniques and various hybrid k-mean algorithms are presented. The comparison of the performance of various hybrid k-means optimization algorithms is done. The comparison of CGA and ABCGA algorithm is done through purity which shows that ABCGA algorithm provides better purity than CGA. The KFFA, KBAT and KFPA k-mean hybrid techniques are also compared in this paper. On these the comparative analysis is done on the basis of CPU time and the results show that the KBAT and KFPA requires less CPU time than KFFA. Also the hybrid KFSS and KPSO algorithm are compared based on accuracy and comparison shows that KFSS provides better Accuracy than KPSO. There is scope for improvement in these hybrid k-mean algorithms to handle high dimension data sets.

5. REFERENCES

1. Ming-Syan Chen, Jiawei Han, Ps Yu, “Data Mining: An overview from database perspective” IEEE Transaction on knowledge and data engineering, Vol. 8, Issue 6, pp. 886-883, 1996.

2. Anil .Jan, “Data Clustering: 50 years beyond k-means” Pattern Recognition Letters, Elsevier, Vol. 31, pp. 651-666, 2010.

3. Lior Rokach and Oded Maimon, “Clustering methods” Data mining and Knowledge handbook, pp. 321-352, 2005.

4. J.B. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations” Proceedings of Fifth Berkeley Symposium on Mathematics Statistics and Probability, University of California Press, Vol. 1, pp. 281-297, 1967.

5. S.J Nanda, G. Panda, “A Survey on nature inspired metaheuristic algorithm for partition clustering” Swarm and Evolutionary Computation, Elsevier, Vol. 16, pp. 1-18, 2014.

6. Xin-She Yang, "Firefly algorithms for multimodal optimization" Stochastic Algorithms: Foundations and Applications, SAGA 2009. Lecture Notes in Computer Sciences, Vol.5792, pp.169–178, 2009.

7. K.Aparana and Mydhili K.Nair, “Enhancement of k-means algorithm using ACO as optimization technique on high dimensional data” 2014 international conference on Electronics and Communication Systems (ICECS) IEEE, pp. 1-5, 2014.

8. Taher Niknam and Babak Amiri, “An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis” Applied Soft Computing, Elsevier, vol. 10, pp. 183–197, 2010.

9. Saida Ishak Boushaki, Kamel Nadjet and Omar Bendjeghaba, “ A New Hybrid Algorithm for Document Clustering based on Cuckoo Search and K-means”, Recent advances on Soft Computing and Data Mining SCDM Springer, pp. 59-68, Vol. 287, 2014.

10. Tang Rui, Fong Simon, Yang Xin-She and Deb Sujay, “Integrating nature-inspired optimization algorithms to k-means clustering” 2012 seventh

International Conference on Digital Information Management ICDIM, IEEE, pp. 116-123, 2012.

11. Karaboga K, Dervis D, Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm" Applied Soft Computing, Springer, Vol. 11, pp. 7-652, 2011.

12. Hamed Nikbakht, Hamid Mirvaziri, "A new clustering approach based on K-means and Krill Herd algorithm" 23rd Iranian Conference on Electrical Engineering, IEEE, 2015.

13. Monica Sood and Shilpi Bansal, "K-Medoids Clustering Technique using Bat Algorithm", International Journal of Applied Information Systems, pp. 20-22, 2013.

14. Yucheng Kao, Szu-Yuan Lee, "Combining K-means and particle swarm optimization for dynamic data clustering problems", IEEE International Conference on Intelligent Computing and Intelligent Systems, 2009.

15. Gungor. Z and Unler. A, "k-Harmonic Means Data Clustering with Tabu Search Method" Applied Mathematical Modeling, Vol. 32, pp. 1115-1125, 2008.

16. Md Anisur Rahman, Md Zahidul Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-Means", Knoweldge based systems, Elsevier, 2014.

17. Parul Aggarwal And Shikha Mehta, "Comparative Analysis Of Nature Inspired Algorithm On Data Mining", IEEE International Conference On Research In Computational Intelligence And Communication Networks, 2015.

18. C.J.A. Bastos-Filho, F.B. Lima Neto, A.J.C.C. Lins, A.I.S. Nascimento, M.P. Lima, "A novel search algorithm based on fish school behavior" , in: Proceedings of the 2008 IEEE International Conference on Systems, Man, and Cybernetics, pp. 2646–2651, 2008.

19. Adriane B.S. Serapiao, Guilherme S. Correa, Felipe B. Goncalves, Veronica O. Carvalho, "Combining K-Means and K-Harmonic with Fish School Search Algorithm for data clustering task on graphics processing units" , Applied Soft Computing, Elsevier, Vol. 41, pp. 290-304, 2016.

20. Minghao Yin, Yanmei Hu, Fengqin Yang, Xiangtao Li, Wenxiang Gu, "A novel hybrid K-harmonic means and gravitational search algorithm approach For clustering" Expert Systems with Applications, Elsevier, Vol. 38, pp. 9319-9324, 2011.

21. A.M. Aibinu, H.Bello Salau, Najeeb Arthur Rahman, M.N. Nwohu, C.M. Akachukwu, "A novel Clustering based Genetic Algorithm for route optimization", Engineering Science and Technology an International Journal, Vol. 19, pp. 2022–2034, 2016.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 391-395, 2017, ISSN:-2319–8656

Model View Mapper Architecture for Software Reusability

Chethana S
Lecturer,Dept.of Computer Science
NMKRV PU College
Bangalore, India

Dr.Srinivasan
Professor
RV Engineering College
Bangalore,India

Abstract: Design Pattern Architecture is a serious issue in the development of any complex software system for Small, Medium and Big Organization. The essential problem is how to incorporate rapidly changing technology and new requirements by composing patterns for creating reusable designs. The main objective of this proposed work is to enhance the performance of enterprise design pattern reuse, to monitor software component problems and to predict the best design pattern for reusability. As a solution for all these problems a new framework called MVM pattern approach is proposed for migrating to a new design approach. It is helpful for all kinds of the organization complex software developments.

Keywords: Software Reuse, Architectures, Framework, Design Patterns, MVC, MVM, MVVM, MVP

1. INTRODUCTION

Software Reuse is characterized as the way toward building or collecting software applications and frameworks from the existing software. By reusing Software pattern there are many advantages includes time can be saved, increase productivity and also reduce the cost of new software development. The proposed work aims on developing a framework for software pattern reuse in enterprise level applications. Frameworks give a standard working structure through which client's primary aim is on creating desired modules than creating lower level points of interest. By utilizing this facility the software designers can invest more time in building up the prerequisite of software, instead of setting up the tools of application development. The framework is a set of the reusable software program that structures the basis for an application. Frameworks help the developers to assemble the application rapidly. At its best code reuse is refined through the sharing of regular classes or collection of methods, frameworks, and techniques.

2. LITERATURE SURVEY

In Neha Budhija [1] expert designers have done an empirical study of the software reuse activity with the concept of object-oriented design. The study concentrated on fundamentally three aspects of reuse : (1) the communication between some design forms (2) the mental procedures required in reuse (3) the mental portrayals developed all through the reuse action. In FENIOSKY PENA-MORA [2] introduces an in-advance improvement of a framework for utilizing design rationale and design patterns for creating reusable programming frameworks. The work describes the use of an explicit software creation procedure to catch and disseminate specific knowledge that augments the depiction of the cases in a library during the development process of software applications by heterogeneous gatherings. B.JALENDER [3], the authors described about how the code level reusable components can be built and how the code level components can be designed. It also provides some coding guidelines, standards and best practices used for creating reusable code level components and guidelines and best practices for making configurable and easy to use. Tawfig M [4] the authors have presented the concept of reuse at design level in more details. Also, the work proposes an approach to improve the reusability of software design by using the concept of directed graph. The outcome of the proposed work is to produce a design to be considered as reusable components which can be adapted in many software systems. Erich Gamma[5] proposed design patterns as a new mechanism for expressing object-oriented design experience and they described that the design patterns can be considered reusable micro-architectures that contribute to an overall system architecture. Authors described how to express and organize design patterns and newly introduced a catalog of design patterns. In Reghu Anguswamy [6] provided a generic list of reuse design principles for component based software development which is based on a preliminary analysis of the literature of software reuse and reuse design over the past few decades. Authors suggested that the proposed list is new since the reuse design principles presented in the past were specific to programming languages, domains, or programming paradigms. William B[7] In their paper authors have briefly summarized about software reuse research, discussed major research contributions and unsolved problems in the proposed area, they provided pointers to key publications. Sajjan G [8] in this paper the authors have done an attempt to answer some unsolvable questions. Authors pointing out that it have been more than three decades since the idea of software reuse was proposed. They have done a research on how far are investigators with software reuse research and practice. In CHARLES W[9] have done a survey on various approaches to software reuse found in the research literature. Some taxonomy has been used to describe and compare the different approaches and make generalizations about the field of software reuse. The taxonomy used in the work is used to characterize each reuse approach by its reusable artifacts and the way how are organized. SathishKumar[10] ,the authors have given a brief summarization of present research status in the field of software reuse and major research contributions. Some future directions for research in software reuse are also discussed.

Model view mapper pattern

There are many different approaches existing for software reusability such as MVC(Model View Controller),MVM(Model View Presenter) and MVVM(Model-View-ViewModel) and all these approaches have its own advantages and disadvantages. So a new framework is proposed for Software design reusability called Model View Mapper Pattern.

The Model represents a set of packages instead of business logic classes i.e. business model as well as database access authority along with server side validation i.e. data model. It also defines business rules for data means how the data can be changed and manipulated then updated in to database along with proper validation. The View represents the UI Components like CSS, jQuery, Angular.js, Ajax, html etc. It is also represents as database access authority along with server side as well as client side validation. It is not only responsible for displaying the data that is received from the mapper and from the database as the result. The Mapper is responsible to process the incoming requests. It receives valid input from users via the View, then process the user's data with the help of Model as well as View and passing the results back to the View. Typically, it acts as the coordinator between the View and the Model. Figure[1] describes the flow of how the communication takes place in the proposed MVM pattern architecture.

Code Classifier & Code Analyzer Algorithm is used in the work to classify and count the number of class name, abstract class name, method name, and abstract method name, interfaces used in the program, iteration counts and the number of lines in the file present in the file. **Tracing Behavioral Dependency algorithm** is used to identify the Dependent Class Name & Depending Class Name in separate table.

Algorithm 3: Hybrid ABC-CM➔Artificial Bee Colony + Naïve Bayes Classifier Model

1: Initialize the population of solutions Bee i, j ,i = 1 ...EL, j = 1 ...P

2: Evaluate the population

3: iteration=1

4: repeat

5: Produce new solutions Val i, j for the employed bees by using (2) and evaluate them

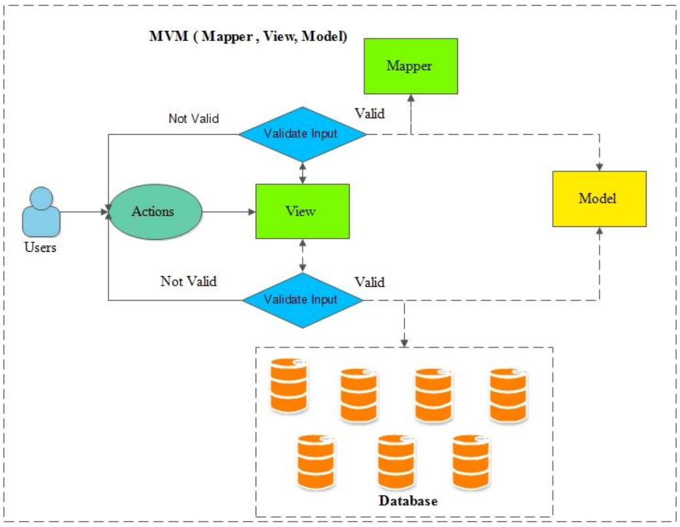
6: Apply the greedy selection process in Bees

7: Calculate the probability values Pop i, j for the solutions Bee i, j by (1)

8: Produce the new solutions Val i, j for the onlookers from the solutions Bee i, j selected depending on Pop i, j and evaluate them

Figure1. Proposed MVM pattern Architecture

9: Apply the greedy selection process



10: Determine the unrestrained solution for the scout, if exists, and replace it with a new randomly produced solution Bee i, j by (3)

11: Memorize the best solution achieved so far.

12. Let **best solution** be a training set of samples, each with their class labels. There are n classes, Cls1, Cls2, . . . ,Clsn.

Figure 2. Sequence Diagram for MVM

Each sample is represented by an n-dimensional vector, X = {x1, x2, . . . ,xn}, depicting n measured values of the n attributes, Attb1, Attb2, . . . , Attbn, respectively.

13. Given a sample X, the classifier will predict that X belongs to the training set taking the highest a posteriori probability, conditioned on X. That is X is predicted to belong to the class Clsi if and only if

$$P(Clsi\ |X) > P(Clsj\ |X) \text{ for } 1 \leq j \leq m,\ j\ != i.$$

Thus we find the class that maximizes P(Clsi |X). The class Clsi for which P(Clsi |X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$$P(Clsi\ |X) = P(X|Clsi)\ P(Clsi)\ P(X)\ .$$

14. As P(X) is the same for all classes, only P(X|Clsi)P(Clsi) need be maximized. If the class a priori probabilities, P(Clsi), are not known, then it is commonly assumed that the classes are equally likely, that is, P(Cls1) = P(Cls2) = . . . = P(Clsk), and we would therefore maximize P(X|Clsi). Otherwise we maximize P(X|Clsi)P(Clsi). Note that the class a priori probabilities may be estimated by P(Clsi) = freq(Clsi , T)/|T|.

15. Given data sets with many attributes, it would be computationally expensive to compute P(X|Clsi). In order to reduce computation in evaluating P(X|Clsi) P(Clsi), the naive assumption of class conditional independence is made. This assumes that the values of the attributes are temporarily independent of one another, given the class label of the trial. Mathematically this means that

$$P(X|Clsi) \approx n\ PI\ k=1\ P(xk|Clsi).$$

The probabilities P(x1|Clsi), P(x2|Clsi), . . . , P(xn|Clsi) can easily be predictable from the training set. Recall that here xk refers to the value of attribute Attbk for sample X.

(a) If Attbk is categorical, then P(xk|Clsi) is the number of samples of class Clsi in T having the value xk for attribute Attbk, divided by freq(Clsi , T), the number of sample of class Clsi in T.

(b) If Attbk is continuous-valued, then we typically assume that the training values have a Gaussian distribution with a mean μ and standard deviation σ defined by

$$g(x,\ \mu,\ \sigma) = 1\ \surd\ 2\pi\sigma \exp - (x - \mu)\ 2\ \ / 2\sigma\ 2\ , \text{ so that}$$

$$p(xk|Clsi) = g(xk,\ \mu Clsi\ ,\ \sigma Clsi\ ).$$

We need to compute μClsi and σClsi , which are the mean and standard deviation of values of attribute Attbk for training samples of class Clsi .

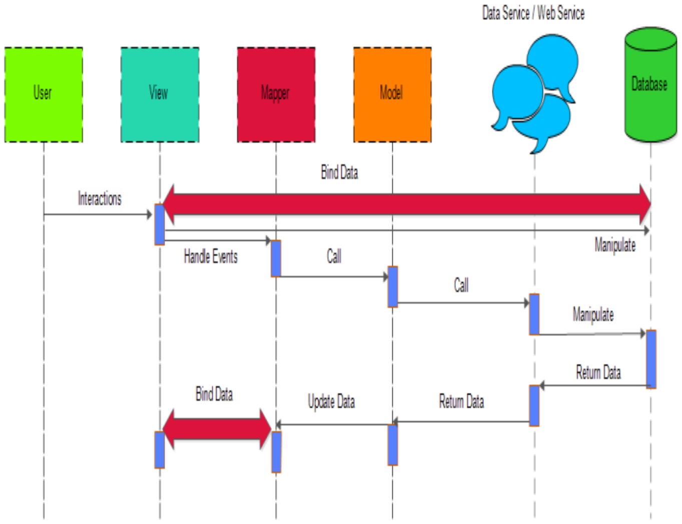


Figure 2.Sequence Diagram of MVM pattern

The sequence diagram in Figure 2shows the difference between the event responses of all patterns.

16. In order to predict the class label of X, P(X|Clsi)P(Clsi) is evaluated for each class Clsi . The classifier predicts that the class label of X is Clsi if and only if it is the class that maximizes P(X|Clsi)P(Clsi).

17: Iteration=Iteration+1

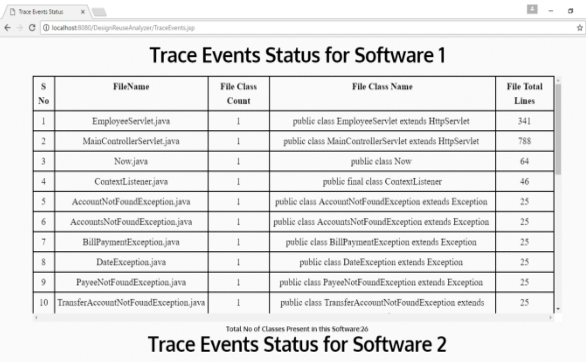
18: until Iteration=MITR

3. RESULTS AND DISCUSSION

3.1. Module 1 Process Flow:

- Upload MVC Software Version 1: MVC Software Version1 have 297 files
- Display Code Analyzer Page for Software version1
- Upload MVC Software Version 2 : MVC Software Version 2 have 299 files.
- Display code Analyzer Page
- Trace events for Software1
- Trace events for Software2

The overall process flow shown in figure 3 to figure 9.Upload both Software1 and Software2.Figure 3 and figure4 shows the Tracing events for both inputs.Figure5 shows the Behavioral dependency table for both software.Figure6 shows the filenames and their reusability percentage by using MVM Pattern. Figure 7 represents the chart which shows the percentage of reusability. Same process continues with proposed MVM pattern and the results are shown in figure 8 and figure9.



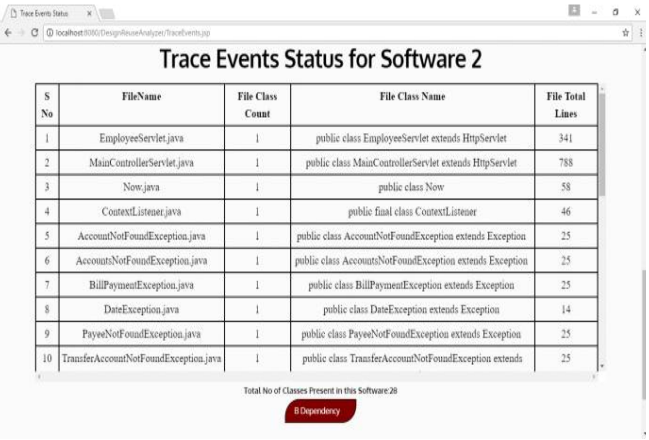
Trace Events Status for Software 1

S No	FileName	File Class Count	File Class Name	File Total Lines
1	EmployeeServlet.java	1	public class EmployeeServlet extends HttpServlet	341
2	MainControllerServlet.java	1	public class MainControllerServlet extends HttpServlet	788
3	Now.java	1	public class Now	64
4	ContextListener.java	1	public final class ContextListener	46
5	AccountNotFoundException.java	1	public class AccountNotFoundException extends Exception	25
6	AccountsNotFoundException.java	1	public class AccountsNotFoundException extends Exception	25
7	BillPaymentException.java	1	public class BillPaymentException extends Exception	25
8	DateException.java	1	public class DateException extends Exception	25
9	PayeeNotFoundException.java	1	public class PayeeNotFoundException extends Exception	25
10	TransferAccountNotFoundException.java	1	public class TransferAccountNotFoundException extends	25

Total No of Classes Present in this Software:26

Trace Events Status for Software 2

Figure 3.Trace events status for Software1



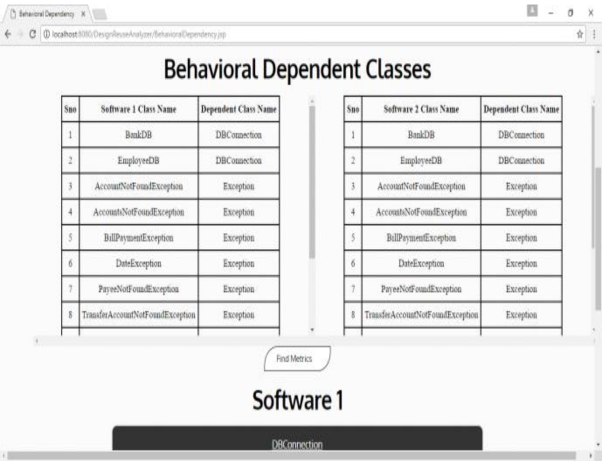
Trace Events Status for Software 2

S No	FileName	File Class Count	File Class Name	File Total Lines
1	EmployeeServlet.java	1	public class EmployeeServlet extends HttpServlet	341
2	MainControllerServlet.java	1	public class MainControllerServlet extends HttpServlet	788
3	Now.java	1	public class Now	58
4	ContextListener.java	1	public final class ContextListener	46
5	AccountNotFoundException.java	1	public class AccountNotFoundException extends Exception	25
6	AccountsNotFoundException.java	1	public class AccountsNotFoundException extends Exception	25
7	BillPaymentException.java	1	public class BillPaymentException extends Exception	25
8	DateException.java	1	public class DateException extends Exception	14
9	PayeeNotFoundException.java	1	public class PayeeNotFoundException extends Exception	25
10	TransferAccountNotFoundException.java	1	public class TransferAccountNotFoundException extends	25

Total No of Classes Present in this Software:28

Figure 4. Trace events for Software2

- ➢ Trace the Behavioral Dependency for both software V1 and V2



Behavioral Dependent Classes

Sno	Software 1 Class Name	Dependent Class Name
1	BankDB	DBConnection
2	EmployeeDB	DBConnection
3	AccountNotFoundException	Exception
4	AccountsNotFoundException	Exception
5	BillPaymentException	Exception
6	DateException	Exception
7	PayeeNotFoundException	Exception
8	TransferAccountNotFoundException	Exception

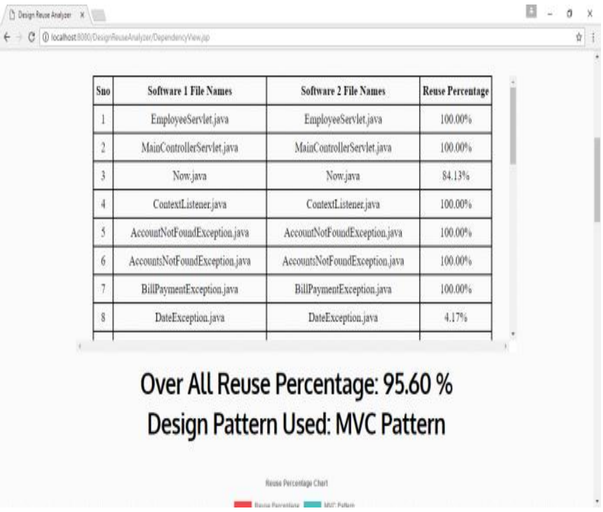
Sno	Software 2 Class Name	Dependent Class Name
1	BankDB	DBConnection
2	EmployeeDB	DBConnection
3	AccountNotFoundException	Exception
4	AccountsNotFoundException	Exception
5	BillPaymentException	Exception
6	DateException	Exception
7	PayeeNotFoundException	Exception
8	TransferAccountNotFoundException	Exception

Software 1

DBConnection

Figure 5. Behavioral Dependent Class

- ➢ Finding out the design reuse level.

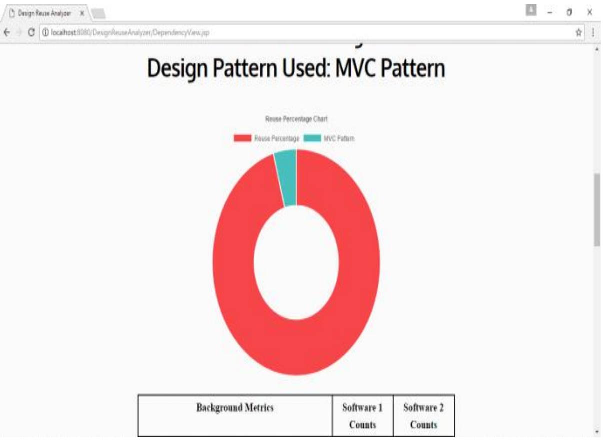


Sno	Software 1 File Names	Software 2 File Names	Reuse Percentage
1	EmployeeServlet.java	EmployeeServlet.java	100.00%
2	MainControllerServlet.java	MainControllerServlet.java	100.00%
3	Now.java	Now.java	84.13%
4	ContextListener.java	ContextListener.java	100.00%
5	AccountNotFoundException.java	AccountNotFoundException.java	100.00%
6	AccountsNotFoundException.java	AccountsNotFoundException.java	100.00%
7	BillPaymentException.java	BillPaymentException.java	100.00%
8	DateException.java	DateException.java	4.17%

Over All Reuse Percentage: 95.60 %

Design Pattern Used: MVC Pattern

Figure 6. Overall Percentage using MVC Pattern



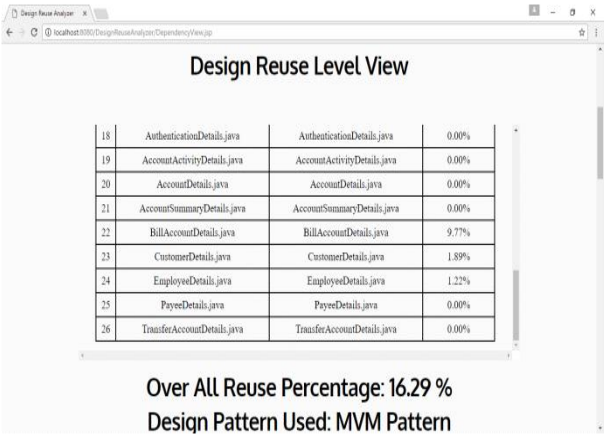
Design Pattern Used: MVC Pattern

Background Metrics	Software 1 Counts	Software 2 Counts

Figure 7. Reusability using MVC pattern

3.2. Module 2 Process Flow – Type 1:

- ➢ Upload MVC Software Version 1:
- ➢ MVC Software Version 1 are having 297 files.
- ➢ Display code analyzer for software V1: Total number of classes present in V1
- ➢ Upload MVM software
- ➢ Code analyzer algorithm works for MVM pattern for softwareV1:Total number of classes present in SoftwareV1 are 26
- ➢ Trace Events will extract both MVC Software Version 1 & MVM Software class names.
- ➢ Behavioral Dependency algorithm gets executed and it will extract all the MVC Software Version 1 &MVM Software Dependent Class Name & Depending Class Name in separate table.
- ➢ Finally displays the design reusability percentage.



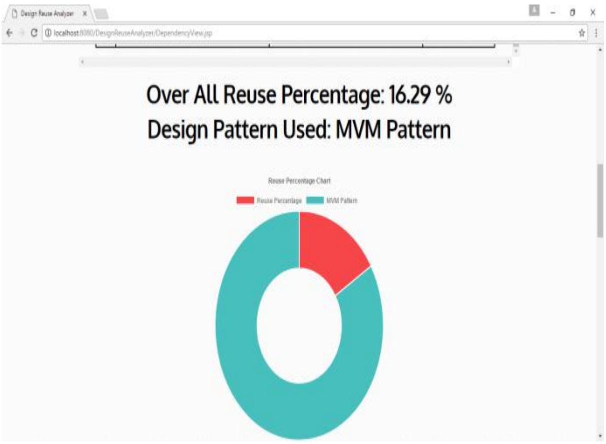
Design Reuse Level View

18	AuthenticationDetails.java	AuthenticationDetails.java	0.00%
19	AccountActivityDetails.java	AccountActivityDetails.java	0.00%
20	AccountDetails.java	AccountDetails.java	0.00%
21	AccountSummaryDetails.java	AccountSummaryDetails.java	0.00%
22	BillAccountDetails.java	BillAccountDetails.java	9.77%
23	CustomerDetails.java	CustomerDetails.java	1.89%
24	EmployeeDetails.java	EmployeeDetails.java	1.22%
25	PayeeDetails.java	PayeeDetails.java	0.00%
26	TransferAccountDetails.java	TransferAccountDetails.java	0.00%

Over All Reuse Percentage: 16.29 %

Design Pattern Used: MVM Pattern

Figure 8. Overall reuse percentage using MVM pattern



Over All Reuse Percentage: 16.29 %

Design Pattern Used: MVM Pattern

Figure 9. Reusability percentage in MVM pattern

In Background Hybrid ABCCM algorithm will works to track the Pattern Reuse Percentage for MVC Software Version 1 & MVM Software. Then extract some additional background metrics. Same process continues with different software modules.

4. CONCLUSION

In order for programmers to be able to reuse those design whose existence is not known to them, a design approach which help them in locating a pattern for reusage and converting them into components is proposed .The outcome of this research is to develop a framework for software pattern reusability at the code level .Methods by which the framework may be used to develop reusability will be proposed in the future research work.

5. REFERENCES

[1] “Review of Software Reusability” NehaBudhija and Satinder Pal Ahujain “International Conference On Computer Science And Information Technology (Iccsit'2011)” Pattaya Dec. 2011

[2] “Design Rationale And Design Patterns In Reusable Software Design” Feniosky Pena-Mora And Sanjeev Vadhavkar

[3] “Designing Code Level Reusable Software Components” ,B.Jalender 1 , DrA.Govardhan 2 , DrP.Premchand ,International Journal of Software Engineering & Applications (IJSEA), Vol.3, No.1, January 2012.

[4] Tawfig M. Abdelaziz, Yasmeen.N.Zada and Mohamed A. Hagal ,” A STRUCTURAL APPROACH TO IMPROVE SOFTWARE DESIGN REUSABILITY”

[5] “Design Patterns: Abstraction and Reuse of Object-Oriented Design” Erich Gamma , Richard Helm , Ralph Johnson, John Vlissides , Conference Proceedings, Springer-Verlag Lecture Notes in Computer Science.

[6] Reuse Design Principles ,ReghuAnguswamy and William B Frakes, International Workshop on Designing Reusable Components and Measuring Reusability Picture held in conjunction with the 13th International Conference on Software Reuse.

[7]”Software Reuse Research: Status and Future” William B. Frakes and KyoKang ,IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 31, NO. 7, JULY 2005

[8] Software Reuse: Research and Practice “Sajjan G. Shiva and LubnaAbouShala ,International Conference on Information Technology (ITNG'07).

[9] “Software Reuse” CHARLES W. KRUEGER ACM Computing Surveys, Vol. 24, No. 2, June 1992.

[10] “A Framework for Software Reuse and Research Challenges” Sathish Kumar Soora, International Journal of Advanced Research in Computer Science and Software Engineering.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 396-400, 2017, ISSN:-2319–8656

Information Communication Technology Training as A Prerequisite for E-Government In County Governments Of Kenya

Mengich Esau Mneria
Department of Information Systems
Jaramogi Oginga Odinga University of Science and Technology
Bondo Town, Kenya

Liyala Samuel
Department of Information Systems, Jaramogi Oginga Odinga University of Science and Technology
Bondo Town, Kenya

Rodrigues Anthony
Department of Information Systems, Jaramogi Oginga Odinga University of Science and Technology
Bondo Town, Kenya

Abstract: Today, the Kenyan government has invested and endeavored to embrace electronic government (e-government) in almost all ministries to enable speedy service delivery and enhance transparency and accountability by eliminating inefficient processes and bureaucracies as much as possible. However, Information Communication Technology (ICT) professionals are lacking skills to manage these projects. The general objective of this study sought to investigate ICT training as a requirement for e-government in Trans Nzoia and Kisumu county governments. This paper assesses the ICT training needs among staff working in Trans Nzoia and Kisumu County governments. Interviews and Document reviews from both case counties and national governments are the data collection methods. The study is underpinned by Structuration theory, from the field of sociology and a constructivist paradigm approach is employed. Purposive sampling technique is used targeting participants in the two county governments. Thematic Analysis's procedures and processes are adopted for data analysis. The findings contribute to knowledge in ICT training and capacity building discipline within the context of Kenyan government both national and county. We recommend a comprehensive and coherent approach to ICT training and capacity building at all educational tiers, country wide roll out of ICT infrastructure and its support for e-government to be fully embraced.

Keywords: ICT, training, e-government, County government, Structuration theory

1. INTRODUCTION

WSIS states, "Everyone should have the necessary skills to benefit fully from the Information Society [23]. Therefore, capacity building and ICT literacy are essential". Thus, it is important to look into sector policies and see how they intend to address human resource training, initial and continuous teacher development, as well as research, evaluation, and monitoring [23]. Private and public organizations form a substantial constituent of the global economy. This is because the importance of ICT for organizations (e.g. private and public) cannot be questioned and several organizations lead in the design and support of ICT implementations [20]. However, to maintain professional relevance, it is important that employees undergo a process of continuous learning and training within the organizations [18]. It appears that organizations face difficulty in developing and implementing adequate ICT training programs [2]. In the Kenyan ICT context, ICT training is lacking despite the efforts being made by the government at both the county and national governments.

Singapore, as it adheres to a developed strategic framework, has mobilized available human and capital resources to build new capability and capacity to sustain the operation of e-Government portal [11]. Capacity and capability imply receiving, filtering, digesting and managing information technology. Singapore must nurture and maintain its IT to continuously improve the living standards of Singaporeans, and to help Singapore meet international standards in public administration. In Sri Lanka, an objective of the ICT Human Resources Development Programme is to inculcate skills and competencies pertinent for the management of e-government projects [9]. Mauritius fares well in ICT literacy outreach as a result of a comprehensive and well-thought-out strategy in ensuring that not only relevant ICT literacy/proficiency programs are offered but also free ICT facilities with broadband access are made available throughout the country [16].

Kenya today is embracing Integrated Financial Management Information Systems (IFMIS) to oversee the implementation of a unified financial management system and its adoption across all Government departments. In empowering IFMIS end users, the National Treasury/IFMIS Department established IFMIS Academy as a key capacity building institution for IFMIS users [12].

The Information Communications Technology Authority (ICTA) of Kenya collaborates with academia and industry to develop structured ICT training for professionals geared towards building technical expertise with high-end skills, competencies and experience required to implement flagship ICT projects [10]. Kenya has government training institutions like the Kenya School of Government (KSG) located in Nairobi, Mombasa, Eldoret, Embu, Baringo, and Matuga. The ICT courses offered by KSG are ICT seminars for top managers, computer application skills for managers, advanced computer skills for secretaries, computer-based record management, computer networking for e-government, Statistical Package for Social Sciences (SPSS) and cyber-crime courses Kenya School of Government [13]. Kenya has also initiated the Digital Literacy Program for schools. This program coordinated at the Ministry of ICT together with the Ministry of Education and other key stakeholders provides ICT tools to primary school learners, and ensures curriculum and other learning resource books are availed through multiple electronic platforms [19]. The government, however, is facing challenges in implementing these initiatives although some progress has been made [13].

Capacity is the ability of individuals, institutions and societies to perform functions, solve problems, and set and achieve objectives in a sustainable manner [21]. Within the public service, typical aspects of capacity are: the quality of public servants, organizational characteristics, the diffusion of ICTs among organizational units, the intergovernmental relations, and the style of interaction between government and its social and economic environment [3]. Capacity Building consists of three elements namely; establishing institutional framework, engaging personnel with requisite experience, and upgrading skill sets through training. The policy thrust in Capacity Building should provide for equitable access to ICTs enabled education and training in all parts of the country including disadvantaged communities. Policies and regulatory mechanisms should be in place to help manage operations in the ICTs sector [14].

In Kenya, the majority (55%) of the youth seeking for employment are between the ages of 21-25 years old. 57% of the unemployed youth lack ICT training. Out of the remaining 43% that have ICT training, 28% have basic computer application packages such as Microsoft office [22]. According to the Ministry of Information, Communication and Technology, there is demand for local high-end ICT professionals from National and County Government that aim to provide e- government services and to automate their internal processes.

Governments both at national and county levels have made huge strides towards the adoption of ICT; however, there is dearth of ICT professionals in the county governments of Trans Nzoia and Kisumu for example.

In this paper we assess ICT training needs as an enabler of good governance in Trans Nzoia and Kisumu county governments. The results contribute to deeper understanding of ICT training in Government of Kenya as it strives to be a leader in the region.

2. OVERVIEW OF STRUCTURATION THEORY

Anthony Giddens, is regarded as one of the worlds' most cited sociologists [1]. As Roberts and Scapens note: "Through being drawn on by people, pattern and shape structures themselves are, however, reproduced only through interaction [17]. Explanations of social phenomena must thus refer to both the role of human action and the effects of existing institutional properties. The approach of Giddens' structuration theory argues that action and structure operate as a duality, simultaneously affecting each other [6]. Giddens defines structure as 'rules and resources recursively implicated in social reproduction; institutionalized features of social systems have structural properties in the sense that relationships are stabilized across time and space' [7]. The structuration-type theory advances our understanding of the use of ICT in the government sector as it provides a means of handling the complexity of the interactions between citizens, organizations, the government and other industry sectors, within the national and international context.

Giddens opines that structure is similar to language [6]. While speech acts are situated temporally and contextually and involve dialogue between humans, language exists outside of space and time. Language is a condition for achievements of dialogue and language is sustained through the ongoing production of speech acts. Social actions are temporary and contextually involve human interaction. Social structures conditions these social practices by providing the contextual rules and resources that allow human actors to make sense on their own acts and those of other people. Conceiving of structure in this way acknowledges both its subjective and objective features. Structure does not merely emerge out of subjective human action; it is also objective because it provided the conditions for human action to occur. Structure provides the means for its own sustenance and structure and action constitute each other recursively. Structuration theory recognizes that "….man actively shapes the world he lives in at the same time as it shapes him" [7].

From structuration theory, the training issues build on what Giddens calls Communication dimension: the structure of *Signification*, modality of *Interpretive schemes* and interaction of *Communication* in depth. The structure of *signification* is available to the social actor as *interpretative scheme(s),* the structure of *domination* is available as authoritative and allocative *resources* through which power can be exercised and the *legitimation* structure is available as norms, which allow for the *sanctioning* of conduct [8].

Interpretive schemes is defined by Giddens as the "standardized, shared stocks of knowledge that humans draw on to interpret behavior and events, hence, achieving meaningful interaction". It is through ICT training that humans are able to interprete their lived experience with interaction of ICT to create structure of *Significance* [6].

3. Problem Formulation

The Kenyan government in recent years has initiated some capital investment towards the setting up and installation of ICT infrastructure through establishment of ICT training programs. ICTs are about information flowing faster, more generously, and less expensively throughout the planet. As a result, knowledge is becoming an important factor in the economy, more important than raw materials, capital, labor, or exchange rates [4]. As much as the Kenyan government is investing heavily on ICT, training must be a prerequisite to manage and maintain the modern systems that are associated with dynamic changes over time. Kenya needs ICT professionals to take advantage of established ICT infrastructure.

4. Methods

4.1 Participants

Using purposive sampling method, we interviewed a sample of fourteen top and middle level management employees from both counties. They were eight males and six females. A pilot study of two participants was done and they were isolated from the final research. All participants were employees with daily duties using ICT devices. Secondly, the participants must have been employees of Trans Nzoia and Kisumu counties or national government employees attached to these counties.

4.2 Research Design

This study adopted a qualitative case study research design. The strengths of a qualitative design is that it emphasizes people's lived experience of a phenomena. A case study was adopted because a county is bounded by time and place (boundary) and this paper was a multiple case study.

4.3 Measures

Interviews, observations and document reviews were used as data collection methods. Face-to-face interviews and observations were the primary data collection methods. In-depth interviews were conducted and audio-recorded through a laptop and later keenly transcribed into word documents. For observations, notes taking were done in the event of observations. Document reviews are secondary sources of data collections. They represent variety of non-personal documents such as minutes of meetings, agendas of such meetings and office memos. In the Kenyan context, government circulars, specific ministry master plans and any other official correspondences were perused to give more information on the use of ICT innovations and or applications.

4.4 Procedures

Consent was sought from Trans Nzoia and Kisumu county governments, particularly the offices of County Commissioners and County Directors of Education of both counties respectively. Data collection started immediately. Top and Middle level management employees from both counties were targeted on a voluntary basis. Confidentiality, the principle of anonymity and the right to de-briefing to participants are examples of ethical issues spelled out before interviewing commenced.

4.5 Data Analysis

In this paper, data analysis means understanding the ways county employees use and make sense of innovative applications, but also identifies and defines the patterns that emerge from that meaning making process. Thematic Analysis which is a type of qualitative analysis was applied. [15] model of Thematic Analysis was adopted comprising of three link stages: data reduction - "A form of analysis that sharpens, sorts, focuses, discards, and organizes data in such a way that "final" conclusion can be drawn and verified"; data display – "displaying the data in a variety of ways e.g. tables, figures and theme maps" and data conclusion – "ideas to generate meaning from the data and they include: the notation of any patterns or themes / Grouping or establishing categories of 'information that can go together".Extracts of this thematic analysis is contained in Table 1 and 2 respectively.

Table 1: Sample of Unordered list of responses to interview open-ended questions

Participant responses	
They initially resisted new system but later accepted	INT 3/ INT 1
We have a website with a lot of information about the county	INT 1
Our county has not integrated much in ICT	INT 5
The SMS platform not being used much despite mobile phones	INT 3
Citizens use emails to make inquiries or feedbacks	INT 1
National government use emails to inform on seminars/workshops/circulars etc	INT 1/INT 3/INT 2

Table 2: Sample Categorization of responses to Interview open ended questions

Inductive responses	Participant responses
Capacity building/Training Only two ICT professionals in my ministry	INT 3
Some staff are qualified but cannot handle practicals	INT 4
IFMIS should have started as a pilot in few counties	INT 2
Lack of competent staff to run ICT equipment /systems	INT3
ICT directorate not fully fledged still under Education Ministry More than 90% of staff used typewriters; Today most of them are still challenged	INT 3

As stated earlier, ‘data conclusion’ is considering ideas to generate meaning from the data and they include the notion of any patterns or themes and grouping them. From the study, ICT Capacity building and training emerged as a major theme. This theme was compared with the literature as shown in the following section under where it either supported or disputed findings.

5. Study Findings

The guiding theoretical framework for this study was Structuration theory which emphasizes that structures enable and constrain three basic aspects of (inter)action: Communication, the exercise of Power, and the Sanctioning of conduct. Each of these three basic activities relates to a particular structural dimension: Communication to the structure of signification, Sanctioning to the structure of legitimation, and power ‘plays’ to the structure of Domination [5].

There was evidence of structuration theory through the Communication dimension: the structure of *Signification*, modality of *Interpretive schemes* and interaction of *Communication* in the study findings. In concurring with study findings, ICT training is linked with *Interpretive schemes* defined by [5] as the “standardized, shared stocks of knowledge that humans draw on to interpret behavior and events, hence, achieving meaningful interaction”. It is through ICT training that humans share stocks of knowledge and stored, for instance, by Knowledge Management for future use. There was evidence of training, seminars, workshops and capacity building organized events periodically by national government for employees.

The Kenyan government also supports and builds capacity of the employees in both county and national governments of Kenya. For example: the IFMIS department, see [12], offers training in all ministries and government agencies to facilitate use of the system by all employees and the Ministry of Education, Science and Technology facilitates ICT training and capacity building initiatives for teachers in primary schools. IFMIS, concurs with this finding [12].

Kenya must emphasize training and capacity building to attain the goal of being a nation that is strategically positioning herself as a technological hub in Africa. Strategic capacity is seriously lacking in Africa and hence a huge constraint in keeping up with technological innovations and global competition. ICT training is still an impediment to the diffusion of ICT in Kenyan county governments calling for practical measures to speed up the process.

6. Conclusions and Recommendations

We have shown that the government of Kenya has made progress in ICT training and capacity building initiatives considering that Kenya is relatively young country and our counties in particular that were established in 2010 effectively became operational only in April-2013. Hence, the full potential of ICT has not been realized to date. Although efforts have been made towards ICT training, there is need to focus on the youth who form the largest percentage of the Kenyan population. The government of Kenya is making positive strides on ICT implementation longitudinal studies are required.

In recommendations, ICT Training and capacity building should be implemented from primary schools, secondary, tertiary colleges and at university levels. This means that the government through Ministry of Education should come up with ICT curriculum that addresses the above levels of training and capacity building in a systematic and coherent manner. The government should emphasize the need for employers to continue training and building capacity of their employees. The solutions to these issues require goodwill by leaders in government, employees, citizens and the private sectors.

For all employees of governments, training and capacity building is essential in order to maximize use of ICT for good governance. Citizens of Kenya are urged to embrace ICT as the government makes strides to provide ICT which enables service delivery anywhere, at any time and in an instance.

7. ACKNOWLEDGEMENTS

The author wishes to acknowledge Dr. Liyala, my first supervisor and Prof. Rodrigues, my second supervisor for facilitating the PhD program. The Lecturers and employees of the School of Informatics & Innovative Systems at JOOUST, Kenya and the University as a whole. Finally, to the Almighty God who made all these happen.

REFERENCES

[1] Bryant, C., G., A., and Jary, D., (2001) Anthony Giddens: Critical Assessments, Routledge, London, UK.

[2] Bocij, P., Chaffey, D., Greasley, A. and Hickie, S., (1999). 'Business Information Systems', Technology, Development and Management.

[3] DAC Network on Governance, (2006). "The Challenge of Capacity Development: Working Towards Good Practice"

[4] Delone, W.,H., and Mclean, E.,R., (2002)-last update, Information Systems Success Revisited http://www.csdl.computer.org/comp/proceedings/hicss/2002/1435/08/14350238.pdf [08.12.2004].

[5] Giddens, A., (1976). New rules of sociological methods. Basic books, New York.

[6] Giddens, A., (1979): Central Problems in Social Theory: Action,Structure and Contradictions in Social Analysis. London: Macmillan.

[7] Giddens, A, (1982). Profiles and critics in social theory. University of California press, p. 21, Berkeley, CA.

[8] Giddens, A. (1984). The Constitution of Society, Polity Press, Cambridge.

[9] GreenTech Consultants (Pvt.) Ltd., (2011). International Development Association (IDA)-Information and Communication Technology Agency of Srilanka (ICTA) Volume 2 Government Organizations Employees Survey (Goes) RFP No.: ICTA/CON/QBS/P1/359 Survey Final Report (Original).

[10] ICTA, (2015). Accessed on 13/07/2015 from http://www.icta.go.ke/ict-human-resource-skills/).

[11] IDA, Infocomm Development Authority of Singapore, (2000): [cited 5September 2005]. http://www.ida.gov.sg/idaweb/media/infopage.jsp?infopagecategory=factsheet:aboutida&versionid=5&infopageid=I853.

[12] IFMIS, (2015). Accessed on 18/07/2015 from http://www.ifmis.go.ke/?page_id=2432/

[13] KSG - Kenya School of Government, (2015). Accessed on 08/07/2015 from http://www.nairobi.ksg.ac.ke

[14] Kundishora, S., M., (2003). "The Role of Information and Communication Technology (ICT) in Enhancing Local Economic Development and Poverty Reduction". Chief Executive Officer, Zimbabwe Academic and Research Network

[15] Miles, M., B., and Huberman, A., M., (1994). *Qualitative data analysis*. Thousand Oaks:
Sage.

[16] Oolun, K., Ramgolam, S., & Dorasami, V., (2012). The Making of a Digital Nation: Toward i-Mauritius.

[17] Roberts, J., and Scapen, R., (1985). Accounting systems of accountability: "Understanding accounting practices in their organizational context." Accounting, organizations and society. 10, 4, 443-456.

[18] Sambrook, S., (2003). 'E-learning in Small Organisations', Journal of Education and Training, 40(8/9): 506-516.

[19] Tiampati, J., (2015). SAP skills for Africa Graduation at Chiromo Campus:http://www/information.go.ke Retrieved 1/10/2015 6.07 pm.

[20] Van Weert, T., J., and Pilot, A., (2003). 'Task-Based Team Learning with ICT, Design and Development of New Learning', Education and Information Technologies, 8(2): 195-214.

[21] Verheijen, A.,J.,G., (2000). "Administrative Capacity Development. A race against time?", The Hague.

[22] Waswa, M., G., (2011). Information and Communication Technology (ICT) Access for Training and Employment Opportunities by Kenyan Youth. A Case Study of Youth Living in Nairobi. Faculty of Information Technology Strathmore University Nairobi, Kenya June.

[23] WSIS –World Summit Of Information Society, (2003). Geneva Declaration of Principles and Geneva Plan of Action.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 401-404, 2017, ISSN:-2319–8656

Web-Based Programming: A Veritable Tool for Security and National Development

Ezeano, A. N
Computer Science Department, Akanu Ibiam Federal Polytechnic, Unwana, Nigeria

Idemudia, O. J
Computer Science Department, Akanu Ibiam Federal Polytechnic, Unwana, Nigeria

Madubuike, C. E.
Computer Science Department, Akanu Ibiam Federal Polytechnic, Unwana, Nigeria

Omoregbee, E.U.
Petroleum Training Institute, Effurun, Delta State, Nigeria

Onuorah, A. C
Computer Science Department, Akanu Ibiam Federal Polytechnic, Unwana, Nigeria

Abstract: In all industrialized countries and increasingly in developing countries web based computer systems are economically critical. More and more products and services are incorporate inside the web-based information system. Education, administration, banking, oil and gas and health care services etc. are all dependent on a web-based-flavoured information communication technology. The effective functioning of a modern economic and political system which is a precursor to national development depend on the skill flaunted by IT expert to produce a flawless and less complex web-based products and services. This paper proposes an effective method of achieving national development and security via the resourcefulness and instrumentality of the numerous activities on the Web platform. The proposed model is anchored on four components: green computing and energy star; cloud-based services; Web start-ups; and adequate government response programmes..

Keywords: web-based computer systems, web-based-flavoured information communication technology, best industry practise, simplified framework

1. INTRODUCTION

The past thirty years, the Internet, has become a critical enabler of social and economic change, transforming how government, business and citizens interact and offering new ways of addressing development challenges. And today, with the scalability of the Web technologies, there are things like Websites, Social networks, affiliate marketing, blogging, tweeting, virtual library, online survey, virtual office. Also, there are new terms like "apps industry", "apps economy", "apps start-ups", "microwork" etc.

This technological transformation has been made possible because the newest generation of Web platform (Web 2.0) enables one to create interest groups on the internet and share information in form of photos, videos, music, logs of plans, web browsers bookmarks etc; creating a rich array of user-generated content. With this crop of activities, which are too numerous to mention, private and public institutions can take advantage of the Web blossom there economy while ensuring national security.

Research shows that Facebook apps alone created over 182,000 jobs in 2011, and that the aggregate value of the Facebook app economy exceeds $$12 billion. Web programming through its startup programme is creating over 46,000 jobs and 12billion dollars economic activity in UK alone [10].

National development could be defined as the ability of a county or countries to improve the social welfare of the people e.g by providing social amenities like quality education, potable water, transportation infrastructure, medical care, etc. Most national development plan seeks to adopt a framework of inclusive growth, which is high growth that is sustained, generates mass employment, and reduces poverty. While national security is the requirement to maintain the survival of the state through the use of economic power, diplomacy, power projection and political power. The concept of national security was developed mostly in the United States after World War II. Initially focusing on military might, it now encompasses a broad range of facets, all of which impinge on the non-military or economic security of the nation and the values espoused by the national society. Accordingly, in order to possess national security, a nation needs to possess economic security, energy security, environmental security, etc. Security threats involve not only conventional foes such as other nation-states but also non-state actors such as violent non-state actors, narcotic cartels, multinational corporations and non-governmental organisations; some authorities include natural disasters and events causing severe environmental damage in this category. Measures taken to ensure national security include: using diplomacy to rally allies and isolate threats; marshalling economic power to facilitate or compel cooperation; maintaining effective armed forces; implementing civil defense and emergency preparedness measures (including: anti-terrorism legislation); ensuring the resilience and redundancy of critical infrastructure; using intelligence services to detect and defeat or avoid threats and espionage; and to protect classified information using counterintelligence services or secret police to protect the nation from internal threats.

2. ASSESSMENT OF WEB PROGRAMMING TO DATE

2.1 Summary of the History of Web Programming

Though early stage of the Web evolution, Web 1.0, which existed between 1990 and 2000 [5] enjoyed some level of growth due to its multiuser interface; single point maintenance and updates; distributed and hyperlinked documents etc. The level of popularity and user activity was still low owing to the fact that most of the websites developed using web1.0 was static and operated in brochure architecture with only professional web designers producing the content for users to access.

The Web platform we enjoy today, Web 2.0 has grown tremendously with resurgence of popularity and interest from millions of companies and billions of users across the world. Web 2.0 operates in architecture of participation were companies only provide the platform and users generate the content. Most sites on the Internet today like wikis, blogs and social media sites all present user generated content bringing the shift from few powerful professionals (programmers) to many empowered users[11]. In-lieu of this development, so many tools have been introduced to enable more user participation such as the Rich Internet Applications (RIA) and AJAX (Asynchronous JavaScript and XML) technologies. These technologies are used to develop web applications, which look and behave like desktop applications. At the root of this Web 2.0 evolution, which is triggered by technologies like AJAX, Document Object Model (DOM), RIA, frameworks etc, is enshrined the concept of Object-Oriented Programmed (OOP). Hence, the knowledge of these tools and OOP cannot be ignored.

2.2 Literatures on web programming

Many of the studies in the area of web application development have mainly focused on the evolution of web application and comparison of web application development languages. Jazayeri wrote on trends and status quo of web application [13]. Ronacher presented security related issues in web application [17]. Voslro and Kourie wrote on concepts and web framework [17]. Purer highlighted some differences, advantages and drawbacks of PHP, Python and Ruby [16]. He compared the languages based on history, evolution, popularity, syntax, semantics, features, security and performance in web application environments. Cholakov analyzed PHP and summarized some drawbacks[3]. Gellersen and Gaedke in their article [8], overviewed object oriented web applications and identified object-oriented model for web applications, they found that XML technology contributes in enabling high level abstractions for design level modeling in a markup language. Mattsson identified the strengths and weaknesses of object oriented frameworks [12]. Finifter and Wagner explored the relation between web application development tools and security [6]. Chatzigeorgiou et al, evaluated object oriented design with link analysis [2]. Paikens and Arnicans explored the use of design patterns in PHP-based web application frameworks [14]. French presents a new methodology for developing web applications and web development life cycle [7]. Copeland et al, in their article titled “Which web development tool is right for you” discussed and compared various tools for web application development [4].

However, not too many studies have been conducted in the area of impact of object oriented programming on web application development. This research aims at discussing the impact of object oriented programming on web application.

2.3 Challenges with web programming

Some of the identified challenges with web programming include:

1. Requires expert knowledge: developers or programmers of Web applications are required to have special knowledge in basic programming concepts, client-side and server-side scripting, Internet technologies and client-server technologies.
2. Provisions of power: Devices which run and support web applications runs on power; hence, there is need for provision of constant supply of power and integration of low power devices.
3. Internet connection
4. misuse of technology and information overload
5. Cyber crimes
6. Virus and worms
7. Environmental pollution: Most of the electronic devices – including Web- support devices – are toxic and bio non-degradable devices; unfortunately, the developing world dump sites for these devices. They are in most cases not disposed properly; thereby causing environmental hazards.

2.4 How Web Programming Can Boost National Development and Security

At a time of slowed growth and continued volatility, many countries are looking for policies that will stimulate growth and create new jobs. Information communications technology (ICT) is not only one of the fastest growing industries – directly creating millions of jobs – but it is also an important enabler of innovation and development.

The number of mobile subscriptions (6.8 billion) is approaching global population figures, with 40% of people in the world already online. In this new environment, the competitiveness of economies depends on their ability to leverage new technologies. Here are the five common economic effects of ICT.

1. Direct job creation: The ICT sector is, and is expected to remain, one of the largest employers. In the US alone, computer and information technology jobs are expected to grow by 22% up to 2020, creating 758,800 new jobs. In Australia, building and running the new super-fast National Broadband Network will support 25,000 jobs annually. Naturally, the growth in different segments is uneven. In the US, for each job in the high-tech industry, five additional jobs, on average, are created in other sectors. In 2013, the global tech market will grow by 8%, creating jobs, salaries and a widening range of services and products.

2. Contribution to GDP growth: Findings from various countries confirm the positive effect of ICT on growth. For example, a 10% increase in broadband penetration is associated with a 1.4% increase in GDP growth in emerging markets. In China, this number can reach 2.5%. The doubling of mobile data use caused by the increase in 3G connections boosts GDP per capita growth rate by 0.5% globally. The Internet accounts for 3.4% of overall GDP in some economies. Most of this effect is driven by e-commerce – people advertising and selling goods online.

3. Emergence of new services and industries: Numerous public services have become available online and through mobile phones. The transition to cloud computing is one of the key trends for modernization. The government of Moldova is one of the first countries in Eastern Europe and Central Asia to shift

its government IT infrastructure into the cloud and launch mobile and e-services for citizens and businesses. ICT has enabled the emergence of a completely new sector: the app industry.

4. Workforce transformation: New “microwork” or “start-ups” platforms, developed by companies like oDesk, Amazon and Samasource, help to divide tasks into small components that can then be outsourced to contract workers. The contractors are often based in emerging economies. Microwork platforms allow entrepreneurs to significantly cut costs and get access to qualified workers.

5. Business innovation: In OECD countries, more than 95% of businesses have an online presence. The Internet provides them with new ways of reaching out to customers and competing for market share. Over the past few years, social media has established itself as a powerful marketing tool. ICT tools employed within companies help to streamline business processes and improve efficiency. The unprecedented explosion of connected devices throughout the world has created new ways for businesses to serve their customers.

3. THE WEB-BASED MODEL FOR NATIONAL SECURITY AND DEVELOPMENT

Our proposed Web-based model for achieving development and national security will be anchored four major components: green computing and energy star; IT start-ups; cloud based infrastructure; and adequate government response project. Our idea is to propose a system that will empower the teaming youths through the start-ups, cloud services and government programmes; as well the use of economically viable and eco-friendly computer through the green computing standards.

1. Green computing and Energy star: Green computing is the design, use and disposal of computers and its associated resources in an eco-friendly manner. In age where world leaders making strong efforts in other to achieve sustainable development, there is every need to respect and promote our ecosystem [19] – the Igbo man will say “Ndu bu isi”, meaning life is the head.

   Green computing aims at providing economic viability of improved computing devices. Green IT practices include the development of environmentally sustainable production practices, energy efficient computers and improved disposal and recycling procedures. Energy star is one the program that implements green computing.

2. Web-based start-ups: These are small businesses that derive their benefit from the numerous activities on the Web platform. They range from digital marketers, social media specialist, Web designers, Network administrators, Web programmers, Media content developers, Hardware engineers, Affiliate marketers etc

   In London, the Silicon Roundabout and City Tech Initiative has generated over 46,000 jobs and addition 12 billion Euros to the economic activities over the past four years via their clusters of start-ups. The Web start-ups in UK have generated some many jobs to the extent that they have “apps economy” [10, 15]. Also in 2012, oDesk alone had over 3 million registered contractors who performed 1.5 million Web-related tasks. This trend had spill over effects on other industries, such as online payment systems. The Web has also contributed to the rise of entrepreneurship, making it much easier for self-starters to access best practices, legal and regulatory information, and marketing and investment resources

3. Cloud based services: These are services which are delivered through the internet instead of a local computer/device. In the simplest terms, cloud computing means storing and accessing data and programs over the Internet instead of your computer's hard drive. The cloud is just a metaphor for the Internet. When you store data on or run programs from the hard drive, that's called local storage and computing.

   With cloud computing, a user form large corporation or from small start-ups can log into a Web-based service which hosts all the programs he or she would need. Remote machines owned by these companies would run everything from e-mail to word processing to complex data analysis programs from the internet; thereby saving cost and space, without losing standards [9, 1].

4. Government Response Project: since our independence in 1960, different Nigeria government have implemented one project or the other geared towards addressing the plight of the poor masses as well as improving the standard of people. This in FEDAMA project; the NEEDS project etc. How many of this project have addressed technology and to what extent? What about things like: IT Development Banks; IT Micro Loans; Subsidy programme for common IT components like breadboards, resistors; IT- research centres/institutes; IT local content bill (mention but a few) ?

   In our Information – driven era, our government should provide adequate response programmes and policies to groom small IT start-ups and to strengthen the bigger ones.

4. CONCLUSION

In this paper we have discussed how national development and security can be achieved through Web programming and the numerous activities of the Web. Our idea is to propose a system that will empower the teaming youths through the start-ups, cloud services and government programmes; as well the use of economically viable and eco-friendly computer through the green computing standards.

5. RECOMMENDATIONS

In this paper we have discussed how national development and security can be achieved through Web programming and the numerous activities of the Web. Our idea is to propose a system that will empower the teaming youths through the start-ups, cloud services and government programmes; as well the use of economically viable and eco-friendly computer through the green computing standards.

6. REFERENCES

[1] Beal, V (2016). Cloud computing (Cloud). Webopedia, QuinStreet Entreprise, http://www.webopedia.com/TERM/C/cloud_computing.html , accessed on 30/05/16

[2] Chatzigeorgiou Alexander, Xanthos Spiros & Stephanides George (2004). Evaluating Object Oriented designs with link analysis; Proceeding of the 26th International Conference on Software Engineering, IEEE Computer Society

[3] Cholakov, N. (2008). On some drawbacks of the php platform, CompSysTech '08: Proceedings of the 9th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing (New York, NY, USA), ACM, pp. II.7.

[4] Copeland Dennis R., Corbo Raymond C., Falkenthal Susan A., Fisher James L., & Sandler Mark N. (2000). Which Web Development Tool Is Right for You? IT Pro IEEE

[5] Deitel, P. J. & Deitel, H. M. (2007), Java How To Program, USA, Pearson Inc., 7th Ed., pp. 421 - 423.

[6] Finifter Mathew & Wagner David (2010). Exploring the Relationship between Web Application Development Tools and Security

[7] French M. Aaron (2011). Web Development Life Cycle: New Methodology for Developing Web Applications, Journal of Internet Banking and Commerce, vol.16, no.2

[8] Gellersen, H. & Gaedke, M. (1999). Object-Oriented Web Application Development. IEEE Internet Computing. Accessed from http://computer.org/internet on 3/8/15

[9] Griffith, E. (2016). What Is Cloud Computing? PcMag Digital Group, http://www.pcmag.com/article2/0,2817,2372163,00.asp accessed on 30/05/16

[10] Innes, G. (2014). How we can turn London's thriving tech startups into billion dollar digital giants accessed from http://www.cityam.com/1416196266/how-we-can-turn-london-s-thriving-tech-startups-billion-dollar-digital-giants on 18/05/16

[11] James, G.,Hinchcliffe, D. & Nickull, D. (2009). Web 2.0 Architectures. 1st ED, O'REILLY Media Inc. ISBN: 978-0-596-51443-3

[12] Mattsson, M. (1996). Object-Oriented Frameworks: A Survey of Methodological ssues. CODEN:LUTEDX(TECS-3066)/1-131

[13] Mehdi, J. (2007) Some trends in web application development, FOSE '07: 2007 Future of Software Engineering (Washington, DC, USA), IEEE Computer Society, pp. 199

[14] Paikens An dris and Arnicans Guntis (2008). Use of Design Patterns in PHP-Based Web Application Frameworks, Latvijas Universitātes Raksti:Datorzinātne Un Informācijas Tehnoloģijas

[15] Papadopoullos, C. (2016).UK's "app economy" boosts demand for web designers and programmers from http://www.ctyam.com/1416197554/app-creators-drive-tech-skills-demand accessed on 19/05/16

[16] Purer, K. (2009). PHP vs. Python vs. Ruby - The web scripting language shootout; Vienna University of Technology, Institute of Computer Languages, Compilers and Languages Group, 185.307 Seminar aus Programmiersprachen

[17] Ronacher, A.(2006) Sicherheit in Webanwendungen, http://dev.pocoo.org/blackbird/fachbereichsarbeit.pdf

[18] Techopedia (2016) What is Green Computing? - Definition from Techopedia *https://www.techopedia.com/definition/14753/**green-computing accessed on 25/05/16***

[19] United Nations (2015). Biodiversity and Ecosystem., United Nations, https://sustainabledevelopment.un.org/topics/biodiversityandecosystems, accessed on 25/05/16

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 405-409, 2017, ISSN:-2319–8656

Image Indexing Using Color Histogram and K-Means Clustering for Optimization CBIR

Juli Rejito
Department of Computer Science
Padjadjaran University
Bandung, Indonesia

Deni Setiana
Department of Computer Science
Padjadjaran University
Bandung, Indonesia

Rudi Rosadi
Department of Computer Science
Padjadjaran University
Bandung, Indonesia

Abstract: Retrieving visually similar images from image database needs high speed and accuracy. The researchers are investigating various text and content based image retrieval techniques to match the image features accurately. In this paper, a content-based image retrieval system (CBIR), which computes colour similarity among images, is presented. CBIR is a set of techniques for retrieving semantically relevant images from an image database based on automatically derived image features. The colour is one important visual elements of an image. This document gives a brief description of a system developed for retrieving images similar to a query image from a large set of distinct images with histogram colour feature based on image index. Result from the histogram colour feature extraction, then using K-Means clustering to produce the image index. Image index used to compare to the histogram colour element of a query image and thus, the image database is sorted in decreasing order of similarity. The results obtained by the proposed system apparently confirm that partitioning of image objects helps in optimization retrieving of similar images from the database. The proposed CBIR method is compared with our previously existed methodologies and found better in the retrieval accuracy. The retrieval accuracy is comparatively good than previous works offered in CBIR system.

Keywords: *CBIR, Image Features, Colour Histogram, K-Means clustering*

1. INTRODUCTION

In CBIR, the feature vector is extracted from images. Query feature vector is matched with the stored feature vector on one to one basis, which resulted in slow down the processing time. To improve the speed of executing and better result, many researchers are paying attention to uses the clustering algorithm for CBIR. Clustering is a process of separating a data set into groups in such a way that the object in one group is more similar to those objects in the other group [1].

Chin-Chin Lai et.al. [2] have proposed an interactive genetic algorithm (IGA) to reduce the gap between the retrieval results and the users' expectation called semantic gap. They have used HSV color space that corresponds to the human way of perceiving the colors and separate the luminance component from chrominance ones. They have also used texture features like the entropy based on the grey level co-occurrence matrix and the edge histogram. They compared this method with others approaches and achieved better results.

Kannan A, et al [3] have proposed Clustering and Image Mining Technique for fast retrieval of Images. The primary objective of the image mining is to remove the data loss and to extract the meaningful information to the human expected needs. The images are clustered based on RGB Components, Texture values and Fuzzy C mean algorithm.

Chakravarti and Meng [4] have published a paper on Color Histogram Based Image Retrieval. They have used color histogram technique to retrieve the images. This method allows retrieval of images that have been transformed regarding their size as well as translated through rotations and flips.

Kumar, et.al [5] have published on Content Based Image Retrieval using Color Histogram. They have used Color Histogram technique to retrieve the similar images. To speed up the recovery, they have used the proposed grid-based indexing to obtain the nearest neigh hours of the query image, and accurate images are retrieved. Indexing can be performed in vector space to improve retrieval speed. Mainly, they have implemented CBIR using color histogram technique and refined with the help of grid method to improve the image retrieval performance.

CBIR with clustering algorithm is an alternative that is expected to improve the performance and accuracy of searches in the query image. K-means is the core clustering algorithm, but in this case, k-means is very sensitive for first grouping and delicate to outliers and noise [1,6]. Also, to form the clusters, K-mean depends on initial condition, which causes the algorithm to give a suboptimal solution. As compare to k-means ant colony algorithm are more prominent for initializing the cluster. Due to the optimal global nature of particle swarm optimization algorithm give the optimum solution for clustering. These most prominent features of ant colony algorithm and particle swarm optimization are used for clustering.

The implementation of a query optimization proposed in this paper was related to CBIR in image databases aimed at obtaining the image in the database with a high image content similarity level during an image searching process. In this proposal, proposed by carrying out a color extraction process of each image database, a method of clustering with the k-means algorithm, and results of clustering of each image database were used to a filtering process based on query images.

2. LITERATURE REVIEW

2.1 CBIR

CBIR is a method that is used to look at image features like (color, shape, texture) to find a query image from the database. The difficulties of CBIR lie in reducing the differences of contents based feature and the semantic based functions. This problem in giving useful retrieval images and channelize the researchers to use (CBIR) system, to take global color and texture features to achieve, the right recovery, where others used local color and texture features [6]. The method in [7] presented the holistic representation of spatial envelope with a very low dimensionality for making the incident image.

The method in [8] proposed a modern approach for image classification with the open field design and the concept of over-completeness methodology to achieve an excellent result. As reported in [8], this method produced the best classification performance with much lower feature spatiality compared to that of the former schemes in image classification task. For similarity search [9] the user needs to enter keywords along with the query image that might appear in the text of patents. Higher average precision and recall rates compared to the traditional Dominant Color method were obtained successfully [10]. The texture and color attributes are computed in a way that model the Human Vision System (HSV) [11].

2.2 Color Histogram Feature Extraction

Obtained by extraction of the color histogram feature extraction of image pixels for each color component R, G, and B then calculated the frequency of each color index from 0 to 255 and raised in the form of histogram value for each color component, and is written as a vector shown in the equation as follows:

$$H = \{H[0], H[1], H[2], H[3], \ldots, H[i], \ldots, H[n]\} \quad (1)$$

Where i is the color in the color histogram storage and H [i] indicates the number of pixels of color i in the image, and n is the number of colors used in the storage of the color histogram.

Results histogram value for each color component (x, y, z) of the image is sought (H_q) and record images (H_i) then calculate resemblance to calculate the distance to the known color histogram. Histogram Intersection Technique (HIT) [12], using the formula in the equation as follows:

$$S(H_q, H_i) = \frac{\sum_{x \in X, y \in Y, z \in Z} \min\left(H_q(x,y,z), H_i(x,y,z)\right)}{\sum_{x \in X, y \in Y, z \in Z} H_q(x,y,z)} \quad (2)$$

Possess the formula distance values tend to have small differences so that the formula developed into the equation as follows:

$$S(H_q, H_i) = \frac{\sum_{x \in X, y \in Y, z \in Z} \min\left(H_q(x,y,z) H_i(x,y,z)\right)}{\min\left[\sum_{x \in X, y \in Y, z \in Z} H_q(x,y,z), \sum_{x \in X, y \in Y, z \in Z} H_i(x,y,z)\right]} \quad (3)$$

2.3 K-Means Algorithm

The k-means algorithm is effective in producing good clustering results for many applications [13]. The reasons for the popularity of k-means are ease and simplicity of implementation, scalability, speed of convergence and adaptability to sparse data [14]. K-means clustering is a partitioning clustering technique in which clusters are formed with the help of centroids. From these centroids, groups can vary from one another in different iterations. Also, data elements can differ from one cluster to another, as groups are based on the random numbers known as centroids [15]. The k-means algorithm is the most extensively studied clustering algorithm and is effective in producing good results. The k-means algorithm is computationally expensive and requires time proportional to the product of the number of data items, the number of clusters and the number of iterations. K-means is formally described by Algorithm 1.

```
Algorithm 1: Basic K-means algorithm.
1: Select K points as initial centroids.
2: repeat
3: Form K clusters by assigning each point to its   closest
   centroid.
4: Recomputed the centroid of each cluster.
5: until Centroids do not change.
```

2.4 Measurement Image Quality

Measurements with this model are widely used because of the ease of calculation, have a physical understanding and mathematically easy to use for optimization purposes, although not very useful in matching visual quality. This measurement consists of several formulas that MAE (Maximum Absolute Error), MSE (Mean Square Error), RMSE (Root Mean Square Error), SNR (Signal to Noise Ratio), and PSNR (Peak Signal to Noise Ratio) [16].

2.4.1 Maximum Absolute Error (MAE)

MAE is the highest value of the absolute value difference between the input image $f(x, y)$ and the output image $g(x, y)$. *MAE* calculation is mathematically written in equation form:

$$MAE = max\,|f(x, y) - g(x, y)| \quad (4)$$

where $f(x, y)$ is the value of the initial / original image intensity in the position *(x, y)* and $g(x, y)$ is the value of the image's intensity in the position (x, y).

2.4.2 Mean Square Error (MSE)

MSE is the average value of squared error between the input image $f(x, y)$ to the output image $g(x, y)$, where both of the images have the same values. Good MSE values are where the value near zero (*MSE* ≈ 0) [17]. Mathematically *MSE* value calculation using the equation

$$MSE = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} [f(x, y) - g(x, y)]^2 \quad (5)$$

where M, N are the width and height of the image, $f(x, y)$ is the initial image intensity use values/position (x, y) and $g(x, y)$ native to the image of the intensity value at the position (x, y).

2.4.3 Peak Signal to Noise Ratio (PSNR)

PSNR is a value of the ratio between the maximum image reconstruction results with square root value of *MSE* or equal to the value of *RMSE* [17]. For 8-bit image pixel, the maximum value is 255. The criteria of image quality will be better if the result of the greater *PSNR* values and mathematically generated from *MSE* value shown in the equation:

$$PSNR = 10\, log_{10} \left[\frac{255^2}{MSE}\right] dB \quad (3)$$

3. PROPOSED APPROACH

To solve the above problem, the necessary stages of completion as shown in Figure 1.

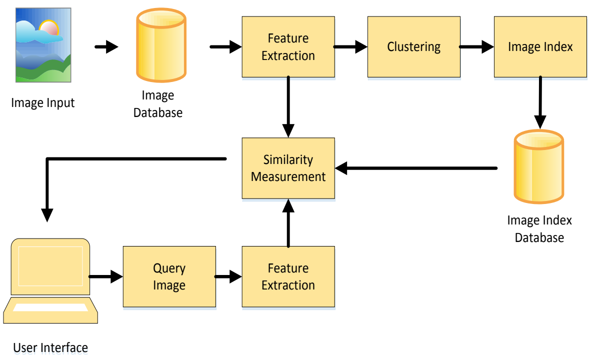


Figure 1. Architecture Systems

From the figure 1 can be explained the sequence of the process is

1. Put the whole picture of WANG database to a database of images. Using equation 1 did the color feature extraction process to obtain the value vector R, G, B based on a base image and calculate the difference in distance using the formula of HIT in equation 2 and 3.
2. Result values $S(H_q, H_i)$ was used to classify the image by using the k-means algorithm, Save the cluster for each image as the image index.
3. To perform image searches done by inserting the image of the user interface applications that have been created in the form of a query image. By using the image index was then determined the position of the cluster image and serve as a filter record by his cluster group.
4. Using equation 4 calculating the similarity between the query image and the image record.

4. RESULTS AND DISCUSSION

The database image used in proposed method is WANG The database, and the Delphi program has been implemented. The WANG database contains 1,000 images in JPEG format. The size of each image is 256x386 and 384x256. It consists of 10 classes such as (Africa, beach, monuments, buses, dinosaurs, elephants, flowers, horses, mountains and food) each class contains 100 images. Figure 2 shows the sample of WANG database.

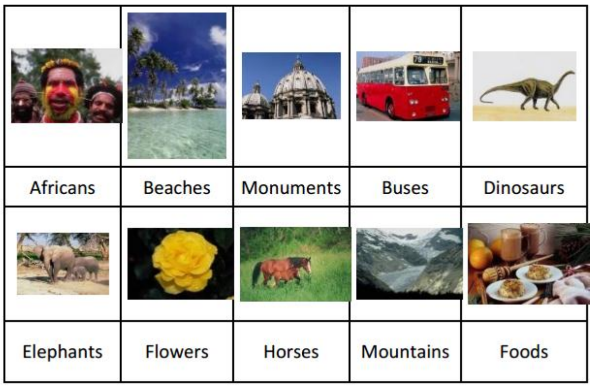


Figure 2. Example Image from each of the 10 Classes of WANG Database

The retrieval effectiveness of the proposed method is measured by using 30 different query images from each class. It is tested for 300 query images. In the beginning, the image size is changed to [128,128] to get a similar size for each image. The color features are extracted according to HSV histogram values. Euclidean distance is used to measure the similarity between query image and database images. Top 100 images are retrieved depending on the minimum distance. Next stage, the shape features are extracted from 100 images that resulted from the first stage and similarity measurements is performed between shape features of query image and the 100 images. Top 50 images are retrieved depending on the minimum distance. The final stage in the proposed method is extracted the first order features from 50 images and compared with the query image. The nearest ten images of the query image are retrieved from the image database.

4.1 The Result of K-Means

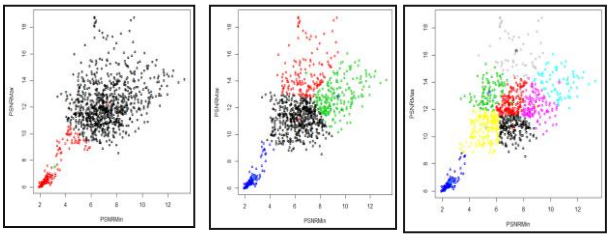
The implementation of the clustering was applied in several cluster groups; namely, 2 clusters, 4 clusters, 8 clusters, and the amounts of iteration of each cluster and the values of minimum PSNR and maximum PSNR for each cluster were shown in Table 1 and Figure 3.

Table 1. Clustering of 1000 WANG Database Records in 2, 4, 8 clusters

Cluster		Centroid PSNR (dB)		Record Count
Cluster	N	Minimum	Maximum	
2	1	7.238144	12.298757	838
2	2	3.217335	7.589639	162
4	1	6.318496	11.190640	529
4	2	6.894847	14.625369	155
4	3	2.638533	6.818268	119
4	4	9.449756	12.881852	197
8	1	10.119414	13.778683	110
8	2	8.000599	11.858669	209
8	3	5.688476	11.963560	188
8	4	6.923792	10.427970	147
8	5	7.305264	17.609415	24
8	6	4.749914	9.994102	95
8	7	6.782889	13.985441	121
8	8	2.518710	6.560355	106

Table 1 shows that the record is processed in WANG database of 1,000 records with the results of the process for the second cluster to cluster-1 has a value of 7.238144 dB PSNR minimum centroid and centroid PSNR maximum of 12.298757 dB with a record number of 838, while the cluster 2 has a value of 3.217335 dB PSNR minimum centroid and centroid PSNR maximum amounted to 7.589639 dB with an unprecedented number of 162. Likewise, for the formation of 4 and 8, as shown in Table 1.

Figure 3 shows the graph plots each record based on the minimum and maximum PSNR value on the formation of 2, 4, and 8 clusters using the K-Means algorithm database. Each cluster in the plot with a different color, for example in the formation of two clusters to cluster groups to plot-1 is shown in black, while for cluster 2nd plot shown in red. It is also shown at 4 and 8 clusters.



2 Clusters 4 Clusters 8 Clusters

Figure 3. The plot for K-Means Clustering in 2,4, and 8 clusters for WANG Database.

4.2 The Result of Precision

The precision and recall are used to measure the performance of retrieval. The recall is used to measure the system's ability to retrieve all the images that are relevant, while precision is used to measure the system's ability to extract only the images that are relevant. The equation of the recall and precision are illustrated in the following:

$$precision = \frac{Number\ of\ relevant\ images\ retrieved}{the\ total\ number\ of\ images\ retrieved} = \frac{N}{N+B} \quad (6)$$

$$recall = \frac{Number\ of\ relevant\ images\ retrieved}{the\ total\ number\ of\ relevant\ retrieved} = \frac{N}{N+C} \quad (7)$$

Where the number of retrieved images is represented by N and the irrelevant images are represented by B, while the C accounts for the number of relevant images not retrieved. The similarity measurement used in this work is Euclidean distance. Table 1 shows the results of precession when using color, shape and texture features. The best results are obtained in using cascade color, form and texture features.

Table 1 shows the results of precession value group by categories obtained by my proposed method (histogram color feature with 5 clusters) with other retrieval systems: Histogram color, GLCM (Gray-level co-occurrence matrix) texture, histogram color + GLCM texture, and histogram color + GLCM texture with sub-block. The best results are obtained in using histogram color feature with 5 clusters.

Table 2. The result of precession value group by categories obtained by proposed method with other retrieval systems

No	Categories	Histogram Color	GLCM Texture	Hist.Clr + GLCM Texture	Hist.Clr+ GLCM Texture +sub-block	Histogram Color with 5 Clusters
1	Africa	0.36	0.21	0.34	0.41	0.92
2	Beaches	0.27	0.35	0.21	0.32	0.42
3	Building	0.38	0.50	0.24	0.37	0.38
4	Bus	0.45	0.22	0.51	0.66	0.52
5	Dinosaur	0.26	0.29	0.39	0.43	1.00
6	Elephant	0.30	0.24	0.26	0.39	0.68
7	Flower	0.65	0.73	0.81	0.87	0.92
8	Horses	0.19	0.25	0.28	0.35	0.97
9	Mountain	0.15	0.18	0.20	0.34	0.25
10	Food	0.24	0.29	0.25	0.31	0.78
Average		**0.33**	**0.33**	**0.35**	**0.45**	**0.68**

In Table 5 shows that the average value of precision in a row are 0.33, 0.33, 0.35, 0.45 and 0.68. Value 0.68 is the greatest value of the proposed proposal.

Figure 4 shows the precession obtained by the different methods. It indicates that the histogram color feature with 5 clusters gave the best result in image retrieval.

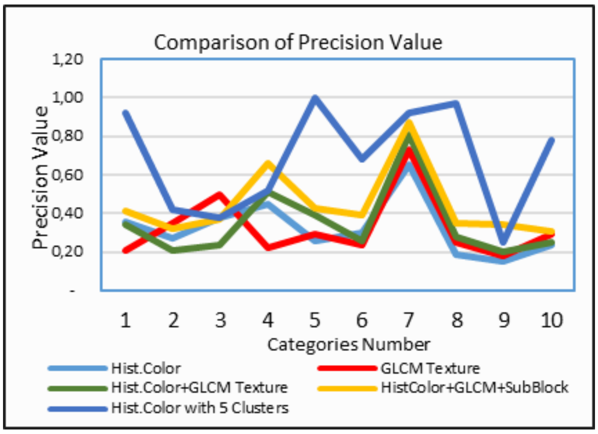


Figure 4. The Result of Precision Value Group by categories obtained by proposed method with other recovery systems

5. CONCLUSIONS

Cluster initialization process would be the main key in an information retrieval process of a query using PSNR minimum and PSNR Maximum. For each record in WANG database used as a base of record order, and then the distance between clusters was determined by computing total record divided by the number of clusters and ended by establishing each cluster that was taken from ordered records based on changes in their distances. The results in WANG database by using color histogram that was taken randomly by an amount of 1,000 showed that highest level of accuracy in 5 clusters shown on the precision value of 0.68.

6. REFERENCES

[1] Rejito J., Wardoyo R., Hartati S., Harjoko, 2012, A.,"Optimization CBIR using K-Means Clustering for Image Database," International Journal of Computer Science and Information Technologies, Vol. 3 (4), 4789-4793.

[2] Chin-Chin Lai, and Ying - Chuan Chen, ‖ A, 2011, “User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm," IEEE transactions on instrumentation and measurement, vol. 60, no. 10,

[3] Kannan, A, Mohan, V., Anbazhagan, N., 2010, “Image Clustering and Retrieval using Image Mining Techniques” IEEE Conference.

[4] Chakravarti, R, and Meng, X, 2009. “A Study of Color Histogram Based Image Retrieval," Sixth International Conference on Informational Technology.

[5] Kumar A.R, and Saravanan, D., 2013 “Content Based Image Retrieval Using Color Histogram," International Journal of Computer Science and Information Technologies, Vol. 4 (2), 242–245.

[6] Datta R., Joshi D., Li J., and Wang J. Z., 2008, "Image Retrieval," ACM Comput. Surv., Vol. 40, no. 2, Apr. pp. 1–60.

[7] Oliva A., and Torralba A, 2001, "Modeling the shape of the scene: "A holistic representation of the spatial envelope," Int. J. Comput. Vis., vol. 42, no. 3, pp. 145 –175.

[8] Jia Y., Huang C., and Darrell T., 2012, "Beyond spatial pyramids: "Receptive field learning for pooled image features," in Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), pp. 3370–3377.

[9] Tiwari A. and Bansal V., 2004, "PATSEEK: Content Based Image Retrieval System for Patent Database," Proceedings of international conference on electronic business, pp. 1167-1171.

[10] Krishnan N., Banu M.S., and Callins C., 2007, "Content Based Image Retrieval Using Dominant Color Identification Based on Foreground Objects," International Conference on Computational Intelligence and Multimedia Applications, Vol. 3, pp. 190-194.

[11] Ahmed H.A., Gayar N.E., Onsi H., 2008, "A New Approach in Content-Based Image Retrieval Using Fuzzy Logic" INFOS.

[12] Smith, J. R., 1997, "Integrated Spatial and Feature Image Systems: Retrieval, Analysis, and compression." Ph.D. thesis, Columbia University, New York, NY.

[13] Nikman T.A, Bakbak A., 2009, "An efficient hybrid approach based on PSO, ACO, and K-Means for Cluster Analysis," Elsevier.

[14] Ghosh A., Parikh J., Sangar V. and Haritsa J., 2002, "Query Clustering for Plan Selection, Tech Report," DSL/SERC, Indian Institute of Science.

[15] Kumar, M. Varun, M. Chaitanya V., and Madhavan, M., 2012. "Segmenting the Banking Market Strategy by Clustering." International Journal of Computer Applications 45.

[16] Ahmed H.A., Gayar N.E., Onsi H., 2008, "A New Approach in Content-Based Image Retrieval Using Fuzzy Logic" INFOS.

[17] Datta R., Joshi D., Li J., and Wang J. Z., 2008, "Image Retrieval," ACM Comput. Surv., Vol. 40, no. 2, pp. 1–60.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 410-416, 2017, ISSN:-2319–8656

Prediction of Phosphorus Content in Different Plants: Comparison of PLSR and SVMR Methods

Sushma D Guthe
Dept. Of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India

Dr.Ratnadeep R Deshmukh
Dept. Of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, India

Abstract: Phosphorus is one of the important biochemical components of plant organic matter and it helps to maintain the health of the plants. This study is conducted in some part of Aurangabad region, aimed to compare the PLSR and SVMR methods for predicting the phosphorus (Cp) content available in leaves in different plants using spectroscopy in Vis-NIIR reflectance spectroscopy. A total 38 leaf samples taken from jawar and maize plants were collected. In the plant 35% to 40% phosphorus content were found.

Keywords: Spectroscopy, phosphorus content (Cp), leaf, Vis-NIR, fieldspec4

1. INTRODUCTION

Plant organic matter such as nitrogen phosphorus and potassium are important biochemical components for metabolic processing. Therefore it is important to estimate the biochemical components. Earlier the analysis can be done using chemical method gives the accurate result but the chemical method is very much time consuming and complex.
The use of imaging spectroscopic analysis methods to estimate the nutrient status of maturing crops may save time, and scale back the value related to sampling and analysis. Imaging spectrographic analysis is a technology that result in getting data in narrow (<10 nm) and contiguous spectral bands. These narrow spectral bands allow the detection of some spectral features that masked within the broader bands of the multispectral scanner. [1]
The correct determination of phosphorus content in soils and plants is extremely important for agricultural science and practice. Phosphorus participates in a number of processes determining the growth, development and the productivity of the plant: formation of cell nucleus and cell multiplication, synthesis of lipids and specific proteins, transmission of hereditary properties, breathing and photosynthesis, energy transmission from richer to poorer energetic compounds, etc. It is very important to know the phosphorus status of soils to determine the necessity for phosphorus fertilizer use. Phosphorus is absorbed by the plants from soil in the form of phosphate ions. Phosphorus is a constituent of cell membranes, certain proteins, all nucleic acids and nucleotides, and is required for all phosphorylation reactions.[2] Phosphorus is known to be unique for its sensitivity and stability as a human activity indicator. Its content in soil represents a great interest to archaeologists, giving them information on type and intensity of human activity. Total phosphorus measurement gives quantitative results in contrast to mobile phosphorus and represents the best indicator for variation caused by human activity. [3]
L.K. Christensen used partial least square regression. PLSR was used on continuous spectra in the range 400-750 nm and the total N and P contents were established through chemical analyses and used as references. He was predicted P content with 74% accuracy based on the canopy spectral reflectance. He investigates the spectral reflectance in the visual range as a potential indicator for estimating nitrogen and phosphorus content in spring barley at three early growth stages. [4]

To develop calibration equations, multiple linear regression, and partial least-squares regression (PLSR) were used. This study compares between MLR and PLSR in the spectral range 1100-2500 nm. [5] Agustin Pimstein used vegetation indices and PLSR. In that study he uses specific narrow band vegetation indices instead of tradition broad band indices. It was observed that a significant improvement is obtained when the mineral total content is considered instead of the relative content. Therefore it was suggested that the biomass should also be retrieved from the spectral data. [6]
Yanfang Zhai study was based on eight different plants. He used the PLSR and SVMR. For implementation of PLSR used the parles 3.0 software and for SVMR uses the LIBSVM library. For some crop, he used the PLSR and for other SVMR. He concluded that the SVMR method combined with Vis-NIR reflectance has the potential to estimate the contents of biochemical components of different plants. [7]
Liu Yanli studied that upper side or lower side of leaves give the better spectral signature. The linear of PLS model and nonlinear of LS-SVM model fit better with spectral data of the

upper side and lower side of leaves, respectively. He used 400-1000nm wavelength sensitive to phosphorus content. [8] Stepwise linear regression was used to select wavelengths to investigate relationships between laboratory analysis results and spectral data.400-900nm wavelength sensitive to phosphorus content. [9] The results of the experiment demonstrated that radiometric measurements can be used for monitoring of N, P, S and K status in a wheat crop. Correlation analysis of nutrient status with leaf and canopy reflectance showed presence of responsive wavelengths to variable N, P, S and K status in wheat. [10]

Osborne et al. used linear regression models that included reflectance at 730 and 930 nm for predicting P concentration in corn. [11]

Name	Algorithm	Elements	Spectral range	Plant
L.K. Chirstensen	PLS	Nitrogen and phosphorus	400-750nm	Spring Barley
C. petisco	MLR and PLSR	nitrogen, phosphorus and calcium	1100 to 2,500 nm	Woody plant species
Agustin Pimstein	Vegetation indices and PLSR	potassium and phosphorus	1400–1500nm and 1900–2100nm	Wheat
Yanfang Zhai	PLSR and SVMR	Nitrogen, phosphorus and potassium	350–730 nm and 1420–1800 nm	rice, corn, sesame, soybean, tea, grass, shrub, and arbour
Liu Yanli	PLS and LS-SVM	Nitrogen and phosphorus	400-1000nm	Citrus leaves

Table 1.1: different method that estimate biochemical content

2. MATERIALS AND METHODOLOGY

2.1 Study Area

The study area is Aurangabad city. Samples of leaf mainly collected from three areas Devlai, Waluj (19.850670, 75.263130) and BAMU University with jawar, maize crop. Aurangabad city is located within the Maharashtra.

2.2 Spectral Measurements and Pre-Processing of Leaf Samples

A total 38 samples were collected between 24 to 30 march 2017, from two types of plants including jawar and maize between 1pm to 5pm. The sampling sites were randomly selected based on the land areas of these plants. The leaves without leaf stalk on upper layer of canopy were cut and kept fresh in plastic bag for less than 8 hours before spectral reflectance were measured. The spectral reflectance of leaf samples measured using a Fieldspec4 spectroradiometer. [12] The spectroradiometer has a spectral range from 350 to 2500nm. It's sampling interval 3nm (from 350 to 1000nm) to 10nm (from 1000 to 2500nm). For illumination purpose a 50 W quartz halogen lamp was set over, the leaf samples at a distance of 30cm and a distance between lamp and gun is 37 cm. The samples were put evenly on black sheet paper. The white reflectance panel was used to optimize the signal and calibrate the accuracy and detector response. Each sample was scanned 10 times within 180° rotation. Taking the mean of these 10 readings for further processing.to remove the noise values below 400nm and above 2400nm were rejected.

Before predicting the models several preprocessing transformation techniques such as first derivative of original reflectance and reflectance transformation. Data visualization and preprocessing can be done using ViewSpec pro software (ASD Inc). [13] The prediction accuracy of the model for the calibration and validation datasets was been evaluated through parameters such as R2 (Coefficient of Determination) is the correlation between predictable and observable variables, RMSE (Root Mean Square Error) provides direct assessment of modelling error expressed with original measurement unit and RPD(Residual Prediction Deviation).[14]

$$R^2 = \sum_{i=1}^{N} \frac{(\hat{y} - \bar{y})^2}{(y - \bar{y})^2}$$

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y} - y)^2}{N}}$$

$$RPD = \frac{\sigma_{val}}{RMSEv}\sqrt{n/n-1}$$

Where
$\hat{y}$is the predicted value,
y is the observed value,

$\bar{y}$ is the mean of observed values,
N is the number of sample data.

2.3 Methodology

In this study partial least square regression and support vector machine regression these two methods used. These methods are as follows:

2.3.1 Plsr

Partial least squares (PLS) is a method for constructing predictive models when the factors are many and highly collinear. Note that the emphasis is on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables. [15]The PLSR method is used to find the hyperplanes of maximum variance between predictable and observable variables and develops a linear model by projecting predictable and observable variables to a new space. [16][17]

$$Y = X\beta + \varepsilon s$$

where **Y** is the vector of predictable variables (biochemical component contents in this study), **X** is the matrix of observable variables, which is a linear combination of a few latent potential factors (spectral reflectance in this study), β is the matrix of regression coefficients, and ε is the error matrix of the relationship between **X** and **Y.**

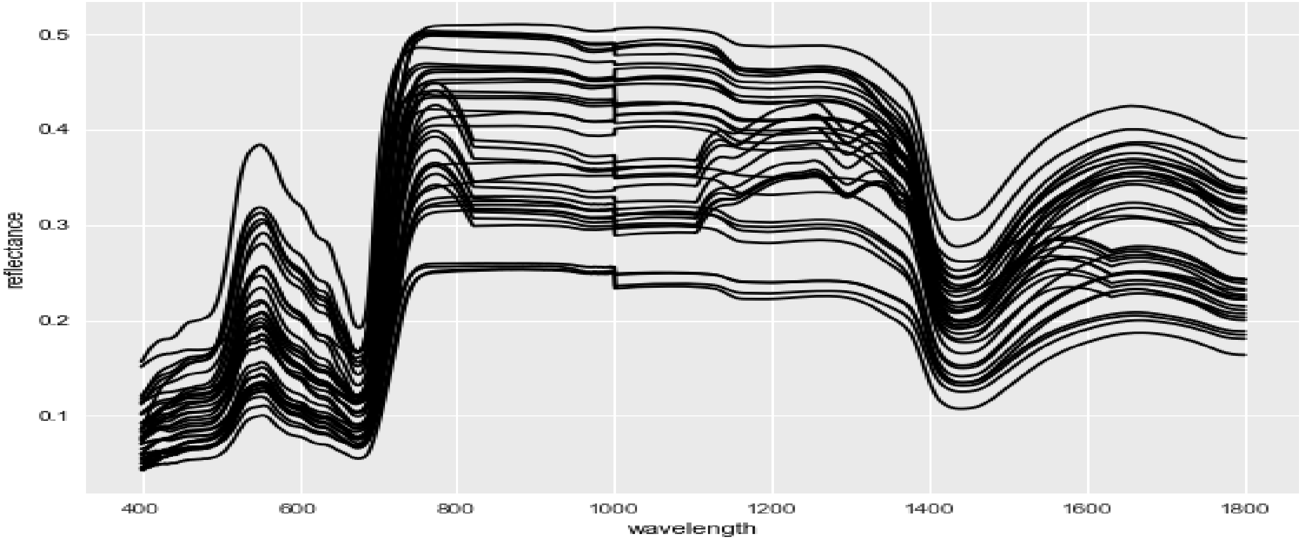
2.3.2 Svmr

The SVMR function can be expressed as [18]

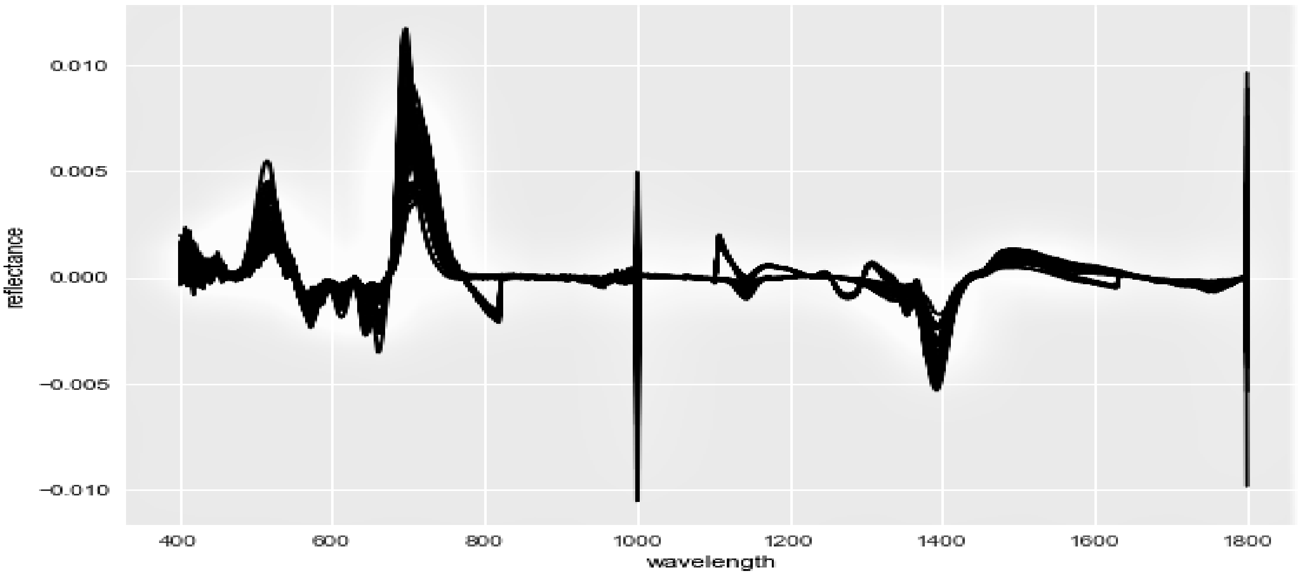
$$f(x) = \sum_{i=1}^{M} \text{aiyiK}(\text{xi}, x) - b\ 0 \le \text{ai} \le C$$

where **K**(**xi**, **x**) is the kernel function, **xi** is the input vector, **x** is an item used to create higher-dimensional feature space, **yi** is the corresponding output vector, **ai** is the Lagrange multiplier (called the support value), M is the number of samples, and **b** is the bias term. [19] The radial basis function (RBF) [20] was employed as the kernel function in this study:

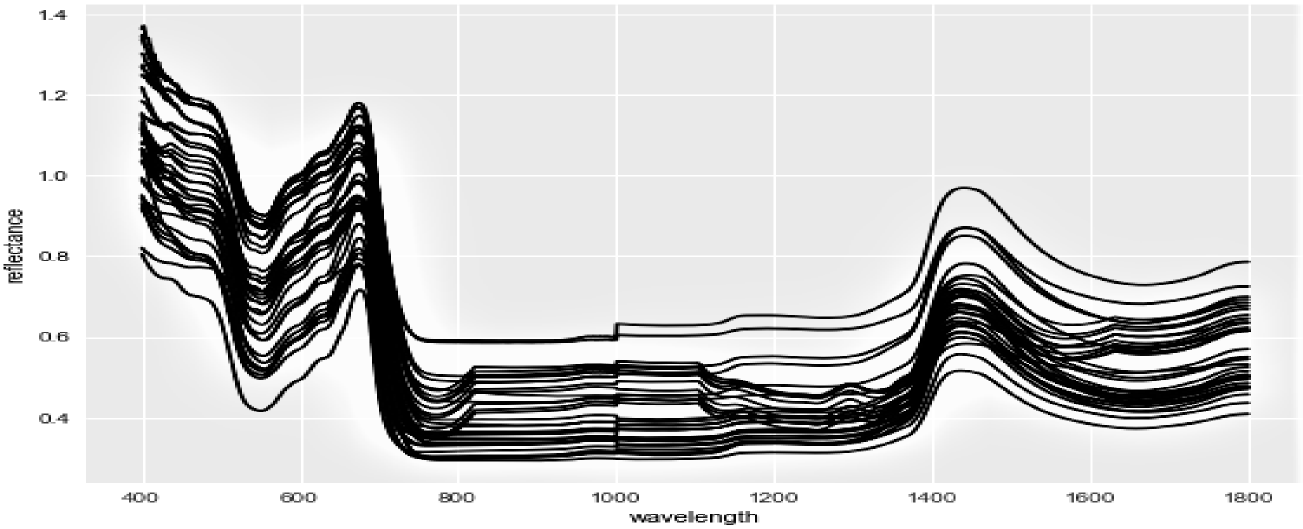
$$\boldsymbol{K}(\boldsymbol{x_i}, \boldsymbol{x}) = \boldsymbol{e}^{-y(x_i - x)^2}$$



(a)



(b)



(c)

Figure 2.1: Average Spectral Reflectance curves (a) original spectral reflectance, (b) the first derivative of reflectance and (c) Reflectance transformation of the reflectance

2.4 Data Analysis

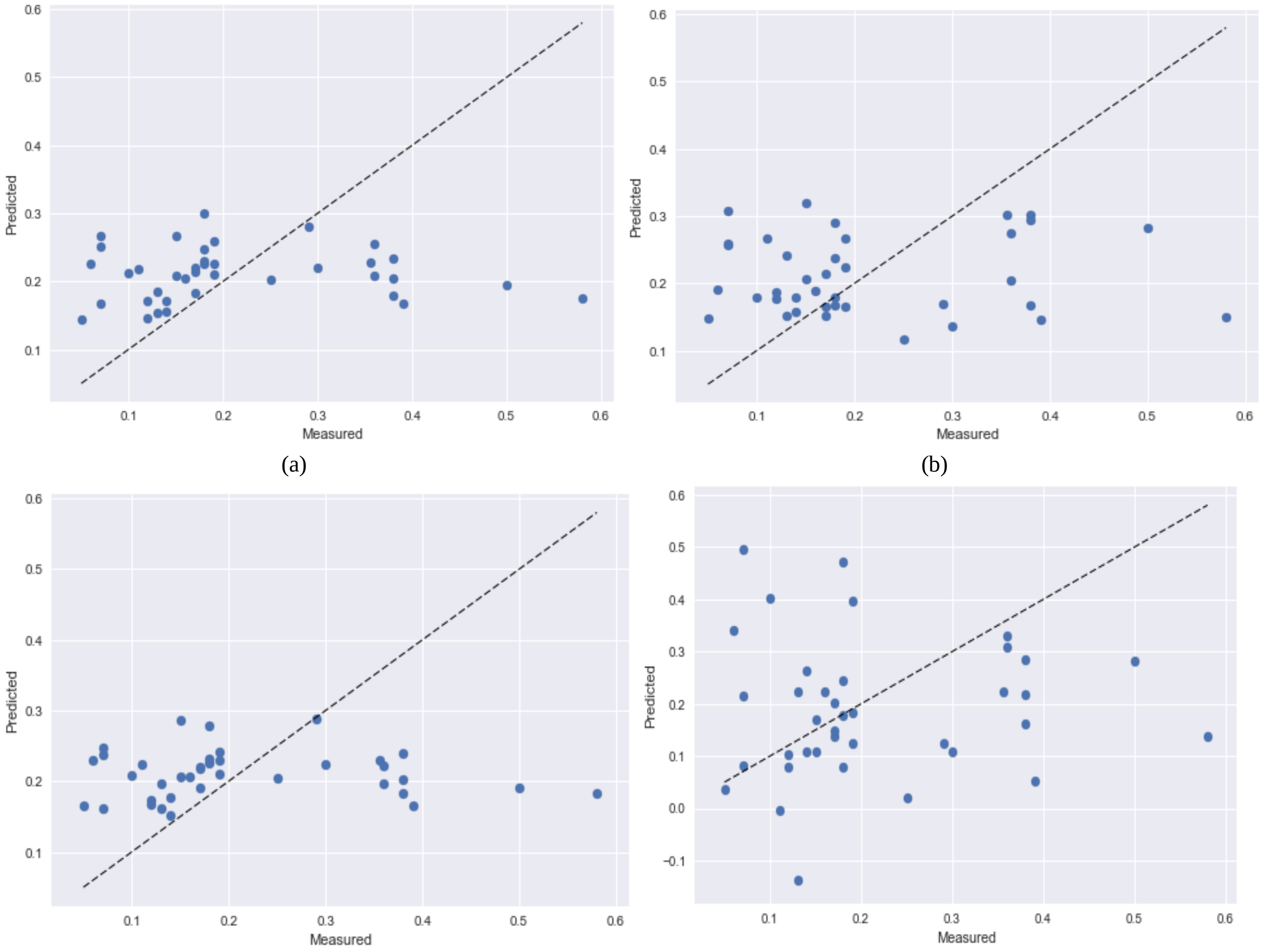
In this study, the PLSR and SVMR models for leaf sample was implemented within the Anaconda2 (32 bit) software used. [21] This software provides all packages that are required to implement model. In Anaconda2 Spyder 3.1.2 environment was used for python programming .this environment used the python 2.7.13 version. Wavelength range between 400-1000nm we sensitive to phosphorus content

3. RESULTS AND DISCUSSION

The average raw reflectance spectra of leaves of various plants had similar patterns as shown in figure 2.1. The descriptive statistics of Cp of the 38 leaf samples are shown in table 3.1. The maximum value is almost ten times the minimum value of Cp.

	Mean	SD	Median	Max	Min
Cp(%)	0.25	0.14	0.24	0.50	0.04

Table 3.1: Descriptive statistics of phosphorus (Cp) content in 38 leaf samples

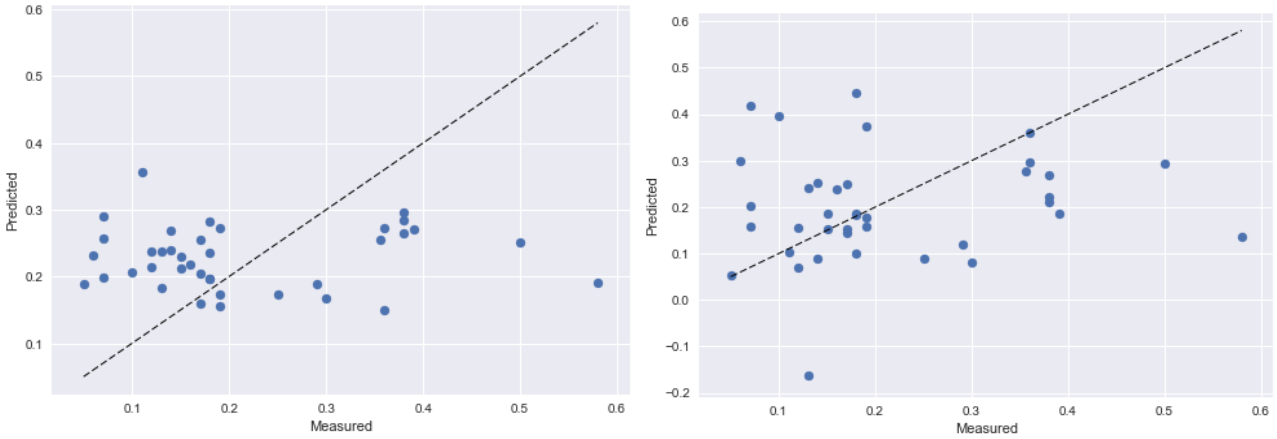


(a)

(b)

(c)

(d)



(e) (f)

Figure 3.1: correlations between measured and estimated values of phosphorus contents in Partial Least Square Regression (PLSR) (a) original reflectance (b)1st derivative (c)reflectance transformation and Support Vector Machine Regression (SVMR) (d) original reflectance (e) 1st derivative (f) reflectance transformation

Pre processing tech	RMSE		R^2	
	Cal	Val	Cal	Val
Original reflectance	0.0793	0.0920	0.0573	0.1072
First derivative	0.1331	0.0926	0.2050	0.0947
Log(1/R)	0.1456	0.0935	0.0477	0.0780

(a)

Pre processing Tech	RMSE		R^2	
	Cal	Val	Cal	Val
Original reflectance	0.1449	0.0776	0.7172	0.3649
first derivative	0.0889	0.08095	0.6450	0.3085
Log(1/R)	0.0818	0.07865	0.6995	0.3473

(b)

Table 3.2: the results of model calibration and validation (a) PLSR (b) SVMR

In the PLSR model, considering raw spectral reflectance it gives poor performance for this model. When we done the preprocessing it gives better results as shown in the table 3.2(a). Log(1/R) preprocessing method gives the poor performance (R^2_v =0.0780) while first derivative give better performance than Log(1/R). In the SVMR model, considering raw spectral reflectance it gives poor or moderate results when estimating Cp(R^2_v=0.365). Preprocessing method results as shown in table 3.2(b).

The SVMR model gives the better results when estimating Cp as compared to PLSR model.

4. CONCLUSION

In this study, we compared PLSR and SVMR models for estimating the phosphorus content of leaves of two different crops with laboratory based Vis-NIR reflectance data. We concluded that the SVMR method after applying preprocessing method estimates biochemical content of different plants.

ACKNOWLEDGEMENT

This work is supported by Dept. of Computer Science and Information Technology under the funds for Infrastructure under science and Technology (DST-FIST) with sanction no. SR/FST/ETI- 340/2013 to Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India. The authors would like to thank Department and University Authorities for providing the infrastructure and necessary support for carrying out the research.

REFERENCES

1. Sushma D Guthe, Ratnadeep R Deshmukh, May-June 2017, “Estimation of Phosphorus Content in Leaves of Plants using PLSR and SVMR: A Review”, IJARCS, Volume 8, No. 4.
2. NCERT Solutions - Biology for Class 11th, MINERAL NUTRITION
3. Krasimir Ivanov, Penka Zaprjanova, Violina Angelova, Georgi Bekjarov and Lilko Dospatliev, August 2010, “ICP determination of phosphorous in soils and plants”, © 2010 19th World Congress of Soil Science, Soil Solutions for a Changing World.
4. L.K. Christensen, B.S. Bennedsen, R.N. Jorgensen, H. Nielsen, May 2004, “Modelling Nitrogen and Phosphorus Content at Early Growth Stages in Spring Barley using Hyperspectral Line Scanning”, Biosystems Engineering, Volume 88, Issue 1, Pages 19–24.
5. C. Petisco , B. Garcı´a-Criado , B. R. Vxa´ zquez de Aldana, 2005, “Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species”, Springer-Verlag , Anal Bioanal Chem (2005) 382: 458–465.
6. Agustin Pimstein, ArnonKarnieli, Surinder K. Bansal, David J. Bonfil, February 2011, “Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy”, Field Crops Research Volume 121, Issue 1, Pages 125–135
7. Yanfang Zhai , Lijuan Cui , Xin Zhou , Yin Gao , TengFei and WenxiuGao, April 2013, “Estimation of nitrogen, phosphorus, and potassium in the leaves of different plants using laboratory-based visible and near-infrared reflectance spectroscopy: comparison of partial least-square regression and support vector machine regression methods”, International Journal of Remote SensingVol. 34, Issue 7, 2502–2518.
8. Liu Yanli, LyuQiang, He Shaolan, Yi Shilai, Liu Xuefeng, XieRangjin, ZhengYongqiang, Deng Lie , April 2015, “Prediction of nitrogen and phosphorus contents in citrus leaves based on hyperspectral imaging”, Int J Agric&BiolEng, Vol. 8,Issue 2 80－88.
9. Y. Özyigit, and M. Bilgen, *2013*, “Use of Spectral Reflectance Values for Determining Nitrogen, Phosphorus, and Potassium Contents of Rangeland Plants”, J. Agr. Sci. Tech. Vol. 15: 1537-1545.
10. G. R. Mahajan, R. N. Sahoo, R. N. Pandey, V. K. Gupta, Dinesh Kumar, 2014, “Using hyperspectral remote sensing techniquesto monitor nitrogen, phosphorus, sulphur and potassium in wheat (Triticum aestivum L.)”, Precision Agric , 15:499–522.
11. Osborne, Schepers, Francis, & Schlemmer, 2002, “Detection of phosphorus and nitrogen deficiencies in corn using spectral radiance measurements”, Agronomy Journal, 94, 1215–1221.
12. Fieldspec4 user manual, ASD Document 600979, Rev. A, December 2011.
13. ViewSpec Pro™ User Manual, ASD Document 600555 Rev. A © 2008 by ASD Inc.
14. Vasques, G. M., S. Grunwald, and W. G. Harris, 2010, “Spectroscopic Models of Soil Organic Carbon in Florida, USA”, *Journal of Environmental Quality* 39: 923–34.
15. Randall D. Tobias, 1995, “An Introduction to Partial Least Squares Regression”, SAS Institute Inc., Cary, NC.
16. Efron, B., and G. Gong, 1983, “A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation”, The American Statistician 37: 36–48.
17. Lecture, Modeling Input/output Data: Partial Least Squares (PLS), cited on www.sbcny.org/pdfs.
18. Zhang, X. J., and M. Z. Li, 2008, “Analysis and Estimation of the Phosphorus Content in Cucumber Leaf in Greenhouse by Spectroscopy.” *Spectroscopy and Spectral Analysis* 28: 2404–8.
19. Zhang, X. G (2000), “Introduction to Statistical Learning Theory and Support Vector Machines”, Acta Automatica Sinica 26: 32–42.
20. Burges, C. J. C (1998), “A Tutorial on Support Vector Machines for Pattern Recognition.” *Data Mining and Knowledge Discovery* 2: 121–67.
21. “Anaconda software Download 32-bit” cited on https://www.continuum.io/download