International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 64-69, 2012

Healthcare Web Services by E-Governance

Rajan Datt
Institute of Technology,
Nirma University,
Ahmedabad, India

Priyanka Tripathi
Institute of Technology,
Nirma University,
Ahmedabad, India

Abstract: As India is one of the fastest developing countries in the world, it is important to improve the quality of our health maintenance management and preventive medical care to extend healthy life expectancy. Today's scenario for health care in Indian e governance is in the limit of contacting 75 hospitals of the ISRO Telemedicine network only. Whilst this is currently working best of it, the limitation of this can be retarded by introducing the health care web services to each individual of the country. We believe advanced implementation of Information and Communications Technologies (ICT) may improve the medical services and health maintenance management. As medical science is fast developing and information resource is pouring in, there is urgent need for dissemination knowledge by interlinking primary, secondary and tertiary level health centers by ICT applications. This will help health personal to deliver high quality services. Moreover, IT systems have been built to support different work flows in the health sector, but the systems are rarely connected and have become islands of data. From 2006 onwards corporate IT giants are experimenting for ICT application in health sector both in Government and private hospitals. In this paper, we discuss the potentialities and expansibility of the XML Web Services based on the Adaptive Collaboration (AC) which can be aggregated by the Indian e governance system as a health care web services. We would like to present ways of improving health maintenance service and regional medical services. In order to realize better health maintenance and prevention of disease, we would like to prove that incorporating medicine, life, and work through the XML Web Services is highly effective.

The developed system is using data agent concept in transferring the format of information from different medical database systems to be an international standard format of metadata known as HL7 v3.0 using XML based cloud services called the Medical Cloud system which can take advantage of the Indian cloud revolution.

Keywords: Directory Services, Interoperability, HL7, Healthcare services, E Governance

1. INTRODUCTION

Most of the health information systems today are proprietary and often only serve one specific department within a healthcare institute resulting in difficult interoperability problems. To complicate the matters worse, a patient's health information may be spread out over a number of different institutes which do not interoperate. This makes it very difficult for clinicians to capture a complete clinical history of a patient. [1]

The benefits of utilizing the XML Web Services are the following: [15]

1. It is platform independent therefore it is usable regardless of the type of hardware and software,
2. the connection is highly flexible, collaborative, and compatible with other systems,
3. It avoids overlapping investments of the ICT utilization and development
4. It enables the sharing of the ICT sources, and
5. It offers more flexibility in data process and exchange.

Introducing Web services to the healthcare domain brings many advantages:

1. It becomes possible to provide the interoperability of medical information systems through standardizing the access to data through WSDL [2] and SOAP [3] rather than standardizing documentation of electronic health records.
2. Medical information systems suffer from proliferation of standards to represent the same data. Web services allow for seamless integration of disparate applications representing different and, at times, competing standards.
3. Web services will extend the healthcare enterprises by making their own services available to others.
4. Web services will extend the life of the existing software by exposing previously proprietary functions as Web services.

However it has been generally agreed that Web services offer limited use unless their semantics are properly described and exploited [4-7].

Evidence based clinical practice needs sufficient knowledge [9] on latest development in medical science. Automated information management tools like internet, web based libraries, CME, Electronic Medical Record (EMR), Electronic Health records (EHR), and computerized prescriptions are important components. [10]

1.1 E-Governance

E-Governance is the application of Information and Communication Technology (ICT) for delivering government services, exchange of information communication transactions, integration various stand-one systems and services between Government-to-Citizens (G2C), Government-to Business(G2B),Government-to-Government(G2G) as well as back office processes and interactions within the entire government frame work.[1] Through the e-Governance, the government services will be made available to the citizens in a convenient, efficient and transparent manner. The three main target groups that can be distinguished in governance concepts are Government, citizens and businesses/interest groups. In E-governance there are no distinct boundaries. [13]

1.2 Privacy vs. Safety

Health care records often contain sensitive data, which could potentially harm a person's reputation or private life, should it be exposed to unauthorized people. More seriously, though, these records are the basis on which a patient receives care, and errors caused by negligence, malicious intent, or the like can potentially cause physical harm. [8]

For these reasons, health care records are surrounded by security measures. Ensuring the confidentiality of information while in transit from one practitioner to the next, and while being stored, is imperative to avoid eavesdropping by unauthorized individuals.

Thus organizations that handle sensitive data and the authorized personal who are given the right to access those data should be bounded by law [11 - 12] to ensure that only authorized staff gains access. [8] Moreover the system should have proper, faster and simpler authentication measure.

2. ARCHITECTURE

2.1 Security architecture

Various components of this architecture are

1. A trusted system Security Token Service(SSTS) having a predefined maximum limit of validity
2. Web Services Client (WSC) which are the client computers from where the authenticated personals can refer the system
3. Web Services Providers(WSP) are the provider systems which provides the web services in demand
4. SAML tokens

Security Assertion Markup Language (SAML) is an XML-based open standard for exchanging authentication and authorization data between security domains, that is, between an identity provider (a producer of assertions) and a service provider (a consumer of assertions). SAML is a product of the OASIS Security Services Technical Committee. [14]

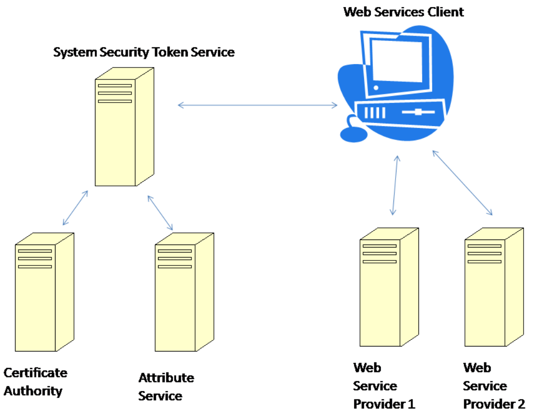


Figure 1 Security and Authentication Architecture

Following steps can be followed for authentication process of the authorized personnel on the system providing the private and secure data about the patients and other details.

1. The user to the system by logging in from a general Web Service Client (WSC). This Client then builds an SAML with attributes and credentials of that user.
2. The System Token Services(SSTS) checks that
   a. the credential of the WSC system is valid
   b. the Web Service Provider(WSP) system certificate is valid and not revoked
   c. the user’s credential is valid
   d. the user’s certificate is valid and not revoked
3. The SSTS now seeks to verify that the client-specified core attributes are valid by using backend attribute services. Some of these verified attributes are cached for a short period for optimization purposes.
4. If everything is OK, the security token is digitally signed by the SSTS and returned to the WSC.
5. The security token can now be used in interactions with different WSPs until it expires.
6. Upon receipt, the WSPs validate the security token by verifying the SSTS credentials and leverage the embedded attributes for logging and authorization.
7. Finally a result, i.e. business information or an error is returned.

2.2 Service Oriented architecture for healthcare

There are standards that expose the business logic in the healthcare domain such as HL7 [16], which use the messaging technique.

Electronic Healthcare Record (EHR) based standards such as CEN TC251 [17], ISO TC215 [18] and GEHR [19], on the other hand, define and classify clinical concepts that make up the patient records. Such standards offer significant value in developing ontologies to express the semantics of Web services.

But HL7 events are usually very complex containing innumerous segments of different types and options. Moreover the party invoking the Web service must be HL7 compliant. All or some of this data may be coming from different systems that do not interoperate. This in turn, creates the need to retrieve these partial results probably through finer granularity Web services.

In order to define the granularity of Web services, we can refer to Electronic Healthcare Record (EHR) based standards from major standard bodies like CEN and GEHR. These standards define metadata about EHR through “meaningful components”. [1]

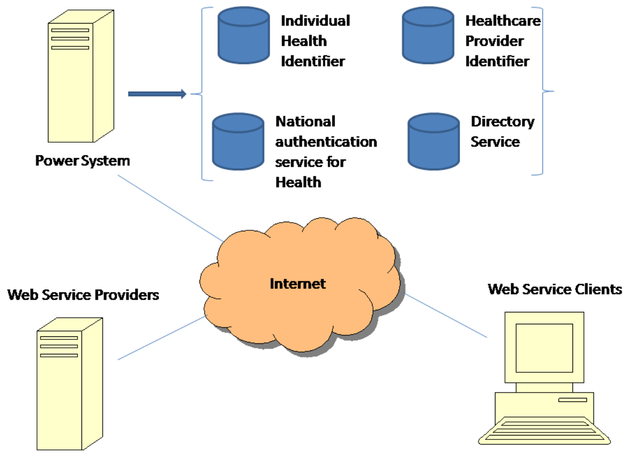

Individual Health Identifier
Healthcare Provider Identifier
National authentication service for Health
Directory Service
Power System
Internet
Web Service Providers
Web Service Clients

Figure 2 Service Oriented Architecture (SOA)

Generally, health information is stored over a number of different WSP. A national Power system must be available for the provision of directory services to determine the distributed locations of the source systems holding the related health records. Our proposal [20] addresses this need by defining a model to support secure communications between healthcare providers and the Power System in the national e-health environment as shown in Figure 2.

Proposed architecture defines the required constructs to share and transfer healthcare information securely between healthcare providers and the authorized national Power System. This architecture proposes that the
Power System should be built on a high trust computer platform.

Since the Power System is itself a critical application under any operating system, so Power System must be protected from even internal threats through the use of modern "flexible mandatory access control (FMAC)" structures. Under such an operating system, and as distinct from the less secure "discretionary access control (DAC)" systems, even a systems manager may not have permission to access the health record data. In simple terms, in these systems there is no "super-user" capable of obtaining access to all system resources at any time. If an individual name server system is "captured", propagation of exposure will not extend beyond the compromised application itself, a vital concern in any e-health record indexing structure. Such systems exist and are commercially available, e.g. the "Secure LINUX (SELinux)" [21] systems, "Solaris/SE" [22] system, etc.

2.2.1 Power System

The load of the national Power System should be relatively lightweight to perform e-health indexing services efficiently. This can mitigate the Power System explosion and traffic bottleneck risks. Such an approach is favorable in a geographically large country such as India. To maximize the efficiency of the indexing services, the proposed Power System does provide network connectivity services, messaging translation, addressing and routing functions and extensive logging of all message access. These services can be performed at the level of the WSP, which is detailed in Section 2.2. The access control and authorization process is best performed close to where the source system is, as each healthcare service provider might implement the service differently based on its own WSP access requirements. There are no centralized network provisions to handle peer-to-peer communications; each service must manage its own interface to the network.

The Power System will be a centralized facility run at a national level. It is envisioned that the directory service is devised in the context of a DNS, which uses hierarchical distributed database architecture. Our proposed national Power System performs common and fundamental functionalities including:

- Identification and authentication, and
- Directory services.

2.2.1.1 Identification and authentication

Identification and authentication services is same as the security architecture given in Section 2.1

2.2.1.2 Directory services

The Directory Service is one of the fundamental services in national e-health infrastructure. Since healthcare data are located at various places, directory services are used to identify and locate the available information. The Directory Service in the Power System provides a mechanism for obtaining the necessary information for invoking a service. This information contains the network location of the service, the digital certificate required to use it and other information required to invoke the service. It is envisaged this will be specified in Web Services Description Language [23] (WSDL) format, which equates to Service Instance Locator (SIL)

2.2.1.3 Operation of the Directory Services

The service patterns can be divided into two broad categories: synchronous and asynchronous services. A synchronous service occurs in direct response to a request. An asynchronous service has no relationship between the events. For example, to request a specific individual's health records is a synchronous service. To send out a discharge summary report to a healthcare provider is an asynchronous service.

With a synchronous service, when interacting with the directory service the requesting entity will provide proof of their identity and the IHI associated with the records they are requesting. Once the requester has been authenticated by the Power Server, it will respond with the following:

1. A signed token attesting to the identity of the requester ({token}SignIS_PrivKey) and
2. A list of service instances containing health records for the person identified by the IHI (Service_Instance_1,...,Service_Instance_N).

The entire response is signed so that the requester can be assured that it is a legitimate response from an authorized Power System and that any alterations to the response will be detectable. The confidentiality of both the requester and the individual identified by the IHI is maintained.

The token is signed independently of the entire response in order that it can be reused with requests to each service instance. The full response is depicted in Figure 3.

{{token}SignIS_ PrivKey,Service_Instance_1,..., Service_Instance_N}EncryptHPI-O_PubKey

Figure 3 Service Instance Response Message Format

The service instance information contained in the response identifies the target system location and information necessary for securely invoking that service.

This may include, but will not be limited to the credentials certificates required to access the service. The signed token provided in the Power System response may be the only credential required, in which case the effort expended by the Power System in authenticating the requester is reused. It is, however, conceivable that additional authentication may be required by a given service instance. For example, the requester may need to prove that they are a member of a given practice or college of medical practitioners.

With an asynchronous service, such as when a discharge summary message needs to be sent to the patient's primary healthcare provider, the healthcare provider issuing the summary queries the Power System for WSP, location and the digital certificate, credentials and then signs and encrypts the discharge message prior to transmission.

2.2.2 WSP

2.2.2.1 Peer-Entity Authentication

Many proposals are only concerned with the authenticity of the requesting entity (i.e. one-way authentication) but fail to address the importance of two-way authentication. Proposed architecture provides a mutual peer-entity authentication service complying with the ISO 7489-2. To authenticate the authenticity of the Power System, the service requesting entity must validate the certificate of the Power System. Once the authenticity of the national Power System is assured, the Power System authenticates the identity of the healthcare service requesting entity. In this sense, the authentication service of the Power System acts as a notarization mechanism in line with the philosophy of peer-entity authentication stated in ISO IS7498-2.

2.2.2.2 Provision of Data Protection

As various healthcare organizations may have their own specific access authorization requirements and processes, access authorization is best performed where the resource system is located. Once the requesting entity's identity is authenticated, the request of particular healthcare information is presented to the target service provider.

The HIP of the target service provider will provide the verified identity and the profile of the requester to the authorization logic unit to perform access decision making. The authorization decision depends upon the requesting entity's profile and defined privilege management policy. The implementation of the authorization logic unit is based on the "Sensitivity Label" function.

2.2.2.3 Interoperability Platform

Health Level 7 (HL7) 4 can be used as the national standard for the electronic exchange of health information. WSP provides an interoperability platform by incorporating an HL7 Interface Engine and Message Mapping Sets conforming to the HL7 v3.0 Message Standards for healthcare information exchange. HIP also incorporates an HL7 Interface Engine and Message Mapping Sets for messaging Interoperability.

HL7 Interface Engine

Any non-HL7-compliant data contents are translated into the HL7 standard format (XML-based data structure) by the HL7 Interface Engine prior to information transmission. The HL7 Interface Engine contains a set of mapping algorithms to map data contents with an appropriate HL7 Message Template to generate an HL7 message.

Message Mapping Sets

The Message Mapping Sets contain a repository of HL7 Message Templates for various clinical and administrative messages. Each set provides one HL7 Message Template to serve for one clinical or administrative message. Message Mapping Sets will be designed and developed to meet the current healthcare service needs and will be imported into WSP. The HL7 Message Template guides and directs data contents to form an HL7 message.

HL7 Clinical Document Architecture (CDA)

HL7 Clinical Document Architecture (CDA) provides a framework for clinical document exchange. WSP imports the HL7 message into a CDA document. This CDA document is also associated with an appropriate style sheet. The CDA document and the style sheet will be sent to the requesting entity through Web services. The requesting entity renders the received document with the style sheet in a human-readable form with a Web browser.

2.2.3 Key Information Flows

2.2.3.1 Peer-Entity Authentication Process

Follow the steps given in Section 2.1

2.2.3.2 Health Record Enquiry Process

1. The service request, containing the patient's IHI and requester's HPI-I, is sent to the Directory Services of the Power System to inquire which health providers hold the health records of the specific patient.
2. The Directory Services of the Power System responds with a token and a list of the service instance information for service invocation to the requesting entity. This token indicates the requester identity assertion to enable single sign on for service invocation.
3. The requester verifies the received information and then contacts each target service provider for service invocation. The requester sends the request including the token with other necessary information to invoke the service.

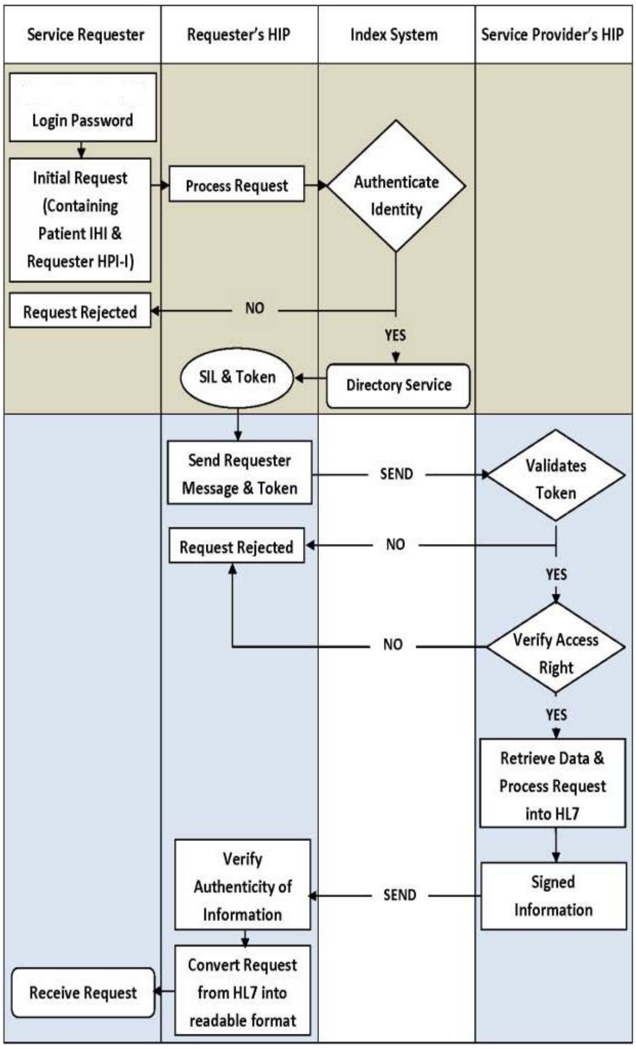


Figure 4 Information Flow

2.2.3.3 Verification and Authorization Evaluation Process

1. Each target service provider validates the request message containing the token and other necessary information for service invocation.
2. In turn, the request is passed to the authorization logic to make an access authorization decision based on the service requester's profile indicated in the ticket and any additional authorization attributes which are mutually agreed by the policy.

2.2.3.4 Provision of Requested Health Record Process

1. If the access is granted, the service provider extracts the health record from the data source.
2. The service provider processes the requested health record into the HL7 message format.
3. The target service provider sends the signed and encrypted information to the requester.
4. The service provider records the information access for auditing purposes.

2.2.3.5 3.5) Reception of Requested Health Record Process

1. The requested information arrives at the service requester's HIP.
2. The service requester's HIP verifies the information arrived and then extracts the requested information which is in HL7 message format.
3. The message must be presented in a human readable format. The representation of HL7 message is rendered and displayed to the requester

3. CONCLUSION

Many people recognize the need of improving the quality and efficiency of health maintenance management. In order to improve the quality of the healthcare management system, sharing information among individuals, patients, hospitals, clinics, medical institutes, and pharmacies is imperative.

XML Web Services enables many people to contact and stay in close touch with physicians and outside mental health professionals at any moment when necessary through network. Therefore, utilization of the XML Web Services would generate innovative ways for the people to maintain and improve their mental and physical health.

In this paper we have presented a healthcare system that uses the Service Oriented Architecture as a basis for designing, implementing, and deploying, managing and invoking healthcare web services. Healthcare requires modern solutions, designed and implemented with modern technologies that encourage healthcare professionals and patients to adopt new procedures that can improve the presentation and delivery of healthcare. Multimedia input and output, particularly graphics and speech, makes the system seem less computer-like and more attractive to users who are not computer-oriented.

This paper proposes following Basic perspectives:

Architecture proposed. A trusted architecture for the Power System which provides the critical solution to determine the locations of distributed health records. This Power System plays a vital role in the national e-health scheme for identification and authentication and directory services. The Power System, therefore, must be a high trust system running on a trusted platform; and

Authentication levels. Users and systems can be authenticated with different degree of certainty, depending on the credentials that the principal presents

Maximum performance. The number of requests/messages is minimized. When trust has been established and the user has logged in to the federation, the WSC and WSP communicate directly with no third party involved.

Presently ICT implementation in health services is in infancy but its further use in both medical education and healthcare industry will revolutionize the healthcare provided by Government hospitals, corporate sector. Finally good quality health care delivery at doorstep in low cost would safeguard national health leading to economic growth.

We believe that our proposal to apply the Web Services would make a substantial contribution to the healthcare and medical field to realize the patient-oriented services.

4. REFERENCES

[1] Dogac, G. Laleci, S. Kirbas, Y. Kabak, S. Sinir, A. Yildiz, Y. Gurcan," Artemis: Deploying Semantically Enriched Web Services in the Healthcare Domain", Software Research and Development Center Middle East Technical University (METU)

[2] Web Service Description Language (WSDL), http://www.w3.org/TR/wsdl

[3] Simple Object Access Protocol (SOAP), http://www.w3.org/TR/SOAP/

[4] S. A. McIlraith, T. C. Son,H. Zeng,"S emantic Web Services", IEEE Intelligent Systems, March/April 2001, pp. 46-53.

[5] S. A. McIlraith,T . C. Son,H. Zeng, "Mobilizing the SemanticWeb with DAMLEnaled Web Services",Semantic Web Workshop 2001, Hongkong, China.

[6] E. Motta, J. Domingue, L. Cabral,M. Gaspari,"I RS II: A Framework and Infrastructure for Semantic Web Services",2n d International Semantic Web Conference,Flor ida, USA, October 2003.

[7] M. Paolucci, T. Kawamura,T . Payne,K . Sycara, "Semantic Matching of Web Services Capabilities", in Proc. of Intl. Semantic Web Conference,S ardinia,Italy , June 2002.

[8] Esben Dalsgaard, Chair, SOSI steering committee Digital Health Denmark (SDSD), Kåre Kjelstrøm Solution Architect Silverbullet A/S Skovsgaardsvaenget, Jan Riis Solution Architect / Project Manager, "A Federation of Web Services for Danish Health Care"

[9] Lele R.D (2008) ," ICT in day-to-day Clinical Practice Postgraduate medicine" API and ICP 2008 Vol. XXII. pp. 3-9.

[10] Subash Chandra Mahapatra (Department of Medicine, MKCG Medical College, Berhampur, Orissa, India) , Rama Krushna Das (National Informatics Centre, Berhampur, Orissa, India) and Manas Ranjan Patra(Department of Computer Science, Berhampur University, Berhampur, Orissa, India), "Current e-Governance Scenario in Healthcare sector of India"

[11] Blobel B, Nerdberg R et al, Modelling privilege Management and access control, "International Journal of Medical Informatics", 2006, 75:597

[12] Han Song, Skinner Geoff et al, "A Framework of Authentication and Authorisation for e-Health Services" . 2006 ACM 1-59593546-0/06/0011 Pages: 105-6

[13] http://en.wikipedia.org/wiki/E-Governance

[14] http://en.wikipedia.org/wiki/Security_Assertion_Markup_Language

[15] Mayumi Hori & Masakazu Ohashi, "Applying XML Web Services into Health Care Management", 0-7695-2268-8/05/$20.00 (C) 2005 IEEE

[16] Health Level 7 (HL7), http://www.hl7.org

[17] CEN TC/251 (European Standardization of Health Informatics) ENV 13606, Electronic Health Record Communication http://www.centc251.org/

[18] ISO TC215, International Organization for Standardization, Health Informatics Technical Committee http://www.iso.ch/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=4720

[19] The Good Electronic Health Record, http://www.gehr.org

[20] Min Hui Lee, Zi Hao Ng, Jin Hong Foo and Weihao Li, Vicky Liu, William Caelli, Jason Smith, Lauren May, "A Secure Architecture for Australia's Index Based E-health Environment"

[21] http://docs.redhat.com/docs/enUS/Red_Hat_Enterprise_Linux/6/pdf/Security-Enhanced_Linux/Red_Hat_Enterprise_Linux-6-Security-Enhanced_Linux-en-US.pdf

[22] http://www.oracle.com/us/products/servers-storage/solaris/solaris11/overview/index.html

[23] WDSL is used for describing how to access the network services in XML format. More detail is available at http://www.w3.org/TR/wsdl#_introduction accessed 30/08/2009.

International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 70-74, 2012

Comparative Analysis of Different Techniques for Novel Class Detection

Patel Jignasa N.
Parul institute of Engineering & Technology,
Waghodia, Vadodara,
Gujarat, India

Sheetal Mehta
Parul institute of Engineering & Technology,
Waghodia, Vadodara,
Gujarat, India

Abstract: Data stream mining is the process of extracting knowledge from continuous data. Data stream can be viewed as a sequence of relational touples arrives continuously at time varying. Classification of data stream is more challenging task due to three major problems in data stream mining: Infinite length, Concept-drift, Arrival of novel class. Novel class detection in stream data classification is interesting research topic and researches available for concept drift problem but not attention on the Novel class detection. In this paper we have discussed various techniques of the novel class detection. And have also covered comparative analysis of various techniques for the same.

Keywords: Data stream, Novel class, Incremental learning, Ensemble Technique, Decision tree.

1. INTRODUCTION

Data mining is the process of extracting hidden useful information from large volume of database. A data stream is an ordered sequence of instances that arrive at any time does not permit to permanently store them in memory. Data mining process has two major functions: classification and clustering. Data stream classification is the process of extracting knowledge and information from continuous data instances. The goal of data mining classifiers is to predict the class value of a new or unseen instance, whose attribute values are known but the class value is unknown [1].Classification maps data into predefined that is referred to a supervised learning because the classes are determined before examining the data and that analyses a given training set and develops a model for each class according to the features present in the data. In clustering class or groups are not predefined, but rather defined by the data alone. It is referred to as unsupervised learning.

There are three major problems related to stream data classification [2].

1. It is impractical to store and use all the historical data for training
2. There may be concept-drift in the data, meaning, the underlying concept of the data may change over time.
3. Novel classes may evolve in the stream.

In data stream classification most of the existing work related to infinite length and concept drift here we focus on the novel class detection. Most of the existing solutions assume that the total number of classes in the data stream is fixed but in real-world data stream classification problems, such as intrusion detection, text classification, fault detection, novel classes may arrive at any time in the continuous stream. There are many approaches to develop the classification model including decision trees, neural networks, nearest neighbor methods and rough set-based methods [4].The data stream classifiers are divided into two categories: single model and ensemble model [1]. Single model incrementally update a single classifier and effectively respond to concept drifting so that reflects most recent concept in data stream. Ensemble model use a combination of classifiers with the aim of creating an improved composite model, and also handle concept drifting efficiently. The traditional tree induction algorithm is that they do not consider the time in which the data arrived. The incremental classifier that reflects the changing data trends effective and efficient so it is more attractive. Incremental learning is an approach to deal with the classification task when datasets are too large or when new examples can arrive at any time [5]. Incremental learning most important in applications where data arrives over long periods of time and storage capacities are very limited. In [7] author Defines incremental tasks and incremental algorithms as follows:

Definition 1: A learning task is incremental if the training examples used to solve it become available over time, usually one at a time.

Definition 2: A learning algorithm is incremental if, for any given training sample *e1... en*, it produces a sequence of hypotheses h_0, $h_{1\ ,\ ...}$,h_n such that h_{n+1} depends only on *hi* and the current example e_i.

As per [8] the learning to be one that is: Capable to learn and update with every new data (labeled or unlabeled), Will use and exploit the knowledge in further learning, Will not rely on the previously learned knowledge, Will generate a new class as required and take decisions to merge or divide them as well

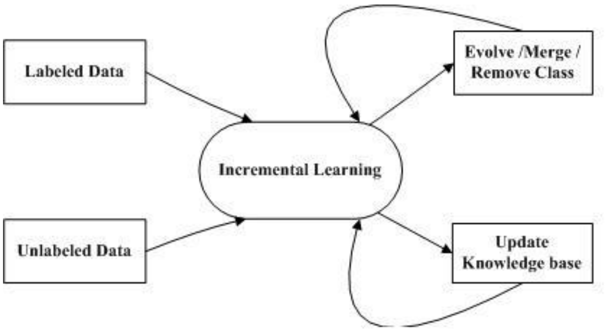


Figure 1: Working of an Incremental learning

Will enable the classifier itself to evolve and be dynamic in nature with the changing environment.

Decision tree that provide the solution for handling novel class detection problem. ID3 is very useful learning algorithm for decision tree.C5.0 algorithm improves the performance of tree using boosting. MineClass that provide solution for Novel Class. ActMiner extends MineClass, and addresses the limited labeled data problem. ECSMiner which stands for Enhanced Classifier for Data Streams with novel class Miner. The stream classification model is enhanced to handle dynamic feature sets. SCANR, which stands for Stream Classifier And Novel and Recurring class detector that address the recurring issue, and propose a more realistic novel class detection technique, which remembers a class and identifies it as "not novel" when it reappears after a long Period of time.

2. NOVEL CLASS DETECTION

Novel class detection in stream data classification is interesting research topic and researches available for concept drift problem but not attention on the Novel class detection. This approach fall into two categories : Single model (Incremental approach), Ensemble Model. Data stream classification and novelty detection recently received increasing attention in many practical real-world applications, such as spam, climate change or intrusion detection, where data distributions inherently change over time[6]. Ensemble techniques maintain an combination of models, and use ensemble voting to classify unlabeled instances. As per [6] In 2011, Masud et al. proposed a novelty detection and data stream classification technique, which integrates a novel class detection mechanism into traditional mining classifiers that enabling automatic detection of novel classes before the true labels of the novel class instances arrive, also In 2011, R. Elwell and R. Polikar introduced an ensemble of classifiers-based approach named Learn++. NSE for incremental learning of concept-drift, characterized by nonstationary environments.

In [9], [10] author gives the definition of the existing class and Novel class.

Definition 1 (*Existing class and Novel class*): Let L be the current ensemble of classification models. A class c is an existing class if at least one of the models Li ∈ L has been trained with the instances of class c. Otherwise, c is a novel class.

To detect a novel class that has the following essential property:

Property 1: A data point should be closer to the data points of its own class (*cohesion)* and farther apart from the data points of any other classes (*separation*).

In [10] show the basic idea of novel class detection using decision tree in Fig 2. That introduces the notion of used space to denote a feature space occupied by any instance, and unused space to denote a feature space unused by an instance. According to *property 1*(*cohesion*), a novel class must arrive in the unused spaces. Besides, there must be strong cohesion (e.g. *closeness*) among the instances of the novel class. Two basic steps for novel class detection.

First, the classifier is trained such that an Inventory of the used spaces is created and saved. This is done by clustering and saving the cluster summary as "pseudo point" (to be explained shortly). Secondly, these Pseudo points are used to detect outliers in the test data, and declare a novel class if there is strong Cohesion among the outliers

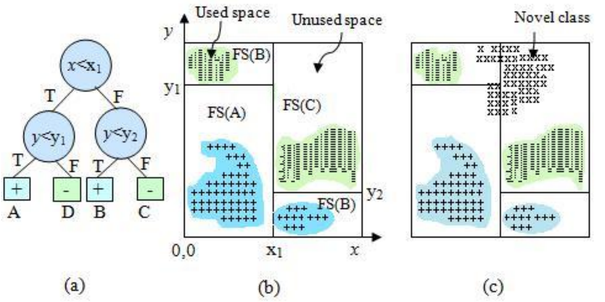


Figure 2: (a) A decision tree, (b) corresponding feature space partitioning where FS(X) denotes the Feature space defined by a leaf node X The shaded areas show the used spaces of each partition. (c) A Novel class (denoted by x) arrives in the unused space.

3. RELATED WORK

Novelty detection techniques into two categories: statistical and neural network based. Statistical approach has two types: parametric, and non-parametric. Some approaches assume that data distributions are known (e.g. Gaussian), and try to estimate the parameters (e.g. mean and variance) of the distribution called Parametric approach. If any test data that outside the normal parameter that detect as Novel.

In [10] author describe "MineClass", which stands for Mining novel Classes in data streams with base learner K-NN (K-nearest neighbor) and decision tree. K-NN based approaches for novelty detection is also non-parametric. Novelty detection is also closely related to outlier/anomaly detection techniques. There are many outlier detection techniques available, some of them are also applicable to data streams However, and the main difference with this outlier detection is that here primary objective is novel class detection, not outlier detection. Outliers are the by-product of intermediate computational steps in Novel class detection algorithm. Recent work in data stream mining domain describes a clustering approach that can detect both concept-drift and novel classa and assumes that there is only one 'normal' class and all other classes are novel. Thus, it may not work well if more than one class is to be considered as 'normal' or 'nonnovel'. Mine class can detect novel classes in the presence of concept-drift, and proposed model is capable of detecting novel classes even when the model consists of multiple "existing" classes.

In [9] ActMiner applies an ensemble classification technique by addressing the limited labeled data problem. ActMiner extends MineClass, and addresses the Limited labeled data problem in addition to addressing the other three Problems thereby reducing the labeling cost. It also applies active learning, but its data selection process is different from the others. An unsupervised novel concept detection technique for data streams is proposed, but it is not applicable to multi-class classification. As per previously mention work MineClass addresses the concept evolution problem on a multi-class classification framework. MineClass does not address the limited labeled data problem, and requires that all instances in the stream be labeled and available for training.

In [11] author describes ECSMiner for Novel class detection. Novel class detection using ECSMiner is different from traditional one class detection technique. This approach offers a "multiclass" framework for the novelty detection problem that can distinguish between different classes of data and discover the emergence of a novel class. This technique is a nonparametric approach, and therefore, it is not restricted to any specific data distribution. ECSMiner is different from other technique in three aspects: (I) It not only considers difference of test instance from training data but also similarities among them. Technique discovers novelty collectively among several coherent test points to detect the presence of a novel class. (II) It is "multiclass" novelty detection technique, and also discover emergence of a novel class. (III) Approach can detect novel classes even if concept-drift occurs in the existing classes. "ECSMiner" (pronounced like ExMiner).This technique on two different classifiers: decision tree and k-nearest neighbor. When decision tree is used as a classifier, each training data chunk is used to build a decision tree. K-NN strategy would lead to an inefficient classification model, both in terms of memory and running time. ECSMiner detect novel classes automatically even when the classification model is not trained with the novel class instances.

In [12] author proposed a *recurring class* is a special case of concept-evolution. A *recurring class* is a special and more common case of concept-evolution in data streams.It occurs when a class reappears after long disappearance from the stream. ECSMiner identifies recurring classes as novel class. Each incoming instance of data stream is first check by primary ensemble if it is outlier called it primary outlier (P-outlier) than again check through auxiliary ensemble if it is outlier than called *secondary outlier(S-outlier)*, and it is temporarily stored in a buffer for further analysis. When there are enough instances in the buffer, the *novel class detection* module is invoked. In this technique compute a unified measure of cohesion and separation for an S-outlier *x*, called *q*-NSC (neighborhood silhouette coefficient), range of q-NSC is [-1, +1]. The *q*-NSC(*x*) value of an S-outliers *x* is computed separately for each classifier $Mi \in M$. A *novel class* is declared if there are S-outliers having positive *q*-NSC for all classifiers $Mi \in M$. Recurring class instance, they should be P-outliers but not S-outliers because the primary ensemble does not contain that class, but secondary ensembles shall contain that class. The instances that are classified by the auxiliary ensembles are not outliers. The technique for Classification with novel and recurring class is called SCANR (Stream Classifier and Novel and Recurring class detector).

ERR is calculated using the following equation:

$$M_{new} = \frac{F_n * 100}{N_c} \quad (1)$$

$$F_{new} = \frac{F_p * 100}{N - N_c} \quad (2)$$

$$ERR = \frac{(F_p + F_n + F_e) * 100}{N} \quad (3)$$

Where,

Fn = Total novel class instances misclassified as existing class, *Fp* = Total existing class instances misclassified as novel class, *Fe* = Total existing class instances misclassified (other than *Fp*), *Nc* = total novel class instances in the stream, *N* = total instances the stream, *Mnew* = % of novel class instances Misclassified as existing class, *Fnew* = % of existing class instances falsely identiified as novel class, *ERR* = Total misclassification error (%) (Including *Mnew* and *Fnew).*

In [12] using (3), authors have demonstrated that OW (*OLINDDA-WCE)* has highest *ERR* rate followed by EM (ECSMiner). The main source of error for OW is *Mnew*, since it fails to detect most of the novel class instances. Therefore, the Fnew rates of OW are also low. The main source of higher error for EM compared to SC (SCANR) can be contributed to the higher Fnew rates of EM, which occurs because EM misclassifies all recurring class instances as novel ("false novel" error). Since SC can correctly identify most of the recurring class instances, the Fnew rates are low. Here describe that *ERR* rate of EM increase with increasing number of recurring classes. This is because EM identifies the recurring classes as novel. Therefore, more recurring class increases its *Fnew* rate, and in turn increases *ERR* rate. For

SC, the *Fnew* rate increases when drift increases, resulting in increased *ERR* rate. The *Fnew* rate (and *ERR*) of EM is almost independent of drift, i.e., whether drift occurs or not, it misclassifies all the recurrent class instances. However, the *Fnew* rate of SC is always less than that of EM. *Fnew* rate increases in OW because the drift causes the internal novelty detection mechanism to misclassify shifted existing class instances as novel However, for EM, here that describe *ERR* increases with increasing chunk size. The reason is that *Fnew* increases with increasing chunk size For OW, on the contrary, the main contributor to *ERR* is the *Mnew* rate. It also increases with the chunk size because of a similar reason, i.e., increased delay between ensembles update SCANR Need Extra running time because of auxiliary ensemble.

In [1] authors have proposed New decision tree learning approach for detection of Novel class. In this approach calculate the threshold value based on the ratio of percentage of data points between each leaf node in the tree and the training dataset t and also cluster the data points of training dataset based on the similarity of attribute values. If number of the data points classify by a leaf node of the tree increases than the threshold value that calculated before, which means a novel class arrived. IN [6] paper describe the decision tree learning algorithm The ID3 (Iterative Dichotomiser) technique builds decision tree using information theory. The C5.0 algorithm improves the performance of building trees using boosting, which is an approach to combining different classifiers. CART (classification and regression trees) is a process of generating a binary tree for decision making. CART handles missing data and contains a pruning strategy. The SPRINT (Scalable Parallelizable Induction of Decision Trees) algorithm uses an impurity function called gini index to find the best split .In this they introduce decision tree classifier based novel class detection in concept drifting data stream classification, which builds a decision tree from data stream. The decision tree continuously updates with new data points so that the most recent tree represents the most recent concept in data stream. Using (3), Compare the traditional decision tree and new decision tree learning approach and demonstrated the efficacy of New approach with less *ERR* rate.

4. COMPARATIVE ANALYSIS FOR NOVEL CLASS DETECTION.

The Table 1 below describes comparative analysis between different techniques of Novel class detection based on Learning Approach, type of classifier, advantages and disadvantages or limitation.

Table 1: Comparative Analysis of Various Techniques for Novel Class Detection

Algorithm	Learning Approach	Classifier	Advantage	Disadvantage
ACT Miner [9]	Ensemble	Active classifier work with K-NN and decision tree.	Work on the less label instance. It saves 90% or more labeling time and cost.	Not directly applicable to multiclass. Not work for the multi label classification.
Mine Class [9][10]	Ensemble	Decision tree and K-NN (Train and create inventory baseline techniques.)	Nonparametric. Does not require data in convex shape.	That requires 100% label instance.
ECS miner [11][12][13]	Ensemble	Classical classifier Work with K-NN and decision tree.	Non parametric Does not require data in convex shape	Not efficient in terms of memory and run time. It Identifies recurring class as Novel class.
SCANR [12]	Ensemble	Multiclass classifier	Remembers a class and identifies it as “not novel” when it reappears after a long disappearance.(Detect Recurring class)	Auxiliary ensemble is used so running time is more than other detection method
Decision tree[1][6]	Incremental	Decision tree based classifier	Detect the arrival of new class and update the tree with new recent concept	Does not work for dynamic attribute sets

5. CHALLENGES

- Concept drift and Arrival of Novel class is the challenging task for stream data mining
- Multiclass classification is challenging problem in stream data mining. [9]
- Work with less label instances and detection of recurring class is the challenging for stream data mining [10], [11].

6. CONCLUSION

Novel class detection is the more challenging task in data stream classification. In this paper we have studied the different approach that provides the solution for novel class detection with Incremental learning and Ensemble Technique. Supervised learning algorithm that has several advantages such as it is easy to implement and requires little prior knowledge, so it is very popular. Incremental approach in decision tree classifier that represent most recent concept in data stream.

7. REFERENCES

[1] Amit Biswas, Dewan Md. Farid and Chowdhury Mofizur Rahman A New Decision Tree Learning Approach for Novel Class Detection in Concept Drifting Data Stream Classification JOURNAL OF COMPUTER SCIENCE AND ENGINEERING, VOLUME 14, ISSUE 1, JULY 2012.

[2] Mohammad M. Masud, Jing Gao, Latifur Khan Integrating Novel Class Detection with Classification for Concept-Drifting Data W. Buntine et al. (Eds.):ECML PKDD 2009,Part II, LNAI 5782, pp. 79-94, Springer-Verlag Berlin Heidelberg 2009.

[3] S.PRASANNALAKSHMI, S.SASIREKHA INTEGRATING NOVEL CLASS DETECTION WITH CONCEPT DRIFTING DATA STREAMS *International Journal of communications and Engineering Volume 03– No.3, Issue: 04 March2012.*

[4] Ahmed Sultan Al-Hegami Classical and Incremental Classification in Data Mining Process IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[5] Prerana Gupta, Amit Thakkar, Amit Ganatra Comprehensive study on techniques of Incremental learning with decision trees for streamed data International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012.

[6] Dewan Md. Farid, Chowdhury Mofizur Rahman Novel Class Detection in Concept-Drifting Data Stream Mining Employing Decision Tree.

[7] Bassem Khouzam ECD Master Thesis Report INCREMENTAL DECISION TREES.

[8] Prachi Joshi, Dr. Parag Kulkarni Incremental Learning: Areas and Methods – A Survey International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012.

[9] Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, Bhavani Thuraisingham Classification and Novel Class Detection in Data Streams with Active Mining M.J.Zaki et al. (Eds.):PAKDD 2010,Part II, LNAI 6119, pp. 311-324 Springer-Verlag Berlin Heidelberg 2010.

[10] Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, Bhavani Thuraisingham Integrating Novel Class Detection with Classification for Concept-Drifting Data Streams W. Buntine et al. (Eds.):ECML PKDD 2009,Part II, LNAI 5782, pp. 79-94 Springer-Verlag Berlin Heidelberg 2009.

[11] S.Thangamani DYNAMIC FEATURE SET BASED CLASSIFICATION SCHEME UNDER DATA STREAMS *International Journal of Communications and Engineering Volume 04 – No.4, Issue: 01 March2012.*

[12] Mohammad M. Masud, Tahseen M. Al-Khateeb, Latifur Khan, Charu Aggarwal, Jing Gao, Jiawei Han, Bhavani Thuraisingham Detecting Recurring and Novel Classes in Concept-Drifting Data Streams icdm, pp.1176-1181, 2011 IEEE 11th International Conference on Data Mining, 2011.

[13] S.PRASANNALAKSHMI, S.SASIREKHA INTEGRATING NOVEL CLASS DETECTION WITH CONCEPT DRIFTING DATA STREAMS *International Journal of Communications and Engineering Volume 03, No.3, Issue: 04 March2012.*

Enhancing the vision segment architecture for AI-Robots using an orthogonal algorithm

M.Mohamed Sirajudeen
Periyar University,
Salem, India

Abstract: Object capture mechanism for robots is closely related to its vision architecture. Clarity and quality of the incoming object or object come closer to contact help to enhance the identification process. High quality and resolution cameras are used to improve the effectiveness of the acquired objects by the robots. In this paper, we focus on the angular specification of vision segment of the robots by using orthogonal algorithm. Configure the angular movement of eye ball rotation serve to cover all the directions and capture the object in an effective manner.

Key words: Vision, object and Robot

1. INTRODUCTION

Vision sensor in the robots architecture can be used to identify the object closer to its visible region. The inner components of the vision segment comprise of a high resolution cameras. For example, Contact Image Sensors (CIS) are a relatively recent technological innovation in the field of optical flatbed scanners that are rapidly replacing CCDs in low power and portable applications. As the name implies, CISs place the image sensor in near direct contact with the object to be scanned in contrast to using mirrors to bounce light to a stationary sensor, as is the case in conventional CCD scanners.

A CIS typically consists of a linear array of detectors, covered by a focusing lens and flanked by red, green, and blue LEDs for illumination. The use of LEDs allows the CIS to be highly power efficient, allowing scanners to be powered through the minimal line voltage supplied via a USB connection.

CIS devices typically produce lower image quality compared to CCD devices; in particular, the depth of field is greatly limited, which poses a problem for material that is not perfectly flat. However, a CIS contact sensor is smaller and lighter than a CCD line sensor, and allows all the necessary optical elements to be included in a compact module, thus helping to simplify the inner structure of the scanner. With a CIS contact sensor, the scanner can be portable, with a height of only around 30 mm. CIS is a both a key component of, and widely used in, scanners (especially portable scanners), electrographs, bar code readers and optical identification technology.

The proposed architecture insists the angular specification of the inbuilt digital camera in the place of vision for robots.

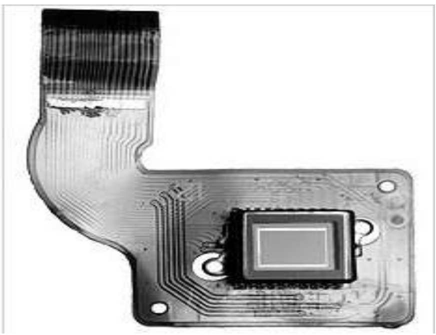


Fig 1 .0 A CCD image sensor on a flexible circuit board

[source:http://en.wikipedia.org/wiki/Contact_image_sensor].

2. RELATED WORK

2.1 Methods of Image capture – an overview

Since the first digital backs were introduced, there have been three main methods of capturing the image, each based on the hardware configuration of the sensor and color filters. The first method is often called *single-shot*, in reference to the number of times the camera's sensor is exposed to the light passing through the camera lens. Single-shot capture systems use either one CCD with a Bayer filter mosaic, or three separate image sensors (one each for the primary additive colors red, green, and blue) which are exposed to the same image via a beam splitter.

The second method is referred to as *multi-shot* because the sensor is exposed to the image in a sequence of three or more openings of the lens aperture. There are several methods of application of the multi-shot technique. The most common originally was to use a single image sensor with three filters (once again red, green and blue) passed in front of the sensor in sequence to obtain the additive color information. Another multiple shot method is called Microscanning. This technique utilizes a single CCD with a Bayer filter but actually moved the physical location of the sensor chip on the focus plane of the lens to "stitch" together a higher resolution image than the CCD would allow otherwise.

A third version combined the two methods without a Bayer filter on the chip. The third method is called *scanning* because the sensor moves across the focal plane much like the sensor of a desktop scanner. Their *linear* or *tri-linear* sensors utilize only a single line of photo sensors, or three lines for the three colors. In some cases, scanning is accomplished by moving the sensor e.g. when using Color co-site sampling or rotate the whole camera; a digital rotating line camera offers images of very high total resolution.

The choice of method for a given capture is determined largely by the subject matter. It is usually inappropriate to attempt to capture a subject that moves with anything but a single-shot system.

However, the higher color fidelity and larger file sizes and resolutions available with multi-shot and scanning backs make them attractive for commercial photographers working with stationary subjects and large-format photographs. Dramatic improvements in single-shot cameras and raw image file processing at the beginning of the 21st century made single shot, CCD-based cameras almost completely dominant, even in high-end commercial photography. CMOS-based single shot cameras remained somewhat common. [Source: http://en.wikipedia.org/wiki/Digital_camera#Methods_of_image_capture].

2.2 Planar Position Recognition and Identification – an overview

The 2D Robot Vision System is a standard system for identification and/or position recognition without direct contact by the system. It delivers reliable information on position, angle, and variation of any part. The system’s key benefit is its extremely high recognition speed at sub-pixel precise contour extraction.

Further development of the 2D technology resulted in an intelligent recognition method that enables determination of a part’s height using just a single camera and a single camera position. This technology, called 2½ D Robot Visions, results in a system that is affordable, and yet, at the same time complete. 3D palletizing is a typical application using of 2½ D Robot Vision. In the past, the available processes and equipment for 3D measurement were too complex, too expensive and thus not economically efficient for many applications. Until now, at east two - and generally more - cameras, including the necessary peripheral equipment, were required to define the positions of objects in a room.
This is why they were only used, primarily, for highly sophisticated tasks that justified the expenditure, such as the position definition of entire auto bodies in the automotive industry. MONO3D is the solution to this problem.

The new 3D robot guidance process operates with only one camera, which opens up entirely new perspectives for cost-effective image processing applications in all the important industries.

It is now possible, from one single captured image, to precisely define a three-dimensional object based on the measurement of only three criteria in all six degrees of freedom (position and orientation).

- Increased accuracy
- Greater flexibility
- Cost-effective automation

Many modern production processes in manufacturing automation are highly complex. A high level of dimensional and fitting accuracy as well as flexibility is demanded as an extensive range of product variations must be produced on a single line in a short period of time.

The patented, revolutionary system combines information from several camera systems to determine the position of large objects in any given space at extremely high precision and speed. Its function includes the recognition of entire car bodies. 3D Robot Vision identifies variations of a given position in all six degrees of freedom.

The optical 3D stereo sensor can determine positions and measure objects in all six degrees of freedom by using edges, holes, curves or parts of the object that can be described by characteristic elements, edges, lines or polynome-like path contours.

The sensor can be mounted stationary as well as on the robot arm or handling system. To achieve even higher precision, multiple sensors can be used simultaneously.

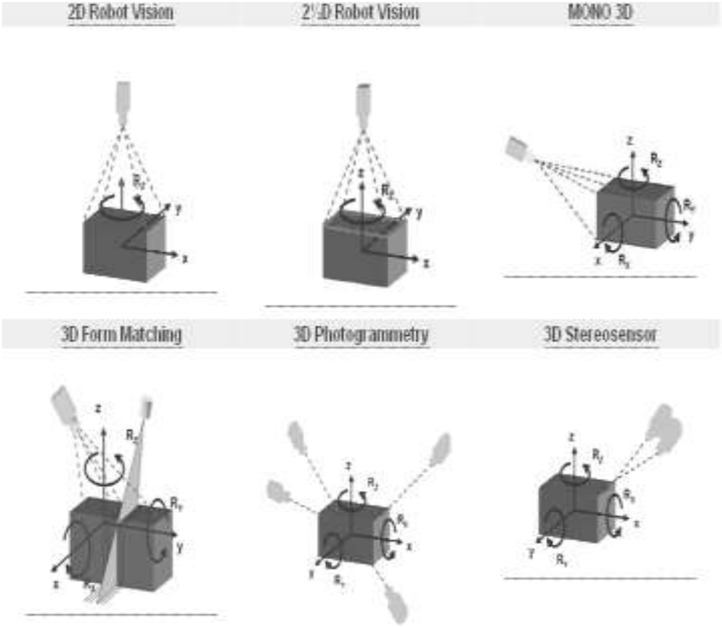


Fig 1.1 Existing vision segment structure [Source: http://isravision.com]

3. PROPOSED WORK

3.1 Architectural layout for the proposed vision segment :

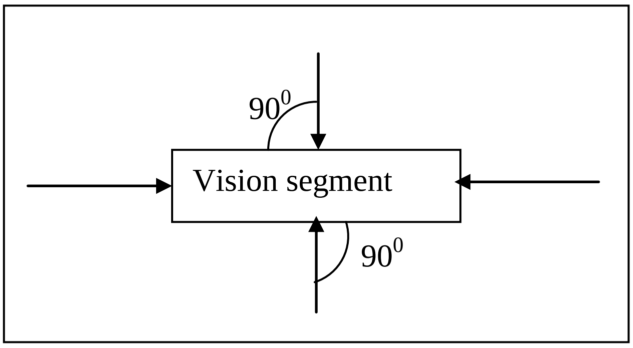


Fig 1.2 angular specification for the vision segment

The enhancement for the object/image capturing depends on the incidental and reflectional angle position for the incoming object/image. Identify the place to get the effective object/image structure help to improve the quality of the captured object.

The storage method for the incoming object helps us to reduce the complexity of the identification process. The boundary level points are used to perform the segmentation regarding to identify the layout of image.

The similarity between the physical segment of the stored object and the logical segment of the incoming object yields the result for identification process.
From the classification for the object storage make simplify the processes for identification. For example, ground based objects are occupied in one part of the memory; fly – based objects are resided in another part of the memory.

Initially we determine the nature of object enter into the visible region such as stable or moving sate, two-dimensional or three-dimensional objects either in different direction will be taken in to the account for further processing.

In the first stage any one of the segment will be taken into account for the comparison and get the probability to reach the reality of the image.

The identification process will start from a guess. Before, we start the comparison; the method of object storage will be specified by the algorithm without any ambiguity.

For this reason, the 100 percentage of success will depends on the storage classification process and if the object is new , then immediately allocate a new storage for the incoming object and display the message “ it is new in this zone.. “ along with the similarity for the image which one is already occupied into the memory.

A number of large-scale experiments [2], involving separate test images that evaluate performance with increasing number of items in the database, in the presence of clutter, background change, and occlusion, and also the results of some generic classification experiments where the system is tested on objects never previously seen or modelled.

Therefore, to avoid the process of store the entire image in the permanent storage part. The visual part of the device is deigned to cover all the directions like a rotating eye ball.

3.2 Specification for the Orthogonal Image Segment Comparison algorithm:

In order to use the system with an object, it’s any one of the view to be stored in the memory. Currently, this is done by obtaining a number of segments of the object from different directions. About a set of views are needed to cover the entire viewing sphere for the curve-based keys we have used.

Here, the searching technique for text and image will be performed by using a rotation of the image segment or text instead of using whole part of it. The way to segment and to store it will determine the efficiency of the algorithm.

The searching process will enter into the appropriate region is reduce our work for more than 80%, unless the progress will become a complicated and not possible to produce the expected result at end.

The categorization of the object will also be performed in a vey careful manner based on the standard characteristics of the nature of the objects.

An orthogonal matrix [9] is the real specialization of a unitary matrix, and thus always a normal matrix. Although we consider only real matrices here, the definition can be used for matrices with entries from any field. However, orthogonal matrices arise naturally from inner products, and for matrices of complex numbers that leads instead to the unitary requirement. Orthogonal matrices preserve inner product. So, for vectors “u, v” in an *n*-dimensional real inner product space,

$$< u, v> = <Q_u ,Q_v> \text{ [10]}$$

To see the inner product connection, consider a vector **v** in an *n*-dimensional real inner product space. Written with respect to an orthonormal basis, the squared length of “b” is “b^T b”. If a linear transformation, in matrix form Q_b, preserves vector lengths, then

$$v^T v= (Q_v)^T (Q_v)=b^T Q^T Q_v$$

The simplest orthogonal matrices are the 1×1 matrices [1] and [−1] which we can interpret as the identity and a reflection of the real line across the origin.[11]

The 2×2 matrices have the form

$$\begin{pmatrix} P & t \\ q & u \end{pmatrix}$$

Which orthogonality demands satisfy the three equations, [12]

$$1 = p^2 + q^2$$
$$1 = t^2 + u^2$$
$$0 = pt + qu$$

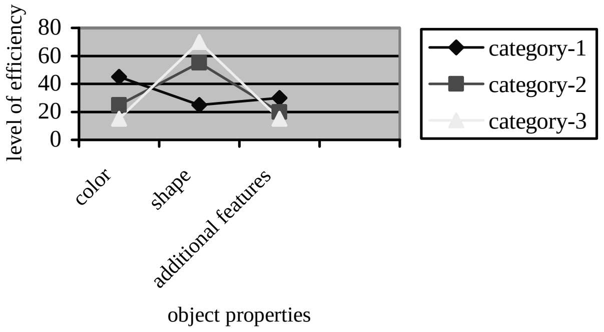
In consideration of the first equation, without loss of generality let $p = \cos\theta$, $q = \sin\theta$; then either $t = -q$, $u = p$ or $t = q$, $u = -p$. We can interpret the first case as a rotation by θ (where $\theta = 0$ is the identity), and the second as a reflection across a line at an angle of $\theta/2$.[13]

The reflection at 45° exchanges x and y; it is a permutation matrix, with a single 1 in each column and row (and otherwise 0):[13]

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The identity is also a permutation matrix [15].A reflection is its own inverse, which implies that a reflection matrix is symmetric (equal to its transpose) as well as orthogonal. The product of two rotation matrices is a rotation matrix, and the product of two reflection matrices is also a rotation matrix.

Fig 1.3 Efficiency of the object classification



4. Conclusion

The main objective of the work is to reduce the storage space for the incoming object and also avoid the complexity of multiple view storage representations in the AI-Robot's memory architecture. In general, the required storage space for the entire object take place huge amount of memory space and increase the complexity for the searching process. Instead of this, we consider only the segment of the incoming object for the identification process by using Orthogonal Image segment comparison algorithm.

References

[1]. Andrea Selinger and Randal C. Nelson, ``A Perceptual Grouping Hierarchy for Appearance-Based 3D Object Recognition", Computer Vision and Image Understanding, vol. 76, no. 1, October 1999, pp.83-92. Abstract, gzipped postscript (preprint) .

[2]. Randal C. Nelson and Andrea Selinger ``Large-Scale Tests of a Keyed, Appearance-Based 3-D Object Recognition System", Vision Research Special issue on computational vision, Vol. 38, 15-16, Aug. 1998.

[3]. Abstract, gzipped postscript (preprint) Randal C. Nelson and Andrea Selinger ``A Cubist Approach to Object Recognition", International Conference on Computer Vision (ICCV98), Bombay, India, January 1998, 614-621.

[4]. Abstract, gzipped postscript, also in an extended version with more complete description of the al

[5]. Randal C. Nelson, Visual Learning and the Development of Intelligence, In Early Visual Learning, Shree K. Nayar and Tomaso Poggio, Editors, Oxford University Press, 1996, 215-236.

[6]. Abstract, Randal C. Nelson, ``From Visual Homing to Object Recognition'', in Visual Navigation, Yiannis Aloimonos, Editor, Lawrence Earlbaum Inc, 1996, 218-250.

[7]. Abstract, Randal C. Nelson, ``Memory-Based Recognition for 3-D Objects", Proc. ARPA algorithms, and additional experiments. Image Understanding Workshop, Palm Springs CA.

[8]. The Architecture of Brain and Mind Integrating Low-Level Neuronal Brain Processes with High-Level Cognitive Behaviours, in a Functioning Robot2.

[9]. Diaconis, Persi; Shahshahani, Mehrdad (1987), "The subgroup algorithm for generating uniform random variables", Prob. in Eng. and Info. Sci. 1: 15–32, ISSN 0269-9648.

[10]. Dubrulle, Augustin A. (1999), "An Optimum Iteration for the Matrix Polar Decomposition", lect. Trans. Num. Anal. 8: 21–25, http://etna.mcs.kent.edu/

[11]. Golub, Gene H.; Van Loan, Charles F. (1996), Matrix Computations (3/e ed.), Baltimore: Johns Hopkins University Press, ISBN 978-0-8018-5414-9

[12]. Higham, Nicholas (1986), "Computing the Polar Decomposition—with Applications", SIAM J. Sci. Stat. Comput. 7 (4): 1160–1174, doi:10.1137/0907079, ISSN 0196-5204,http://locus.siam.org/SISC/volume- 07/art_0907079.html

[13]. Higham, Nicholas; Schreiber, Robert (July 1990), "Fast polar decomposition of an arbitrary matrix", SIAM J. Sci. Stat. Comput. 11 (4): 648–655, doi: 10.1137/0911038, ISSN 0196-5204,http://locus.siam.org/SISC/volume-11/art_0911038.html

International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 79-81, 2012

EXPLOITING DYNAMIC RESOURCE ALLOCATION FOR EFFICIENT PARALLEL DATA PROCESSING IN CLOUD-BY USING NEPHEL'S ALGORITHM

Kavya Jakkula
Department of Computer Science and Engineering
Kottam College of Engineering
Chinnatekur, Kurnool, A.P, India

Abstract: In recent years ad hoc parallel data processing has emerged to be one of the killer applications for Infrastructure-as-a-Service (IaaS) clouds. Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. However, the processing frameworks which are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. Consequently, the allocated compute resources may be inadequate for big parts of the submitted job and unnecessarily increase processing time and cost. In this paper, we discuss the opportunities and challenges for efficient parallel data processing in clouds and present our research project Nephele. Nephele is the first data processing framework to explicitly exploit the dynamic resource allocation offered by today's IaaS clouds for both, task scheduling and execution. Particular tasks of a processing job can be assigned to different types of virtual machines which are automatically instantiated and terminated during the job execution. Based on this new framework, we perform extended evaluations of Map Reduce-inspired processing jobs on an IaaS cloud system and compare the results to the popular data processing framework Hadoop.

.**Keywords**: Iaas-Infrasturcture-as-a-service, PACTs- Parallelization Contracts.

1. INTRODUCTION

The main goal of our project is to decrease the overloads of the main cloud and increase the performance of the cloud. In recent years ad-hoc parallel data processing has emerged to be one of the killer applications for Infrastructure-as-a-Service (IaaS) clouds. Major Cloud computing companies have started to integrate frameworks for parallel data processing in their product portfolio, making it easy for customers to access these services and to deploy their programs. However, the processing frameworks which are currently used have been designed for static, homogeneous cluster setups and disregard the particular nature of a cloud. The main objective of our project is to decrease the overloads of the main cloud and increase the performance of the cloud by segregating all the jobs of the cloud by cloud storage, job manager and task manager. and perform the different task using different resources as the infrastructure needed.

2. Companies providing cloud-scale services have an increasing need to store and analyze massive data sets such as search logs and click streams. For cost and performance reasons, processing is typically done on large clusters of shared-nothing commodity machines. It is imperative to develop a programming model that hides the complexity of the underlying system but provides flexibility by allowing users to extend functionality to meet a variety of requirements. In this paper, we present a new declarative and extensible scripting language, SCOPE (Structured Computations Optimized for Parallel Execution), targeted for this type of massive data.

2. NEPHELE/PACT ALGORITHM

We present a parallel data processor centered around a programming model of so called Parallelization Contracts (PACTs) and the scalable parallel execution engine Nephele. The PACT programming model is a generalization of the well-known map/reduce programming model, extending it with further second-order functions, as well as with Output Contracts that give guarantees about the behavior of a function. We describe methods to transform a PACT program into a data flow for Nephele, which executes its sequential building blocks in parallel and deals with communication, synchronization and fault tolerance. Our definition of PACTs allows applying several types of optimizations on the data flow during the transformation. The system as a whole is designed to be as generic as (and compatible to) map/reduce systems, while overcoming several of their major weaknesses: 1) the functions map and reduce alone are not sufficient to express many data processing tasks both naturally and efficiently. 2) Map/reduce ties a program to a single fixed execution strategy, which is robust but highly suboptimal for many tasks. 3) Map/reduce makes no assumptions about the behavior of the functions. Hence, it offers only very limited optimization opportunities. With a set of examples and experiments, we illustrate how our system is able to naturally represent and efficiently execute several tasks that do not fit the map/reduce model well.The term Web-Scale Data Management has been coined for describing the challenge to develop systems that scale to data volumes as they are found in search indexes, large scale warehouses, and scientific applications like climate research. Most of the recent approaches build on massive parallelization, favoring large numbers of cheap computers over expensive servers. Current multicore hardware trends support that development. In many of the mentioned scenarios, Parallel databases, the traditional workhorses, are refused. The main reasons are their strict schema and the missing scalability, elasticity and fault tolerance required for setups of 1000s of machines, where failures are common. Many new architectures have been suggested, among which the map/reduce paradigm and its open source implementation Hadoop have gained the most attention. Here, programs are written as map and reduce functions, which process key/value pairs and can be executed in many data parallel instances. The big advantage of that programming model is its generality: Any problem that can be expressed with those two functions can be executed by the framework in a massively parallel way. The map/reduce execution model has been proven to scale to 1000s of machines. Techniques from the map/reduce execution model have found their way into the design of database engines and some databases added the map/reduce programming model to their query interface .The map/reduce programming model has however not been designed for more complex operations, as they occur in fields like relational query processing or data mining. Even implementing a join in map/reduce requires the programmer to bend the programming model by creating a tagged union of the inputs to realize the join in the reduce function. Not only is this a sign that the programming model is somehow unsuitable for the operation, but it also hides from the system the fact that there are two distinct inputs. Those inputs may be treated differently, for example if one is already partitioned on the key. Apart from requiring awkward programming, that may be one cause of low performance .Although it is often possible to force complex operations into the map/reduce programming model, many of them require to actually describe the exact communication pattern in the user code, sometimes as far as hard coding the number and assignment of partitions. In consequence, it is at least hard, if not impossible, for a system to perform optimizations on the program, or even choose the degree of parallelism by itself, as this would require modifying the user code. Parallel data flow systems, like Dryad [10], provide high flexibility and allow arbitrary communication patterns between the nodes by setting up the vertices and edges correspondingly. But by design, they require that again the user program sets up those patterns explicitly. This paper describes the PACT programming model for the Nephele system. The PACT programming model extends the concepts from map/reduce, but is applicable to more complex operations. We discuss methods to compile PACT programs to parallel data flows for the Nephele system, which is a flexible execution engine for parallel data flows (cf. Figure 1).The contributions of this paper are summarized as follows:• We describe a programming model, centered around key/value pairs and Parallelization Contracts (PACTs). The PACTs are second-order functions that define properties on the input and output data of their associated first-order functions (from here on referred to as “user function”, UF). The system utilizes these properties to parallelize the execution of the UF and apply optimization rules. We refer to the type of the second-order function as the Input Contract. The properties of the output data are described by an attached Output Contract.• We provide an initial set of Input Contracts, which define how the input data is organized into subsets that can be processed independently and hence in a data parallel fashion by independent instances of the UF. Map and Reduce are representatives of these contracts, defining, in the case of Map, that the UF processes each key/value pair independently, and, in the case of Reduce, that all key/value pairs with equal key form an inseparable group. We describe additional functions and demonstrate their applicability. • We describe Output Contracts as a means to denote some properties on the UF’s output data.

3. PERFORMANCE EXPERIMENTS

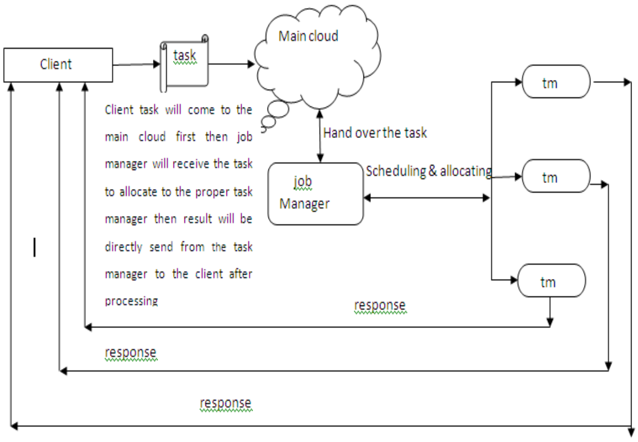
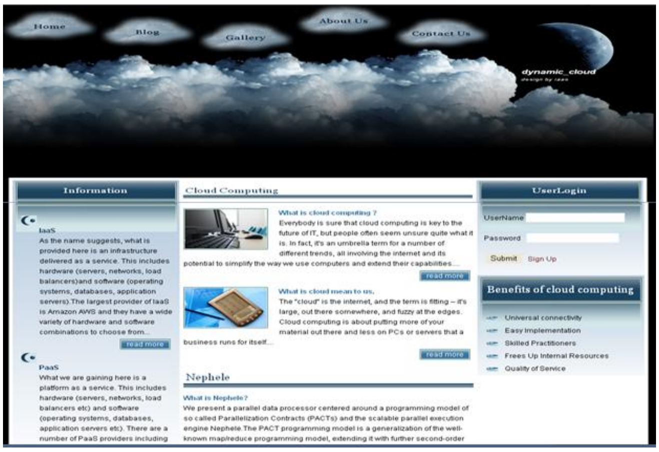

Client
task
Main cloud
Client task will come to the main cloud first then job manager will receive the task to allocate to the proper task manager then result will be directly send from the task manager to the client after processing
Hand over the task
job Manager
Scheduling & allocating
tm
tm
tm
response
response
response

Figure 1. Flow of Dynamic resource Allocation



Information

IaaS

As the name suggests, what is provided here is an infrastructure delivered as a service. This includes hardware (servers, networks, load balancers)and software (operating systems, databases, application servers).The largest provider of IaaS is Amazon AWS and they have a wide variety of hardware and software combinations to choose from...

PaaS

What we are gaining here is a platform as a service. This includes hardware (servers, networks, load balancers etc) and software (operating systems, databases, application servers etc). There are a number of PaaS providers including

Cloud Computing

What is cloud computing ?

Everybody is sure that cloud computing is key to the future of IT, but people often seem unsure quite what it is. In fact, it's an umbrella term for a number of different trends, all involving the internet and its potential to simplify the way we use computers and extend their capabilities....

What is cloud mean to us.

The "cloud" is the internet, and the term is fitting – it's large, out there somewhere, and fuzzy at the edges. Cloud computing is about putting more of your material out there and less on PCs or servers that a business runs for itself...

Nephele

What is Nephele?

We present a parallel data processor centered around a programming model of so called Parallelization Contracts (PACTs) and the scalable parallel execution engine Nephele.The PACT programming model is a generalization of the well-known map/reduce programming model, extending it with further second-order

UserLogin

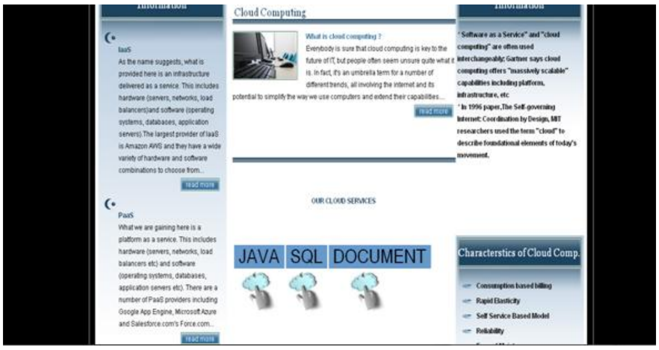
UserName

Password

Benefits of cloud computing

- Universal connectivity
- Easy Implementation
- Skilled Practitioners
- Frees Up Internal Resources
- Quality of Service

Figure 2. Dynamic resource Allocation



IaaS

As the name suggests, what is provided here is an infrastructure delivered as a service. This includes hardware (servers, networks, load balancers)and software (operating systems, databases, application servers).The largest provider of IaaS is Amazon AWS and they have a wide variety of hardware and software combinations to choose from...

PaaS

What we are gaining here is a platform as a service. This includes hardware (servers, networks, load balancers etc) and software (operating systems, databases, application servers etc). There are a number of PaaS providers including Google App Engine, Microsoft Azure and Salesforce.com's Force.com...

Cloud Computing

What is cloud computing ?

Everybody is sure that cloud computing is key to the future of IT, but people often seem unsure quite what it is. In fact, it's an umbrella term for a number of different trends, all involving the internet and its potential to simplify the way we use computers and extend their capabilities....

OUR CLOUD SERVICES

" Software as a Service" and "cloud computing" are often used interchangeably; Gartner says cloud computing offers "massively scalable" capabilities including platform, infrastructure, etc

" In 1996 paper,The Self-governing Internet: Coordination by Design, MIT researchers used the term "cloud" to describe foundational elements of today's movement.

Characterstics of Cloud Comp.

- Consumption based billing
- Rapid Elasticity
- Self Service Based Model
- Reliability

Figure 3. Technical resource Allocation

4. CONCLUSIONS

With a framework like Nephele at hand, there are a variety of open research issues, which we plan to address for future work. In particular, we are interested in improving Nephele's ability to adapt to resource overload or underutilization during the job execution automatically. Our current profiling approach builds a valuable basis for this; however, at the moment the system still requires a reasonable amount of user annotations. In general, we think our work represents an important contribution to the growing field of Cloud computing services and points out exciting new opportunities in the field of parallel data processing.

5. ACKNOWLEDGMENTS

I would like to thank my HOD Srinivas Murthy and to my beloved guide B Krishna Veni for their continuous support and guidance.

REFERENCES

[1] Amazon Web Services LLC, "Amazon Elastic Compute Cloud (Amazon EC2)," http://aws.amazon.com/ec2/, 2009.

[2] Amazon Web Services LLC, "Amazon Elastic MapReduce," http://aws.amazon.com/elasticmapreduce/, 2009.

[3] Amazon Web Services LLC, "Amazon Simple Storage Service," http://aws.amazon.com/s3/, 2009.

[4] D. Battre´, S. Ewen, F. Hueske, O. Kao, V. Markl, and D. Warneke, "Nephele/PACTs: A Programming Model and Execution Framework For Web-Scale Analytical Processing," Proc. ACM Symp.Cloud Computing (SoCC '10), pp. 119-130, 2010.

International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 82-88, 2012

EFFICIENT ENERGY SAVING AND DATA COLLECTION IN WIRELESS SENSOR NETWORKS

R. Evangelin Hema Mariya.
Kingston Engineering College,
Vellore-59, India

Abstract: Energy efficiency operations are essential in increasing the lifetime of Wireless Sensor Networks. Clustering sensor nodes is an energy saving based interesting technique for achieving these goals. The main challenge in wireless sensor network is to optimize energy consumption when collecting data from sensor nodes. Efficiently collecting the data in wireless sensor networks plays a key role in power conservation. In this paper, a new energy-efficient approach for clustering nodes in adhoc sensor networks is proposed to collect data from sensor nodes and to reduce energy consumption using modified tabu search method, which enables with low communication cost. Dynamic Tabu search uses a local or neighborhood search procedure to iteratively move from a solution x to a solution x' in the neighborhood of x, until some stopping criterion has been satisfied. Communication between the distributed cluster heads is to be achieved by low cost. This approach is suitable to avoid the energy wastage during the transmission and to prolong the lifetime of sensor networks. The MNL (Maximising Network Lifetime) algorithm is being added to prolong the life time of the distributed network. NS-2 is used to perform the progress. The performance of distributed approach is to be compared with the centralized clustering approach using Dynamic modified tabu search.

Keywords: wireless sensor networks, clustering, efficient energy saving, Distributed cluster heads, maximize network lifetime.

1. INTRODUCTION

Sensor networks have recently emerged as an important computing platform. Sensor nodes are typically less mobile and more densely deployed than mobile ad-hoc networks (MANETs).The wireless sensor network (WSN) technology is a key component for ubiquitous computing. A WSN consists of a large number of sensor nodes. Each sensor node senses environmental conditions such as temperature, pressure and light and sends the sensed data to a base station (BS), which is a long way off in general. Since the sensor nodes are powered by limited power batteries, in order to prolong the life time of the network, low energy consumption is important for sensor nodes. In order to reduce the energy consumption, a clustering and data aggregation approach is being used. In this approach, sensor nodes are divided into clusters, and for each cluster, one representative node, which called cluster head (CH), aggregates all the data within the cluster and sends the data to BS. Since only CH nodes need long distance transmission, the other nodes save the energy consumption and increase the scalability and lifetime of the network.

Clustering is one of the fundamental issues in wireless adhoc and sensor networks. In clustered sensor networks, clusterheads (CH) are responsible for data fusion within each cluster and transmit the aggregated data to the remote Base station (BS). With clustering the network payload has been greatly reduced i.e. battery energy can be considerably saved. In order to prolong the network lifetime, energy-efficient protocols should be designed for the characteristic of WSN. Efficiently organizing sensor nodes into clusters is useful in reducing energy consumption. Many energy-efficient routing protocols are designed based on the clustering structure. The clustering technique can also used to perform data aggregation, which combines the data from source nodes into a small set of meaningful information. Under the condition of achieving sufficient data rate specified by applications, the fewer messages are transmitted, the more energy is saved. Localized algorithms can efficiently operate within clusters and need not to wait for control messages propagating across the whole network. Therefore localized algorithms bring better scalability to large networks than centralized algorithms, which are executed in global structure. Clustering technique can be extremely effective in broadcast and data query. Cluster-heads will help to broadcast messages and collect interested data within their own clusters. During data collection, two mechanisms are used to reduce energy consumption: message aggregation and filtering of redundant data. These mechanisms generally use clustering methods in order to coordinate aggregation and filtering..The essential operation in sensor node clustering is to select a set of cluster heads among the nodes in the network, and cluster the rest of the nodes with these heads. This paper proposes a Distributed clustering mechanism equipped with energy maps and constrained by Quality-of-Service (QoS) requirements. Such a clustering mechanism is used to collect data in sensor networks.

2. PROBLEM STATEMENT

The Central approach is less efficient than the distributed approach in the cluster building phase. The nodes in the centralized approach have to send their information to a central node that collects all of the information and runs the algorithm to build the clusters. Energy consumed by building cluster and the energy consumed during the data collection phase is more in centralized approach.

3. RELATED WORK

There is a large body of related work in cluster formation and the communication between them that attempts to solve similar problems using various techniques**.** Moussaoui et al [1] discusses a “novel energy efficient and reliable clustering (EERC) algorithm” rebuild the clusters when there is a heavy load in the CH. Furthermore, they have a great problem in reliability that cluster-heads are easy to be attacked. It may lead uselessness of the whole cluster, thus greatly reduce the network reliability.

Raghuwanshi et al[2] similarly use the cluster communication .Communication within the cluster takes place over one-hop distance while traffic moves through the network over multi-hops to points that are connected to a much larger infrastructure. A handshake takes place between the broadcasting cluster-head and the non-cluster-head neighbors, before any data transmission can begin. Each time the nodes in the network configure – new/mobile/hibernating nodes get discovered by the local search performed as a part of the dynamic clustering scheme. Nodes that are closer in distance can have lower energy levels than farther nodes and run out of battery power quickly. The broadcast is done to make the presence known to all neighbours at single-hop distance. Based on the assumption that at least one node is awake at one-hop distance, the corresponding cluster-head sets a timer for which it decides to stay as the cluster-head

Younis et al [3] proposes another clustering method for clusters in distributed manner. Network lifetime can be defined as the time elapsed until the first node (or the last node) in the network depletes its energy (dies).energy efficient clustering method is implemented using a protocol, *HEED* (Hybrid Energy-Efficient Distributed clustering), that periodically selects cluster heads according to a hybrid of their residual energy and a secondary parameter, such as node proximity to its neighbors or node degree. Cluster heads are randomly selected based on their residual energy, and nodes join clusters such that communication cost is minimized. Simulation results show that HEED prolongs network lifetime, and the clusters it produces exhibit several appealing characteristics.

Suchismita Chinara et al [4] propose a cluster head selection criteria using an adaptive algorithm. As the selected cluster heads form the routing backbone of the dynamic network, better stability is ensured by preferring low mobile nodes to act as cluster heads. The algorithm weight based distributed mobility adaptive algorithm DMAC aims to distribute the time for which a node is selected as cluster head in an uniform manner so that every node obtains nearly equal opportunity to act as a central router for its neighbor nodes .

El Rhazi et al [5] propose a data collection algorithm using energy maps. Data aggregation and filtering methods, which minimize transmitted messages over a network, are widely used at the moment to reduce power consumption.A new data collection mechanism that uses a distributed clustering method. The new cluster building approach is based on the network energy map and the QoS requirements specified by an application. The energy consumption model determines the sensor lifetime. The energy map, the component that contains information concerning the remaining available energy in all network areas, can be used to prolong the network lifetime. A novel data collection approach for sensor networks that use energy maps to reduce power consumption and increase network coverage is used. The nodes consume more energy compared to TAG.

Heinzelman et al [6] focus on the limits of the scalability of the protocol. For this, LEACH, application-specific protocol architecture is being proposed. LEACH, a protocol architecture where computation is performed locally to reduce the amount of transmitted data, network configuration and operation is done using local control, and media access control (MAC) and routing protocols enable low-energy networking. The advantage of rotating the cluster head position among all the nodes enables LEACH to achieve a longer lifetime than static clustering. LEACH is not as efficient as LEACH-C.

Lee et al [7] define an energy consumption model. It shows the impact the coverage aging process of a sensor network, *i.e.*, how it degrades over time as some nodes become energy-depleted. To evaluate sensing coverage with heterogeneous deployments, we use total sensing coverage, which represents total information that can be extracted from all functioning sensors in a network area. Energy consumption model determines a device lifetime by considering application specific event characteristics, and network specific data extraction model and communication method**.** High-cost devices can function as a cluster-head or sink to collect and process the data from low-cost sensors, which can enhance the duration of network sensing operation.

Liang et al[8] proposes an Energy efficient method for data gathering to prolong network lifetime. The objective is to maximize the network lifetime without any knowledge of future query arrivals and generation rates. In other words, the objective is to maximize the number of data gathering queries answered until the first node in the network fails. The Algorithm MNL significantly outperforms all the other algorithms in terms of network lifetime delivered.

Basu et al[9] discusses about the data dissemination and gathering.. A majority of sensor networking applications involve data gathering and dissemination, hence energy efficient mechanisms of providing these services become critical. However, due to the broadcast nature of the wireless channel many nodes in the vicinity of a sender node overhear its packet transmissions even if those are not the intended recipients of these transmissions .This redundant reception results in unnecessary expenditure of battery energy of the recipients. Turning off neighboring radios during a certain point-to-point wireless transmission can mitigate this cost. To overcome this, Energy –Efficient data gathering and Dissemination algorithm is used.

3. PRELIMINARIES OF PROPOSED ALGORITHM

3.1 Energy Consumption Model

The energy consumption model determines the sensor lifetime. The energy calculation for a single cycle is done by using the following equation:

$$Ecycle = ED + ES + ET + ER$$

Where *ED, ES, ET* and *ER* represent the energy required for data processing, sensing, transmitting and receiving per cycle time, respectively. The quantity of energy spent for each operation depends on the network and the event model.

3.2 Energy Maps

The energy map, the component that contains information concerning the remaining available energy in all network areas, can be used to prolong the network lifetime.

3.3 Qos Requirements

The Qos requirements are based on energy, cost is used.

3.4 Data Collection Mechanism

Generally, sensor networks contain a large quantity of nodes that collect measurements before sending them to the applications. If all nodes forwarded their measurements, the volume of data received by the applications would increase exponentially, rendering data processing a tedious task. Data aggregation and data filtering are two methods that reduce the quantity of data received by Applications. The aim of those two methods is not only to minimize the energy consumption by decreasing the number of messages exchanged in the network but also to provide the applications with the needed data without needlessly overloading them with exorbitant quantities of messages. The aggregation data mechanism allows for the gathering of several measures into one record whose size is less than the extent of the initial records. However, the result semantics must not contradict the initial record semantics. Moreover, it must not lose the meanings of the initial records. The data filtering mechanism makes it possible to ignore measurements considered redundant or those irrelevant to the application needs. An example of a selective query is "SELECT the humidity readings FROM all sensors WHERE the temperature is above 40◦ for DURATION of 2 hours EVERY 5 minutes". There are two main directions in query processing for optimizing the data collection process for selective queries: (1) preventing that a query is sent to nodes that do not fall into the scope of that query and, therefore, are not aware of the query and do not need to respond, and (2) minimizing the number of non-participating nodes in the collection path.

3.5 A Dynamic Tabu Search Approach

In order to facilitate the usage of tabu search for CBP, a new graph called Grow is defined. It is capable of determining feasible clusters. A feasible cluster consists of a set of nodes. Five steps should be conducted in order to adapt tabu search heuristics to solve a particular problem:1.Design an algorithm that returns an initial solution,2.Define moves that determine the neighbourhood N of a solution s,3.Determine the content and size of tabu lists,4.Define the aspiration criteria,5.Design intensification and diversification mechanisms.

Initial solution:

The goal is to find an appropriate initial solution for the problem, in order to get the best solution from tabu search iterations within a reasonable delay.the cost c_a is found.

The Neighborhood definition:

It involves a move involving a regular node,a move involving an active node and a move involving a cluster head.The Cost c_b is found.

Increase Tabu lists:

Our adaptation proposes two tabu lists: a reassignment list and a re-election list. The size of the tabu list is to be increased.when C_a=C_b then the tabu list id increased and updated.

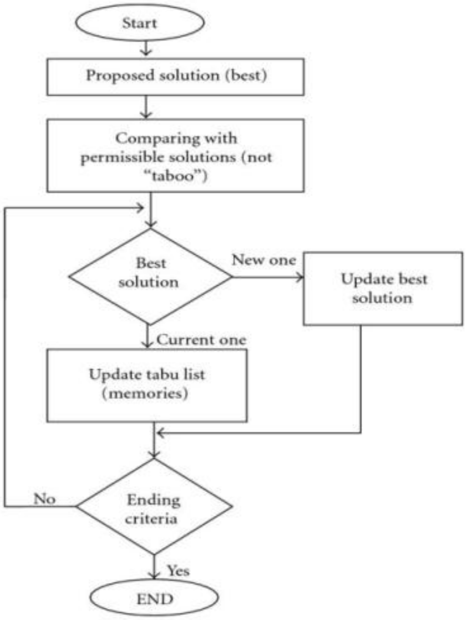


Figure 1 – Flow Diagram for modified Tabu Search for Clustering

Aspiration criteria:

Aspiration criterion, which consists of considering a move inventoried in the tabu list, which in turn, engenders a solution that is superior to the best solution found in the first place.

Design intensification and diversification mechanisms:

Diversification and Intensification are two mechanisms that make it possible to improve tabu search methods. They start by analysing the appropriate solutions. visited and obtain their common properties in order to be able to intensify the search in another neighborhood or to diversify the searches. The long and short term memory is used. The algorithm ends when one of the following 1. All possible moves are prohibited by the tabu lists;

2. The maximal number of iterations allowed has been reached;

3. The maximal number of iterations, where the best solution is not enhanced successively, has been reached.

3.6Mnl (Maximum_Network_Lifetime) Algorithm: The wireless sensor network M(N;A) is treated as a directed graph G(V ;E), where the set of nodes V consisting of sensors and (u; v) € E if and only if u and v are within the transmission ranges of each other. The basic idea is that, once a data gathering query arrives, a data gathering tree for the query is constructed using a greedy policy that maximizes the minimum residual energy among the nodes.

Specifically, the nodes are included into the tree one by one. Initially, only the sink node is included. Each time a node v is included into the tree, either the network lifetime derived from the current tree is at least as long as that without the inclusion of v to the tree or the amount of reduction of the network lifetime is minimized. In other words, a node v is chosen to be included into the tree if it leads to maximizing the minimum residual energy among the tree nodes including itself.

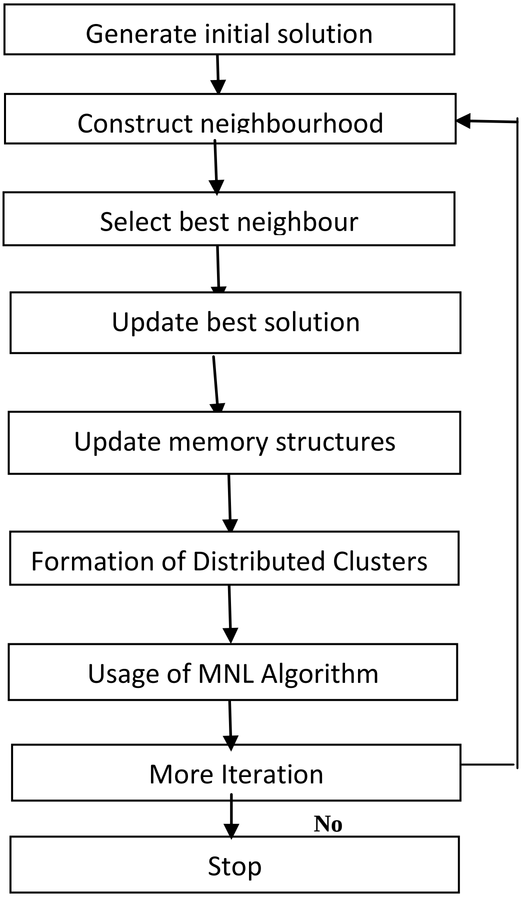


Figure 2-Tabu Search Algorithm and MNL algorithm for Distributed Clustering.

4 Simulation Result and Discussion

4.1 Comparison between Centralized and Distributed approach

The figure shows, the energy consumed to build the clusters by the centralized and distributed approaches using Dynamic tabu search approach. Results show that the distributed approach needs less energy consumption than the centralized approach and the gap between these energies becomes bigger when the network size increases.

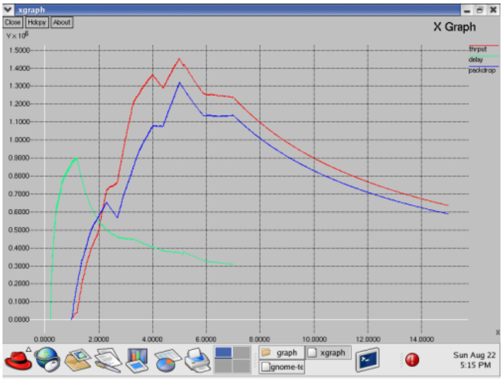


Figure 3 - Centralized Approach

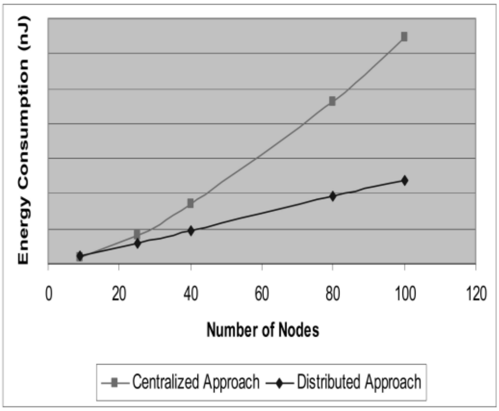


Figure 4- Comparison between Centralized and Distributed Approach

The reason behind this result is that the central node needs to generate a considerable number of messages in order to collect all the node information. The dynamic tabu search is good when compared to the normal tabu search in the memory storage.

5. Conclusions

This paper has presented a heuristic approach based on a energy efficient search to solve clustering problems where the numbers of clusters and cluster heads are unknown beforehand. The tabu search adaptation consists of defining three types of moves that allow reassigning nodes to clusters, selecting cluster heads, and removing existing clusters. Such moves use the largest size clique in a feasibility cluster graph, which facilitates the analysis of several solutions and makes it possible to compare them using a gain function. Performance of distributed approach with those of a centralized approach and we conclude that the central approach is less efficient than the distributed approach in the cluster building phase.

References

[1]. O. Moussaoui, A. Ksentini, M. Naimi, and M. Gueroui, “A Novel Clustering Algorithm for Efficient Energy Saving in Wireless Sensor Networks,” Proc. Seventh Int’l Symp. Computer Networks (ISCN ’06), pp. 66-72, 2006.

[2].S. Raghuwanshi and A. Mishra, “A Self-Adaptive Clustering Based Algorithm for Increased Energy-Efficiency and Scalability in Wireless Sensor Networks,” Proc. IEEE 58th Vehicular Technology Conf. (VTC ’03), vol. 5, pp. 2921-2925, 2003.

[3]. O. Younis and S. Fahmy, “Distributed Clustering in Ad-Hoc Sensor Networks: A Hybrid, Energy-Efficient Approach,” Proc. IEEE INFOCOM, pp. 629-640, 2004.

[4] Suchismita Chinara , Santanu Kumar Rath “Energy Efficient Mobility Adaptive Distributed Clustering Algorithm for Mobile Ad Hoc Network,” Proc ADCOM 2008,pp 265-272,2008

[5].A. El Rhazi and S. Pierre, “A Data Collection Algorithm Using Energy Maps in Sensor Networks,” Proc. Third IEEE Int’l Conf. Wireless and Mobile Computing, Networking, and Comm. (WiMob ’07), 2007.

[6]. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “An Application Specific Protocol Architecture for Wireless Microsensor Networks,” IEEE Trans. Wireless Comm., vol. 1, no. 4, pp. 660-670, Oct. 2002.

[7]. J.J. Lee, B. Krishnamachari, and C.C.J. Kuo, “Impact of Heterogeneous Deployment on Lifetime Sensing Coverage in Sensor Networks,” Proc. IEEE Sensor and Ad Hoc Comm. and Networks Conf. (SECON ’04), pp. 367-376, 2004.

[8].W. Liang and Y. Liu, “Online Data Gathering for Maximizing Network Lifetime in Sensor Networks,” IEEE Trans. Mobile Computing, vol. 6, no. 1, pp. 2-11, Jan. 2007

.[9]. P. Basu and J. Redi, “Effect of Overhearing Transmissions on Energy Efficiency in Dense Sensor Networks,” Proc. Third Int’l Symp. Information Processing in Sensor Networks (IPSN ’04), pp. 196- 204, Apr. 2004.

International Journal of Computer Applications Technology and Research
Volume 1– Issue 3, 89-93, 2012

Development of Maize Expert System Using Ada-Boost Algorithm and Naïve Bayesian Classifier

M.S. Prasad Babu
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

Venkatesh Achanta
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

N.V.Ramana Murty
Dept.of CS
Rayalaseema University,
Kurnool,
A.P,India

Swapna.K
Dept. of CS & SE,
Andhra University,
Visakhapatnam,
A.P, India

Abstract: Machine learning Recent works on ensemble [1] methods like Adaptive Boosting have been applied successfully in many problems. Ada-Boost algorithm running on a given weak learner several times on slightly altered data and combining the hypotheses in order to achieve higher accuracy than the weak learner. This paper deals with the design and development of an expert system to advice the farmers in villages through online. An expert system is a computer program, with a set of rules encapsulating knowledge about a particular problem domain. This is a web based application developed using machine learning techniques. In the present paper, Ada-Boost algorithm technique has been considered and applied to generate conclusion based on the given training data provided by human expert. Here a rule based expert system and machine learning systems are integrated to form the proposed Ada-Boost Algorithm Using Naïve Bayesian Classifier on Maize expert Advisory System. The proposed system examines the symptoms provided by the user and process the information based on the training data and determines the diseases effected to the Maize crop. The system shows a global solution for recognizing the diseases in Maize crop cultivation and also suggests the corresponding treatments to the diseases. This expert system is a web based application for online users with java as front end and MYSQL as backend.

Keywords: Expert Systems, Rule-Based System, Machine Learning, Ada-Boost, Naïve Bayesian Classifier, Maize, JSP and MYSQL

1. INTRODUCTION

A. Maize Crop Information

Maize known in many English-speaking countries as corn or mielie/mealie is a grain domesticated by indigenous peoples in Mesoamerica in prehistoric times. The leafy stalk produces ears which contain seeds called kernels. Maize kernels are technically a fruit but are used in cooking as a vegetable or starch. The Olmec and Mayans cultivated it in numerous varieties throughout central and southern Mexico, cooked, ground or processed through nixtamalization. Between 1700 and 1250 B.C, the crop spread through much of the Americas. Sugar-rich varieties called sweet corn are usually grown for fresh consumption while field-corn varieties are used for animal feed and as chemical feed stocks.

B. Expert Systems

An expert system [5] is software that uses a knowledge base of human expertise for problem solving, or to clarify uncertainties where normally one or more human experts would need to be consulted. Expert systems are most common in a specific problem domain, and are a traditional application and/or subfield of artificial intelligence (AI). A wide variety of methods can be used to simulate the performance of the expert; however, common to most or all are: 1) the creation of a knowledge base which uses some knowledge representation structure to capture the knowledge of the Subject Matter Expert (SME); 2) a process of gathering that knowledge from the SME and codifying it according to the structure, which is called knowledge engineering; and 3) once the system is developed, it is placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or as a supplement to some information system. Expert systems may or may not have learning components. A series of Expert advisory systems [12], [13], [15] were developed in the field of agriculture and implemented in www.indiakisan.net[14].

C. Machine Learning

Machine learning[2, 3, 4 and 6], a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. It is a very young scientific discipline used in various areas including Robotics, Machine Vision, etc. The First Machine Learning Workshop was taken place in 1980 at Carnie-Mellon University (USA).The goal of machine learning is to program computers to use training data or past experience to solve a given problem. Effective algorithms have been invented for certain types of learning tasks. Many practical computer programs have been developed to exhibit useful types of learning and significant commercial applications have begun to appear. Machine learning refers to the changes in systems that perform tasks associated with artificial intelligence (AI). Such tasks involve recognition, diagnosis, planning, robot control, prediction, etc. Some of the machines learning algorithms are Genetic Algorithm, Decision Tree Algorithm, Optimization Algorithm, Adaptive Boosting Algorithm, Bagging Algorithm and Particle Swarm Optimization Algorithm, Bayesian Classifier Algorithm, ID3 and C4.5 Algorithm.

D. Adaptive Boosting (Ada-Boost) Algorithm

Ada-Boost, [7, 8, 11] short for Adaptive Boosting, is a machine learning algorithm, formulated by Yoav Freund and Robert Schapire. It is a meta-algorithm, and can be used in conjunction with many other learning algorithms to improve their performance. Ada-Boost is adaptive in the sense that subsequent classifiers built are tweaked in favor of those instances misclassified by previous classifiers. Ada-Boost is sensitive to noisy data and outliers. However in some problems it can be less susceptible to the overfitting problem than most learning algorithms. Ada-Boost calls a weak classifier repeatedly in a series of rounds t=1,2,…..T from a total T classifiers. For each call a distribution of weights Dt is updated that indicates the importance of examples in the data set for the classification. On each round, the weights of incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

The pseudo code for Ada-Boost algorithm is given as below

- **Input**: a set S, of 'm' labeled examples: S= ((xi,yi), i=(1,2,…,m)), with labels in Y.
- **Learn** (a learning algorithm)
- **A constant L.**

```
[1]  Initialize for all i:  wj(i)=1/m  //  initialize the weights

[2]  for j=1 to L do

[3]  for all i:                    // compute normalized weights

         $p_j(i) = \frac{w(i)}{\sum_i^m w(i)}$

[4]  hj: =Naïve-Bayesian(S,pj)            // call weak Learn
with normalized weights

[5]  Calculate the error of hj

         $\varepsilon_j = \sum_i p_j(i)[h_j(x_i \neq y_i)]$

[6]  if  $\varepsilon_j > \frac{1}{2}$ then

[7]  L=j-1

[8]  go to 12

[9]

         $\beta_j = \frac{\varepsilon_j}{1-\varepsilon_j}$

[10] for all i:                    // compute new weights

$w_{j+1}(i) = w_j(i)\beta_j^{1-[h_j(x_i - y_i)]}$

[11] end for

12]  Output:
```

$$h_{final}(x) = \arg\max_{y\in Y} \sum_{j=1}^{L} \left(\log\frac{1}{\beta_h}\right)[h_j(x=y)]$$

E. Naïve Bayes Classifier (Weak Classifier)

Naïve Bayes Classifier is a simple probabilistic classifier based on Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". Depending on the precise nature of the probability model, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naive Bayes models uses the method of maximum likelihood; in other words, one can work with the Naive Bayes model without believing in Bayesian probability or using any Bayesian methods. In spite of their naive design and apparently over-simplified assumptions, Naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, analysis of the Bayesian classification problem has shown that there are some theoretical reasons for the apparently unreasonable efficacy of Naive Bayes classifiers [9]. Still, a comprehensive comparison with other classification methods in 2006 showed that Bayes classification is outperformed by more current approaches, such as boosted trees or random forests [10].

2. PROPOSED MAIZE EXPERT ADVISORY SYSTEM

A. Maize Crop Information

The Proposed Ada-Boost Algorithm uses the Naïve-Bayes classifier as weak learner and it uses the training data and the weights are initialized based on the number of classifiers i.e., the weights of the each class is equal to the fraction of the total number of classifiers. Select 'T', the number of rounds the algorithm has to run iteratively by adjusting the weights. In each round the weak learner is called based on the given input and the weights for each classifier and it generates a new hypothesis 'hj' in each hypothesis and the weight and the error is calculated based on the obtained hypothesis and based on the error value obtained the new weights are calculated by using the formula given below

$$w_{j+1}(i) = w_j(i)\beta_j^{1-[h_j(x_i - y_i)]}$$

where βj is error coefficient.

The weak learner is called by using the new weights. The process is repeated until the error value greater than ½ or the number of iterations completes. And finally, the hypothesis value is calculated by using the given formula.

$$h_{final}(x) = \arg\max_{y\in Y} \sum_{j=1}^{L} \left(\log\frac{1}{\beta_h}\right)[h_j(x=y)]$$

The flow diagram of the proposed Ada-Boost algorithm used in development of this Expert Advisory System is shown in fig. 1.

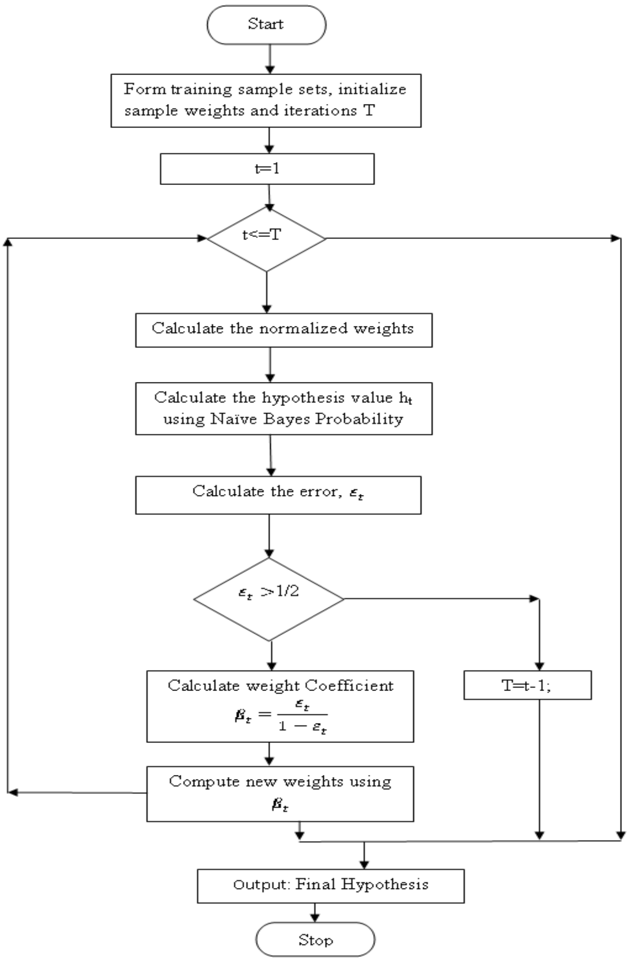

Start
Form training sample sets, initialize sample weights and iterations T
t=1
t<=T
Calculate the normalized weights
Calculate the hypothesis value h_t using Naïve Bayes Probability
Calculate the error, ε_t
ε_t >1/2
Calculate weight Coefficient $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$
T=t-1;
Compute new weights using β_t
Output: Final Hypothesis
Stop

Figure 1 Flow chart of the Ada-Boost Algorithm

B. Simple Example

The working of the proposed system is explained by considering the 10 symptoms as input. It is explained as follows

- Encode Solution: Just use 10 bits (1 or 0).
- Generate input.

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	0	1	0	0	0	1	0	1	0

- Initialize the weights wi based on the classifiers. Consider there are 5 classifiers, wi= 1/5.
- Select the value for 'T', the number of iterations.
- In each and every, iterations the hypothesis value 'hj' is to be calculated.

The probability densities for each disease is calculated using the naïve Bayesian classifier as follows

P(Disease1/s1,..s10)= P(Disease1)*P(s1/Disease1)*P(s2/Disease2)….P(s10/Disease)

By using the above equation for the given input string

P(Corn Streak Virus/1,0,1,0,0,0,1,0,1,0)= 0.20000002

P(Sorghum Down Mildew/1,0,1,0,0,0,1,0,1,0)=0.0

P(Postfloweringstalk rot/1,0,1,0,0,0,1,0,1,0)=0.029626261

P(Phaesophearia Leaf spot/1,0,1,0,0,0,1,0,1,0)=0.12922

P(Alternaria Leaf Spot/1,0,1,0,0,0,1,0,1,0)=0.122

The Probability value for disease Corn Streak Virus is greater than all the remaining diseases hence the hypothesis generated by disease name as "Corn Streak Virus".

- The error value is calculated by adding the probabilities value of the remaining diseases with their corresponding weights.
- Based on the error value the algorithm is repeated repeatedly for 'T' times by adjusting the weights.

The hypothesis values and error values of the given input for **Corn Streak Virus** using Ada-Boost algorithm are shown as below.

Value of 'T'	Hypothesis Value	Error Value
1	0.20001	0.455
2	0.22001	0.346
3	0.22286	0.320
4	0.2272	0.312

Figure 2 Computations Performed During Ada-Boost Algorithm

The final hypothesis generates the disease having the hypothesis value is greater than all other diseases.

Thus for the given example the "Corn Streak Virus" disease is generated.

3. DATABASE DESIGN

This section explains the database design used for the development of the expert system. It explains about the different rules stored in the knowledge base. Generally the rules of the form,

Rule1: S1=0, S2=1, S3=0, S4=1, S5=1, S6=0, S7=1, S8=0, S9= 1, S10=0 resultant disease may be “CORN STREAK VIRUS”.

Rule2: S1=1, S2=0, S3=1, S4=0, S5=0, S6=1, S7=0, S8=1, S9=0, S10=1 resultant disease may be “PRE FLOWERING STALK ROT”.

Rule3: S1=0, S2=1, S3=1, S4=0, S5=1, S6=1, S7=0, S8=1, S9=1, S10=1 resultant disease may be “SORGHUM DOWNY MILDEW”.

Rule4: S1=0, S2=1, S3=1, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be “POST FLOWERING STALK ROT”.

Rule5: S1=0, S2=0, S3=0, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be “ALTENARIA LEAF SPOT”.

Rule6: S1=1, S2=1, S3=1, S4=1, S5=0, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be “TURICUM LEAF BLIGHT”.

Rule7: S1=0, S2=0, S3=1, S4=0, S5=1, S6=1, S7=1, S8=0, S9=1, S10=1 resultant disease may be “FLEA BEETLES AND FLEA ROOTWORMS”.

Rule8: S1=0, S2=0, S3=0, S4=1, S5=0, S6=1, S7=1, S8=1, S9=1,S10=0 resultant disease may be “PHYLLOSTICTAL LEAF SPOT”.

Rule9: S1=1, S2=1, S3=0, S4=1, S5=1, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be “CHARCOAL STALK ROT ”.

Rule10: S1=0, S2=1, S3=0, S4=1, S5=0, S6=1, S7=0, S8=1, S9=1, S10=0 resultant disease may be “MAIZE FINE STRIPE VIRUS”.

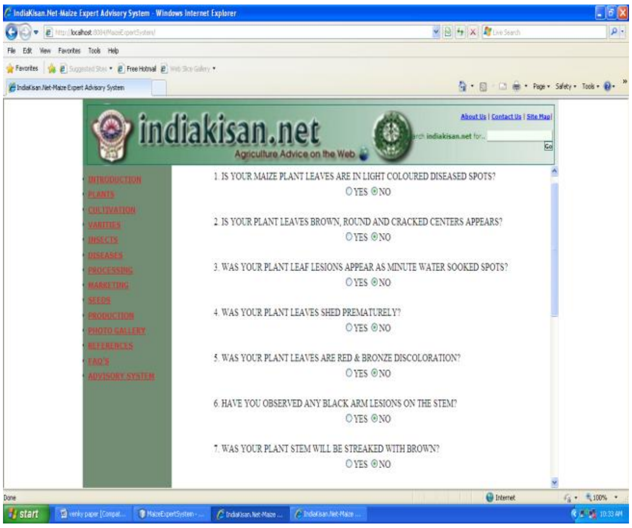
4. RESULTS

Description: In this screen shot, the user can submit the observed symptoms to the maize advisory system through online by selecting the appropriate radio buttons for the processing of the symptoms observed.

From the above screenshot, the following result is observed

Effected With: Sorghum Downy Mildew

Cure is: Spray carbendazim 1.5g and use metalaxyl MXL


IndiaKisan.Net-Maize Expert Advisory System - Windows Internet Explorer
indiakisan.net
Agriculture Advice on the Web
About Us | Contact Us | Site Map
Search indiakisan.net for..
Go
INTRODUCTION
PLANTS
CULTIVATION
VARITIES
INSECTS
DISEASES
PROCESSING
MARKETING
SEEDS
PRODUCTION
PHOTO GALLERY
REFERENCES
FAQ'S
ADVISORY SYSTEM
start

1. IS YOUR MAIZE PLANT LEAVES ARE IN LIGHT COLOURED DISEASED SPOTS?
○ YES ● NO

2. IS YOUR PLANT LEAVES BROWN, ROUND AND CRACKED CENTERS APPEARS?
○ YES ● NO

3. WAS YOUR PLANT LEAF LESIONS APPEAR AS MINUTE WATER SOOKED SPOTS?
○ YES ● NO

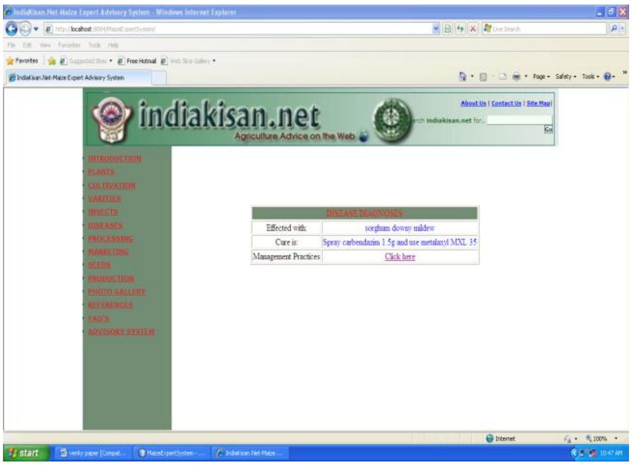
4. WAS YOUR PLANT LEAVES SHED PREMATURELY?
○ YES ● NO

5. WAS YOUR PLANT LEAVES ARE RED & BRONZE DISCOLORATION?
○ YES ● NO

6. HAVE YOU OBSERVED ANY BLACK ARM LESIONS ON THE STEM?
○ YES ● NO

7. WAS YOUR PLANT STEM WILL BE STREAKED WITH BROWN?
○ YES ● NO

Figure 2 Screen for selecting symptoms in Maize Expert System



DISEASE DIAGNOSIS	
Effected with:	sorghum downy mildew
Cure is:	Spray carbendazim 1.5g and use metalaxyl MXL 35
Management Practices	Click here

Figure 3 Displaying advices to the farmer

5. CONCLUSION

An Expert Advisory System entitled “Maize Expert Advisory System Using Ada-Boost Algorithm” is developed using Java Server Pages (JSP) and MYSQL database as backend. This system generates advices based on the symptoms given by the farmer and gives appropriate suggestions to improve the productivity of the crop. This algorithm enhances the performance of the weak learner in iterations by adjusting the weights and reducing the misclassification error values. Thus the performance of the system is enhanced.

6. REFERENCES

[1] Dietterich, T. G., 2000, “An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization,” *Machine Learning*, pp *139-158.*

[2] . Reviews of “Machine Learning by Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell”, Tiago Publishing Company, 1983, ISBN 0-935382-05-4.

[3] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”, Machine Learning, 20(3):273–297, 1995.

[4] C.A.Coello, Gary B.Lamont and David A.Van Veldhuizen, “Evolutionary Algorithms for Solving Multi-Objective Problems”, *2nd Edition, Springer,* 2007.

[5] Dr. B D C N Prasad, P E S N Krishna Prasad and Y Sagar, “A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)” *–CCSIT 2011, Part I, CCIS 131,* pp 570– 576.

[6] Rennie J, Shih L, Teevan J, and Karger D, “Tackling The Poor Assumptions of Naive Bayes Classifiers,” In Proceedings *of the Twentieth International Conference on Machine Learning (ICML).* 2003.

[7] Yoav Freund and Robert E. Schapire, “Experiments with a new boosting algorithm”, In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 148–156, 1996.

[8] A. J. M. Abu Afza, Dewan Md. Farid, and Chowdhury Mofizur Rahman, “A Hybrid Classifier using Boosting, Clustering, and Naïve Bayesian Classifier”, *World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 3,*105-109, 2011.

[9] Rish Irina, "An empirical study of the naive Bayes classifier", *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.*

[10] Atsushi Takemura, Akinobu Shimizu, and Kazuhiko Hamamoto, “Discrimination of Breast Tumors in Ultrasonic Images Using an Ensemble Classifier Based on the AdaBoost Algorithm With Feature Selection”, *IEEE Transactions on Medical Imaging, vol. 29, no. 3,* 2010.,

[11] Atsushi Takemura, Akinobu Shimizu, and Kazuhiko Hamamoto, “Discrimination of Breast Tumors in Ultrasonic Images Using an Ensemble Classifier Based on the AdaBoost Algorithm With Feature Selection”, *IEEE Transactions on Medical Imaging, vol. 29, no. 3,* 2010.,

[12] Prof. M.S. Prasad Babu, N.V. Ramana Murty, S.V.N.L.Narayana, “A Web Based Tomato Crop Expert Information System Based on Artificial Intelligence and Machine learning algorithms”, *International Journal of Computer Science and Information Technologies, Vol. 1 (1), (ISSN: 0975-9646).,* 2010, pp6-15.

[13] Prof. M.S. Prasad Babu, Mrs. J. Anitha, K. Hari Krishna, “A Web Based Sweet Orange Crop Expert System using Rule Based System and Artificial Bee Colony Optimization Algorithm” , *International Journal of Engineering Science and Technology ,vol.2(6),*2010.

[14] www.indiakisan.net

[15] Prof. M.S. Prasad Babu, N.Thirupathi Rao, “Implementation of Parallel Optimized ABC Algorithm with SMA Technique for Garlic Expert Advisory System”, *International Journal of Computer Science, Engineering and Technology (IJCSET), Volume 1, Issue 3,* October 2010.