Study on Deep Learning as a Powerful Technology that Revolutionizing Automation in Industries

Manju Susan Thomas, MPhil
Assistant Professor, Department of Computer Science and Applications
SAFI Institute of Advanced Study
Vazhayoor, Kerala, India

Abstract: Smart production refers to the usage of superior records analytics to complement bodily technology for enhancing device performance and choice making. With the extensive deployment of sensors and Internet of Things, there is a growing need of managing large manufacturing facts characterized through excessive quantity, excessive velocity, and high range. Deep gaining knowledge of present's superior analytics gear for processing and analyzing huge production facts. This paper affords a comprehensive survey of typically used deep mastering algorithms and discusses their programs in the direction of making production "clever". The evolvement of deep mastering technologies and their benefits over traditional gadget gaining knowledge of are first of all mentioned. Subsequently, computational strategies based totally on deep getting to know are provided specifically purpose to improve device overall performance in manufacturing. Several consultant deep mastering models are comparably mentioned. Finally, emerging topics of research on deep learning are highlighted, and destiny trends and challenges related to deep getting to know for smart production are summarized.

Keywords: deep learning, technology, revolution, automation, industries

1. INTRODUCTION

Deep learning is the field of learning deep organized and unstructured portrayal of information. Deep learning is the developing pattern in AI to digest better outcomes when information is vast and complex. Deep learning design comprises of deep layers of neural systems, for example, input layer, shrouded layers, and yield layer. Shrouded layers are utilized to comprehend the mind boggling structures of information. A neural system shouldn't be modified to play out an intricate errand. Gigabytes to terabytes

1.1 Background of Deep Learning

Since 2006, profound organized learning, or all the more regularly called profound learning or progressive learning, has developed as another region of AI explore. Amid the previous quite a while, the strategies created from profound learning research have just been affecting a wide scope of flag and data preparing work inside the customary and the new, extended degrees including key parts of AI and man-made brainpower; see outline articles in and furthermore the media inclusion of this advancement in .A progression of workshops, instructional exercises, and special issues or meeting special sessions as of late have been dedicated exclusively to profound learning and its applications to different flag and data preparing territories. These include:

- 2008 NIPS Deep Learning Workshop;
- 2009 NIPS Workshop on Deep Learning for Speech Recognition and Related Applications;
- 2009 ICML Workshop on Learning Feature Hierarchies;

1.2 Automation in Industry

Much industry is mechanized and exceptionally specialized utilizing programmed framework. Development of in enterprises is colossal and of more extensive degree. Be that as it may, mechanization level differs from industry to industry. Farming, deals and some administration industry are hard to mechanize though industry like correspondence especially phone industry its exceptionally robotized seething from managing, transmission and charging all done naturally. In rural industry is relevant in the preparing and pressing of sustenance's however there are some administration territories which can't be computerized.

Over the beyond century, the producing industry has passed through some of paradigm shifts, from the Ford meeting line (1900s) to Toyota production gadget (1960s), flexible manufacturing (1980s), reconfigurable production (1990s), agent-based production (2000s), cloud production (2010s). Various nations have evolved strategic roadmaps to transform production to take advantage of the emerging infrastructure, as presented by way of Internet of Things (IoTs) and information science. As an instance, Germany added the framework of Industry 4.Zero in 2010, which has been evolved right into a collaborative attempt amongst member international locations within the European Union. Similarly, in 2011 the Smart Manufacturing Leadership Coalition (SMLC) in the U.S. Created a systematic framework for imposing clever production. The plan "China Manufacturing 2025", delivered in China in 2015, goals to sell superior production. As manufacturing machines are more and more geared up with sensors and communication abilties, there is full-size ability to further enhance the condition-awareness of manufacturing machines and approaches, lessen operational downtime, enhance the level of automation and product great and response more well timed to dynamically converting consumer demands . Statistics suggests that 82% of the companies the usage of clever production technology have experienced increased efficiency and forty five% of the organizations of the agencies skilled elevated purchaser delight. Smart production refers to a brand new production paradigm where manufacturing machines are completely connected via wireless networks, monitored by way of sensors, and managed via advanced computational intelligence to enhance product great, device productivity, and sustainability at the same time as lowering costs. Recent advancement of Internet of Things (IoTs), Cloud Computing, Cyber Physical System (CPS) offers key assisting technology to improve contemporary production. By leveraging these new technology in manufacturing, records at distinct stages of a product's existence, ranging from uncooked materials, machines' operations, facility logistics, and even human operators, is amassed and processed . With the proliferation of producing facts, records pushed intelligence with advanced analytics transforms unparalleled volumes of data into actionable and insightful information for clever manufacturing as illustrated in Fig. 1. Data driven intelligence fashions the complex multivariate nonlinear relationships amongst facts, with no in-intensity expertise of gadget physical behaviours required. Data driven intelligence has attracted massive research effort for manufacturing records distilling and choice making.

Fig. 1. The role of data driven intelligence in smart manufacturing.

In statistics mining techniques are categorized into 5 categories, together with characterization and description, affiliation, classification, prediction, clustering and evolution analysis. The obstacles to statistics-pushed choice making in manufacturing are also identified. Typical device mastering strategies are reviewed in for intelligent production, and their strengths and weaknesses also are mentioned in a wide variety of producing packages. A comparative study of device learning algorithms such as Artificial Neural Network, Support Vector Machine, and Random Forest is performed for machining device wear prediction. The schemes, strategies and paradigm of developing selection making guide structures are reviewed for the tracking of machining operations, and those strategies consist of neural networks, fuzzy common sense, genetic algorithms, and hybrid structures. The potential gain and successful software examples of usual machining studying techniques such as Bayesian Networks, instance-based gaining knowledge of, Artificial Neural Network, and ensemble techniques are mentioned in Cloud enabled analysis strategies including records driven method, physics primarily based in addition to model-based totally strategies are reviewed in, with the blessings from both advanced computing functionality and facts sharing for sensible choice making. Traditional system getting to know is generally designed with shallow structures, inclusive of Artificial Neural Network, Support Vector Machine, and logistic regression, and many others. By coping with restricted handcrafted functions, it achieves first rate overall performance in a spread of packages. However, the big information in smart production imposes a diffusion of demanding situations , inclusive of the proliferation of multimodal statistics, high dimensionality of feature space, and multicollinearity among information measurements. These challenges render traditional algorithms struggling and thus greatly hinder their overall performance. As a step forward in artificial intelligence, deep gaining knowledge of demonstrates fantastic overall performance in numerous packages of speech recognition, image recondition, herbal language processing (e.G. Translation, information, check questions & solutions), multimodal photograph-textual content, and games (e.G. Alphago). Deep studying permits robotically processing of information toward notably nonlinear and complicated function abstraction through a cascade of multiple layers, rather than handcrafting the premiere function representation of facts with domain expertise. With automated function gaining knowledge of and excessive-extent modelling abilities, deep learning provides a sophisticated analytics tool for clever manufacturing in the huge information technology. It uses a cascade of layers of nonlinear processing to learn the representations of records corresponding to specific degrees of abstraction. The hidden styles under every other are then recognized and predicted via quit-to-quit optimization. Deep studying gives wonderful capacity to boost records-driven production applications, in particular inside the large records generation In mild of the above challenges, this paper objectives to provide a ultra-modern review of deep mastering strategies and their packages in clever production. Specifically, the deep gaining knowledge of enabled advanced analytics framework is proposed to fulfill the opportunistic need of clever manufacturing. The standard deep gaining knowledge of models are briefly added, and their programs to production are outlined to highlight the trendy development in applicable areas. The challenges and future traits of deep getting to know are mentioned ultimately. The rest of paper is built as follows. Firstly, data-driven artificial intelligence strategies are reviewed in Section 2, with the superiority of deep gaining knowledge of strategies outlined. Next, the challenge and opportunistic need of deep studying in smart production are offered, and standard deep mastering models are in short mentioned in Section 3. Then, the modern packages of deep learning techniques within the context of clever manufacturing are summarized in Section 4. Finally, the challenges as well as destiny developments of deep getting to know in clever manufacturing are discussed.

2. REVIEW OF LITERATURE

Artificial intelligence is considered as a essential manner to own intelligence, and indexed because the first vicinity in Gartner's Top 10 strategic technology tendencies in 2017 . Artificial intelligence has experienced numerous lifecycles, from the infancy period (Forties), via the primary upsurge duration (Sixties) and the second one upsurge period (Eighties), and the prevailing third boom duration (after 2000s). The improvement trend and standard artificial intelligence models are summarized in Table 1.The foundation of Artificial Neural Network started out lower back in 1940s, while MP version and Hebb rule were proposed to talk about how neurons worked in human brain. At the workshops in Dartmouth College, great synthetic intelligence talents like playing chess games and fixing simple logic problems have been evolved. The pioneering paintings brought artificial intelligence to the first upsurge duration (Nineteen Sixties).

In 1956,a mathematical model named Perceptron changed into proposed to simulate the apprehensive system of human getting to know with linear optimization. Next, a network version called Adaptive Linear Unit turned into advanced in 1959 and have been efficaciously utilized in practical programs which includes communication and climate forecasting.The dilemma of early artificial intelligence become additionally criticized because of the difficulty in dealing with non-linear problems, together with XOR (or XNOR) class.

With the improvement of Hopfield community circuit [28], synthetic intelligence advanced to the second one upsurge (Nineteen Eighties). Back Propagation (BP) set of rules was proposed to clear up non-linear problems in complex neural community in 1974 .A random mechanism

Table 1 List of typical artificial intelligence models

Timeline	Proposed models
Infancy period (1940s)	MP model Hebb rule
First upsurge period (1960s)	Perceptron Adaptive Linear Unit
Second upsurge period (1980s)	Hopfield network circuit Back Propagation Boltzmann Machine Support Vector Machine Restricted Boltzmann Machine Auto Encoder
Third boom period (after 2000s)	Recurrent Neural Network Long short-term Memory Convolutional Neural Network Deep Belief Network Deep Auto Encoder Sparse Auto Encoder Deep Boltzmann Machine Denosing Auto Encoder Deep Convolutional Neural Network Generative Adversarial Network Attention-based LSTM

Changed into delivered into Hopfield community and recommend the Boltzmann Machine (BM) in 1985 . With the development of statistical studying, Support Vector Machine (SVM) was advanced with kernel capabilities transformation in 1997, and showed first rate performance on category and regression . However, those traditional device studying strategies require human knowledge for feature extraction to lessen the size of enter, and as a consequence their performance exceptionally is based at the engineered features. The delivery of deep gaining knowledge of blessings now not handiest from the rich accumulation of traditional machine mastering strategies, however additionally the muse of statistical getting to know. Deep learning uses records representation studying as opposed to explicit engineered functions to perform responsibilities. It transforms data into summary representations that permit the functions to be learnt.

In 1986, Restricted Boltzmann Machine (RBM) turned into developed with the aid of obtaining the possibility distribution of Boltzmann Machine , and the hidden layers were used as characteristic vectors to represent the input statistics. Meanwhile, Auto Encoder (AE) changed into proposed using the layer-by means of-layer Greedy gaining knowledge of set of rules to reduce the loss function . In 1995, a neural network with directed topology connections among neurons, called Recurrent Neural Network (RNN), was proposed for function mastering from sequence records .In 1997, an progressed model of recurrent neural network, named Long short-term Memory (LSTM), became proposed to address the vanishing gradient hassle and cope with complex time sequence statistics.

In 1998, Convolutional Neural Network (CNN) changed into recommend to deal with dimensional inputs (e.G. Photograph), in which capabilities gaining knowledge of have been done by way of stacking convolutional layers and pooling layers .As the hierarchical structures of deep mastering models getting deeper, version education and parameter optimization emerge as more hard and time eating, even leading to over fitting or local optimization problems. Many tries have been made to expand deep getting to know fashions, but no quality performance changed into reported earlier than 2006.

Deep Belief Network (DBN) became advanced and achieved fulfillment in 2006.It allowed bidirectional connections in top layer best instead of stacking RBMs directly to lessen computational complexity, and the parameters have been efficaciously learned via layer-smart pre-training and high-quality tuning. Meanwhile, Deep Auto Encoder turned into proposed by using including greater hidden layers to deal with high nonlinear enter . The version parameters were first off pre-trained the use of a greedy layer-with the aid of-layer unsupervised gaining knowledge of set of rules and then first-class-tuned the use of BP algorithm. One 12 months later, Sparse Auto Encoder (SAE) became placed for ward to lessen dimensionality and analyze sparse representations.

Deep gaining knowledge of gained growing popularity. In 2009, Deep Boltzmann Machine with a bidirectional structure turned into proposed to study ambiguous input facts robustly, and the model parameters had been optimized using layer-smart pre-schooling . In 2010, Denoising Auto Encoder changed into supplied to reconstruct the stochastically corrupted input information, and pressure the hidden layer to discover greater robust functions . Deep Convolutional Neural Network (DCNN) become delivered with deep structure of Convolutional Neural Network in 2012 , and it confirmed superior performance in photograph reputation. Generative Adversarial Network (GAN) turned into proposed in 2014 , and it contained two unbiased models performing as adversaries. The generative version became designed to generate random samples just like real samples while the discriminative version changed into used for training and class with each actual and generated random samples.

In 2016, an interest-based LSTM version changed into proposed by way of integrating attention mechanism with LSTM . Nowadays, more and more new models are being advanced even according to week.

2.1 Comparison between deep learning and traditional machine learning

Both deep gaining knowledge of and traditional device learning are records driven artificial intelligence techniques to model the complicated courting between input and output as proven in Fig. 2. In addition to the excessive hierarchical structure, deep learning also has extraordinary attributes over traditional device getting to know in phrases of feature getting to know, model production, and version training. Deep mastering integrates feature mastering and version production in a single version via selecting extraordinary kernels or tuning the parameters via stop to cease optimization. Its deep architecture of neural nets with many hidden layers is essentially multi-level non-linear operations. It transfers every layer's illustration (or functions) from original input into greater abstracted representation in the higher layers to discover the complex inherent systems. For example, the functions which includes area, corner, contour, and object components, are abstracted layer-by way of-layer from an image. These abstracted function representations are then enter to the classifier layer to perform type and regression duties. Overall, deep getting to know is an give up-to-end studying shape with the minimum human inference, and the parameters of deep getting to know version are educated together.

On the contrary, traditional system gaining knowledge of plays function extraction and version construction in a separated manner, and each module is constructed step-by-step. The hand made capabilities are first off extracted by means of reworking uncooked information right into a distinct domain (e.G., statistical, frequency, and time-frequency area) to take the representative statistics requiring expert area knowledge. Next, function choice is accomplished to enhance the relevancy and decrease the spurious redundancy amongst features before feeding into the device mastering version. Traditional gadget getting to know techniques typically has shallow structures with at most 3 layers (e.G. Input, output, and one hidden layer). Thus, the performance of the built model no longer best is predicated on the optimization of adopted algorithms (e.G. BP Neural Network, Support Vector Machine, and logistic regression), however also is heavily stricken by the handcrafted features. Generally, the function extraction and selection are time-ingesting, and highly rely upon domain information.

Therefore, deep mastering has exclusive distinction with conventional machine getting to know strategies as illustrated in Table 2. The high stage summary representation in function studying makes deep studying greater flexible and adaptable to data variety. Because the statistics are abstracted, the numerous records kinds and assets do now not have Sturdy have an impact on at the evaluation results. On the alternative hand, the deep hierarchical structure in deep gaining knowledge of is less complicated to version then on linear relationship using compositional feature evaluating with the shallow structure which is regarded as a everyday function in traditional system getting to know. The superiority of deep community have been established mathematically in . As the scale and sort of dataset grow inside the huge facts context, it will become extra difficult to create new, notably relevant features.

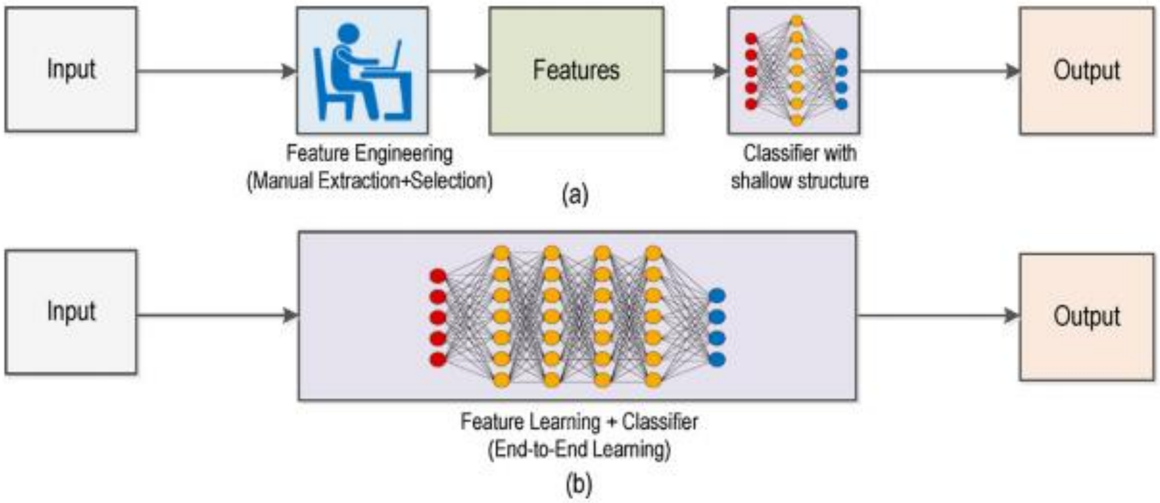

Fig. 2. Comparison between two techniques: a) traditional machine learning, b) deep learning.

Table 2 Comparison between traditional machine learning and deep learning

	Feature learning	Model construction	Model training
Traditional machine learning	Explicit engineered features extracted with expert domain knowledge.	Use extracted features to construct data-driven model, usually with shallow structures.	Each module is trained step-by-step.
Deep learning	Features are learned by transforming data into abstract representations.	An end-to-end high hierarchical model structure with nonlinear combination of multi-layers.	Parameters are trained jointly.

In the context of big information generation in smart production, the capability to keep away from function engineering is regarded as a tremendous advantage because of the demanding situations associated with this process.

3. RESEARCH METHODOLOGY

3.1 Deep Studying for Smart Manufacturing

With new technologies (e.G. IoT, massive facts) embraced in smart production, smart facilities recognition on developing manufacturing intelligence that can have a high quality effect across the complete organization. The production these days is experiencing an unprecedented boom in available sensory information comprised of various formats, semantics, and structures. Sensory information turned into accrued from distinct elements across the producing organization, inclusive of product line, production system, production manner, labor activity, and environmental situations. Data modelling and analysis are the vital a part of smart manufacturing to managing improved high volume information, in addition to assisting actual-time data processing.

From sensory statistics to manufacturing intelligence, deep mastering has attracted tons attention as a breakthrough of computational intelligence. By mining know-how from aggregated facts, deep gaining knowledge of techniques play a key role in automatically getting to know from facts, identifying patterns, and making selections as proven in Fig. Three. Different ranges of statistics analytics can be produced including descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. Descriptive analytics objectives to summarize what takes place by taking pictures the product's situations, surroundings and operational parameters. When the product overall performance is reduced or the equipment failure happens, diagnostic analytics study the foundation cause and document the cause it takes place. Predictive analytics utilizes statistical fashions to make predictions about the possibility of future manufacturing or system degradation with available historical data. Prescriptive analytics is going beyond by using recommending one or more courses of motion. Measures can be identified to enhance production results or correct the problems, showing the likely final results of every choice.

With the superior analytics supplied through deep gaining knowledge of, production is transformed into noticeably optimized smart facilities. The advantages consist of decreasing running expenses, preserving up with converting purchaser demand, improving productivity and lowering downtime, gaining better visibility and extracting greater value from the operations for globally competitiveness.

Up to this point, diverse deep gaining knowledge of architectures had been advanced and the applicable studies subjects are rapid-growing. To facilitate the investigation of producing intelligence, numerous standard deep gaining knowledge of architectures are mentioned inclusive of Convolutional Neural Network, Restricted Boltzmann Machine, Auto Encoder, and Recurrent Neural Network and their variants. The feature learning capability and model production mechanism had been emphasized considering those models are the constructing blocks to construct complete and complex deep studying techniques.

3.2 Convolutional neural network

Convolutional Neural Network (CNN) is a multi-layer feed forward artificial neural community that's first of all proposed for two-dimensional picture processing .It has additionally been investigated for one-dimensional sequential information analysis such as natural language processing and speech popularity recently . In CNN, the feature studying is achieved by way of alternating and stacking convolutional layers and pooling operations. The convolutional layers convolve with uncooked enter statistics the usage of more than one neighborhood kernel filters and generate invariant nearby capabilities. The subsequent pooling layers extract the most sizeable capabilities with a set length over sliding windows of the raw enter facts by using pooling operations which includes max pooling and common pooling. Max pooling selects the most fee of 1 vicinity of the feature map as the most good sized feature. Average pooling calculates the imply cost of one vicinity and takes it as the pooling cost of this place. Max pooling is nicely appropriate to extract sparse features, whilst pooling operation on all samples may not be most desirable. After multi-layer feature gaining knowledge of, fully-linked layers convert a - dimensional function map right into a one dimensional vector and then feed it into a soft max function for version construction. By stacking convolutional layers, pooling layers, and completely-related layers, an ordinary CNN is constructed as proven in Fig. Four. Gradient based totally back propagation is normally used to educate convolutional neural network by using minimizing the minimum suggest squared mistakes or pass-entropy loss feature. CNN has the fine houses which includes sparse interactions with neighborhood connectivity, parameter sharing with reduced numbers, and equivariant illustration which is invariant to object locations.

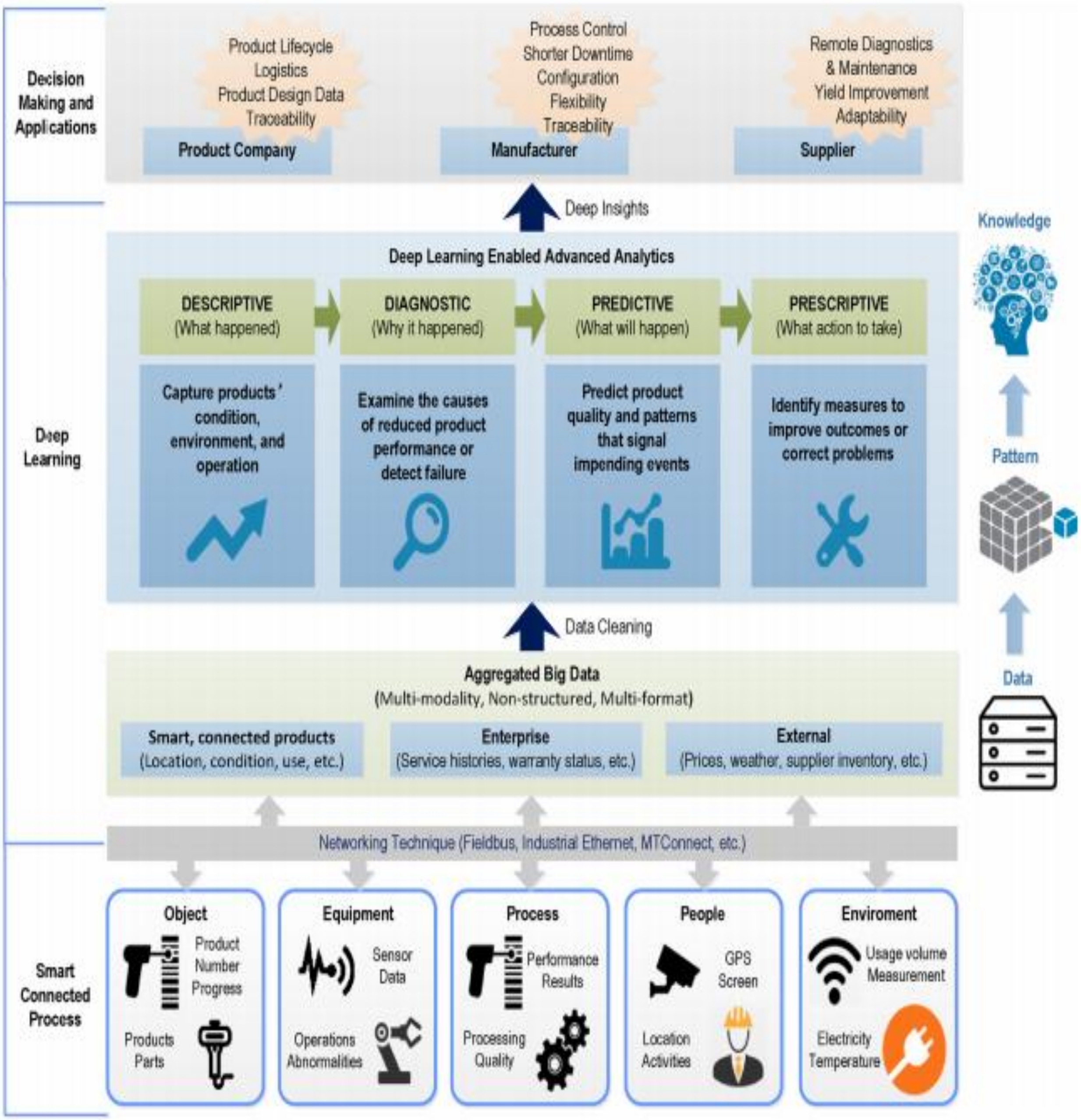

Fig. 3. Deep learning enabled advanced analytics for smart manufacturing.

3.3 Restricted Boltzmann System and its Versions

Restricted Boltzmann Machine (RBM) is a two-layer neural community which includes seen and hidden layer. There exists a symmetric connection between seen and hidden gadgets, however there are not any connections among every neuron inside the equal layer. It is an energy based version wherein the visible layer is used to input records at the same time as the hidden layer is used to extract functions. All hidden nodes are assumed conditionally independent. The weights and offsets of those layers are tuned over iterations so that it will make the output of the seen layer as the approximation of the unique input. Finally, the hidden layers are appeared as special representations of the seen layer. The parameters in hidden layers are handled as the capabilities to symbolize the enter records to comprehend statistics coding and size discount.Then, supervised gaining knowledge of methods which include logistic regression, Naïve Bayes, BP Neural Network, and Support Vector Machine, and many others.May be used to put into effect data class and regression. RBM takes the benefits of extracting required functions from schooling datasets routinely, which avoids the nearby minimal cost and for this reason has received a growing quantity of attentions. Utilizing RBM as the fundamental getting to know module, different version models had been advanced . Deep Belief Network (DBN): DBN is built with the aid of stacking a couple of RBMs, wherein the output of the l th layer in hidden units is used as the enter of the (l + 1)th layer in visible units. For DBN schooling, a quick greedy set of rules is normally used to initialize the network and

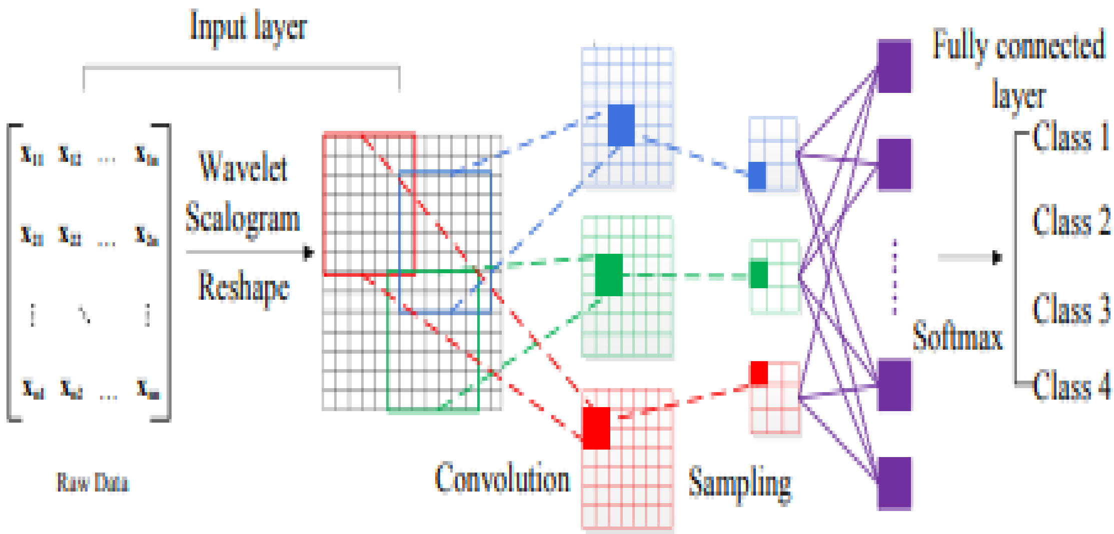

Fig. 4. Architecture of convolutional neural network model.

The parameters of this deep structure are then first-rate-tuned by way of a contractive wake-sleep algorithm .Bayesian Belief Network is carried out to the location that is close to the visible layers, and RBMs are used to the area some distance from the visible layers. That is to say, the best layers are undirected and the opposite lower layers are directed, as proven in Fig. 5. Deep Boltzmann Machine (DBM): DBM can be regarded as a deep established RBMs wherein hidden devices are grouped right into a hierarchy of layers.

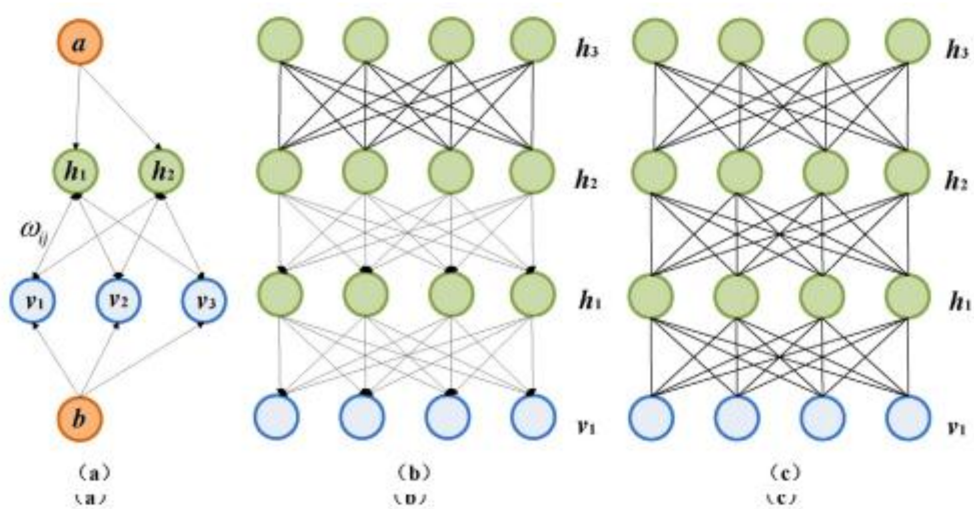

Fig. 5. Architecture of (a) RBM, (b) DBN, and (c) DBM.

The full connections among two adjoining layers are enabled, however no connection is authorized within a layer or between non-neighbouring layers as proven in Fig. Five. By stacking multi-RBMs, DBM can research complex structures and construct excessive-stage illustration of enter information . Compared to DBN, DBM is a fully undirected graphical model even as DBN is a mixed directed/undirected one. Accordingly, the DBM model is trained together and more computationally expensive. On the opposite, DBN may be skilled layer-accurately to be extra correctly.

3.4 Recurrent Neural Network and its Variants

Compared with conventional neural networks, Recurrent Neural Network (RNN) has unique characteristic of topology connections among the neurons formed directed cycles for series facts as shown in Fig. 7. Thus, RNN is suitable for feature getting to know from sequence statistics. It lets in statistics persists in hidden layers and captures preceding states of some time steps ago. An updated rule is implemented in RNN to calculate the hidden states at extraordinary time steps. Take the sequential enter as a vector, the present day hidden country can be calculated through two elements thru a same activation function (e.G. Sigmoid or tan h function). The first part is calculated with the input while the second one component is acquired from the hidden country on the preceding time step. Then, the goal output may be calculated with the present day hidden country via a gentle max characteristic. After processing the whole collection, the hidden country is the learned illustration

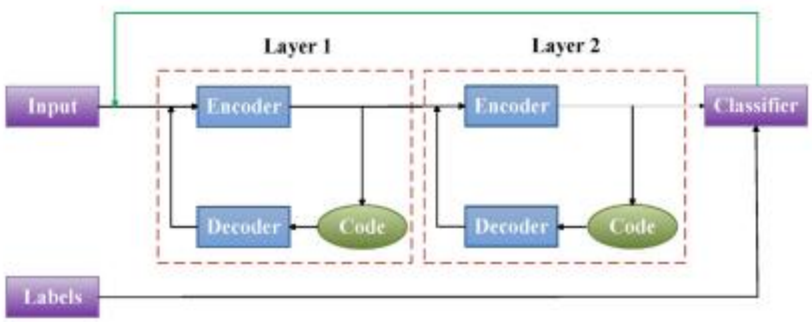

Fig. 6. The architecture of AE.

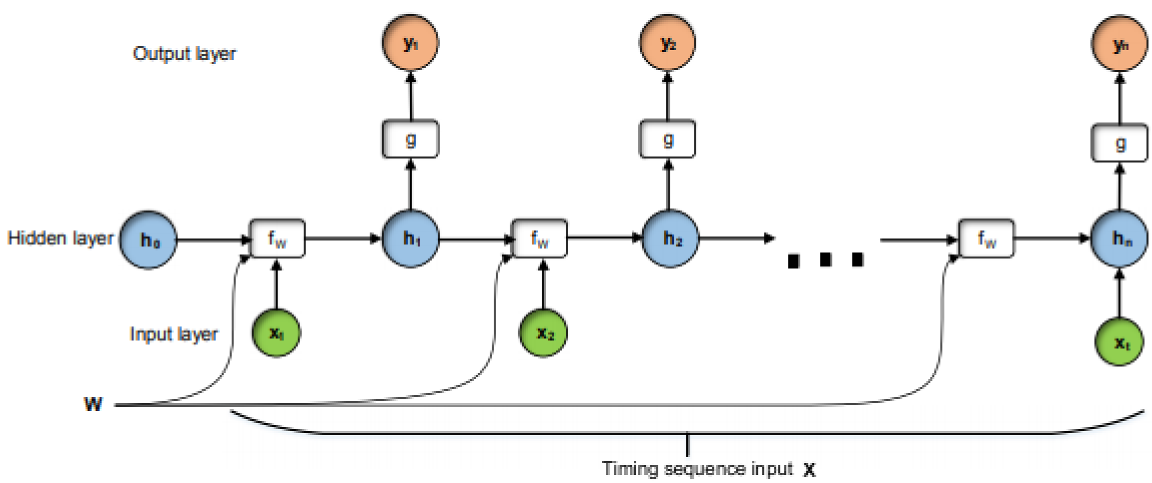

Fig. 7. Architecture of recurrent neural network model.

Of the enter sequential information and a conventional multilayer perceptron (MLP) is brought on top to map the acquired representation to objectives. Different from conventional neural networks, the version education in RNN is done via Back propagation Through Time (BPTT). RNN is first off unrolled in step with time and every unrolled time step is taken into consideration as an additional layer. Then lower back propagation set of rules is implemented to calculate gradients. Due to the vanishing/exploding gradient problem using BPTT for version education, RNN can't capture long-time period dependencies. In different words, RNN has trouble in handling lengthy-time period collection facts. A sort of enhancements are proposed to remedy these troubles, amongst which lengthy quick-time period reminiscence (LSTM) is broadly investigated for its effectiveness. The most crucial concept of LSTM is cell state, which permits statistics waft down with linear interactions. Comparing with single recurrent structure in RNN, the gates including overlook gate layer, enter gate layer and output gate layer, are utilized in LSTM to govern the mobile nation. It permits every recurrent unit to adaptively seize lengthy-term dependencies of different time scales.

4. DATA ANALYSIS

4.1 Applications to Smart Manufacturing

Computational intelligence is an crucial a part of clever production to permit correct insights for higher choice making. Machine getting to know has been broadly investigated in one-of-a-kind tiers of manufacturing lifecycle overlaying concept, layout , evaluation, production, operation, and sustainment as proven in Fig. 8. The programs of data mining in manufacturing engineering are reviewed in , overlaying different categories of manufacturing processes, operations, fault detection, maintenance, selection help, and product first-rate improvement. The evolution and destiny of producing are reviewed in , emphasizing the importance of data modelling and analysis in manufacturing intelligence. The utility schemes of gadget getting to know in manufacturing are diagnosed as summarized in .Smart production also calls for prognostics and health management (PHM) skills to meet the current and destiny needs for efficient and reconfigurable production. Deep learning, as an rising technique, has been investigated for a extensive variety of producing systems recently. To supply a top level view, the applications of latest deep mastering techniques in manufacturing are discussed on this have a look at, especially inside the areas of product nice inspection, fault analysis, and disorder analysis, as highlighted in Table.

Table 3 Comparison between different deep learning models.

Model	Principle	Pros.	Cons.
CNN	Abstracted features are learned by stacked convolutional and sampling layers.	Reduced parameter number, invariance of shift, scale and distortion	High computational complexity for high hierarchical model training
RBM	Hidden layer describes variable dependencies and connections between input or output layers as representative features.	Robust to ambiguous input and training label is not required in pre-training stage	Time-consuming for joint parameter optimization
AE	Unsupervised feature learning and data dimensionality reduction are achieved through encoding	Irrelevance in the input is eliminated, and meaningful information is preserved	Error propagation layer-by-layer and sparse representations are not guaranteed
RNN	Temporal pattern stored in the recurrent neuros connection and distributed hidden states for time-series data.	Short-term information is retained and temporal correlations are captured in sequence data.	Difficult to train the model and save the long-term dependence

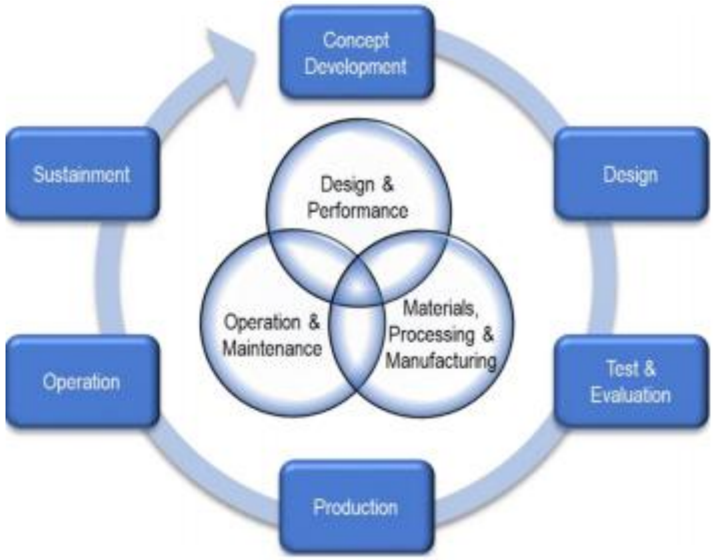

Fig. 8. Typical application scenarios of machine learning in smart manufacturing.

4.1.1 Descriptive analytics for product high-quality inspection

Surface integration inspection is commonly inspected using device imaginative and prescient and picture processing techniques to come across Table 5 A list of deep mastering fashions with packages. Deep studying version Application Scenarios Reference CNN Surface integration inspection [Machinery fault analysis DBN Machinery fault prognosis Predictive analytics & disorder prognosis AE Machinery fault diagnosis RNNs Predictive analytics & disorder diagnosis surface illness for superior product first-rate in production . Traditional system gaining knowledge of has made exquisite development and yields dependable consequences in lots of instances ,but exceptional pre-processing procedures inclusive of structural-based, statistical based, filter out-based, and version based totally techniques are needed to extract consultant capabilities with expert know-how . However, flexible configuration in current manufacturing gadget may want to shift manufacturing from one product to every other quick. The function illustration may need redesign from scratch for classic gadget learning. Additionally, a brand new product can also gift complex texture patterns or intensity variations, and the floor defects could be in an arbitrary length, orientation and form. Therefore, manually designed capabilities in conventional device mastering technique might also lead to inadequate or unsatisfactory inspection overall performance in complicated surface scenarios or dynamic changing technique. To deal with these challenges, deep getting to know has been investigated to examine excessive-stage regular capabilities and implemented to a huge variety of textures or tough-to-locate defects instances. Convolutional Neural Network, firstly designed for photo analysis, is nicely in shape for automatic illness identity in surface integration inspection. In , a Deep Convolutional Neural Network architecture is designed and the hyper-parameters are optimized based totally on back propagation and stochastic gradient descent algorithms. A max-pooling Convolutional Neural Network is offered in to perform characteristic extraction directly from the pixel illustration of metallic disorder pix and shows decrease error costs comparing with multi-layer perceptron and guide vector gadget. The image evaluation is studied with convolutional neural community in to robotically look at dirties, scratches, burrs, and wears on floor parts. The experimental consequences display that CNN works well with one-of-a-kind styles of defects on textured or non-textured surfaces. A popular technique based on CNN is proposed in to extract patch characteristic and are expecting disorder place thru thresholding and segmenting. The results display the pertained CNN version works well on small dataset with advanced accuracy for computerized floor inspection device.

4.1.2 Predictive analytics for defect prognosis

In order to boom production productiveness at the same time as lowering upkeep cost, it's far crucial to increase and implement an wise maintenance strategy that permits manufacturers to decide the condition of in-service systems on the way to predict when upkeep need to be done. The temporal behaviour in the historical facts is critical for prediction, and deep recurrent neural community has proven its capability to model temporal pattern. Recently, a popular recurrent neural network, named lengthy short time period memory, has been investigated to are expecting disorder propagation and estimate final beneficial lifestyles (RUL) of mechanical systems or additives. In,a competitive gaining knowledge of-based RNN has been proposed for lengthy-time period analysis of rolling bearing fitness repute. In,a new nearby characteristic-primarily based gated recurrent unit network has been proposed to learn the representation of the sequence of nearby features and the proposed approach is validated on 3 real gadget fitness tracking tasks. In , an incorporated approach of CNN and bi-directional LSTM is presented for machining device wear prediction, in which CNN is used to extract nearby capabilities from sequential signals and bi-directional LSTM to capture long-term dependence for prediction. Vanilla LSTM is investigated in to estimate the closing beneficial existence of an plane turbofan engine below complex operating situations and sturdy history noise, and the experimental effects verify that Vanilla LSTM affords good prediction accuracy. A stacked LSTM community allows the getting to know of higher level temporal functions, and has been presented for anomaly prediction of space trip and engine.

Deep Belief Network, as the feature getting to know approach in regression models, has also been investigated for predictive analytics. Deep Belief Network is investigated to version the complex relationship between cloth elimination fee and chemical mechanical polishing method parameters in semiconductor manufacturing. An integrative technique of Deep Belief Network and particle clear out is offered in for the RUL prediction of a ceramic bearing. By aggregating the output of ensemble DBNs, Support Vector Regression model is investigated to are expecting electricity load call for . To expect the useful resource request in cloud computing, DBN is proposed in to optimize task agenda and stability the computational load.

4.1.3 Data matter

As the evolution of clever manufacturing, increasingly more machineries are geared up with smart sensors and meshed with Internet of Things. Currently, most companies do not understand what to do with the information they've, and that they lack software and modelling to interpret and analyse them. On the other hand, producers want practical steerage to improve their tactics and products, while the academics increase updated artificial intelligence fashions with out considering how they'll be carried out in exercise. As manufacturing technique turns into extra complicated, extra problem comes along to clean the records and formulate the right troubles to version. Five gaps are diagnosed in clever production innovation together with adopted techniques, progressed information collection, use and sharing, predictive model design, generalized predictive fashions, and related factories and manage tactics . To meet the excessive call for of superior analytics in clever production, deep gaining knowledge of with feature learning and deep network gives terrific capacity and indicates positive residences. To handle overwhelming statistics characterised with the aid of high-quantity, highvelocity and excessive-range,there are nonetheless some demanding situations associated with production enterprise to undertake, put in force, and deploy deep learning for actual-international programs. To deal with the challenges, the destiny improvement trends of deep gaining knowledge of for clever production are discussed in phrases of records remember, version selection, version visualization, prevalent version, and incremental learning.

A common presumption in device mastering is that algorithms can research better with more statistics, and thus the overall performance of deep mastering version heavily relies upon on the dimensions and excellent of datasets. So some distance deep mastering indicates the effectiveness when it's far carried out to limited varieties of statistics (e.G. Pics, speech, and vibration, and many others.) and properly-described responsibilities. Multi-sensory has been instrumented to capture statistics at all ranges of a product's lifestyles. Deep mastering algorithm can be infeasible to immediately deal with such excessive dimensional, multimodality, and non-dependent statistics, or even at risk of the curse of dimensionality. Extracting the relevant facts to lessen the dimensions and making use of appropriating mission-unique regularization time period may additionally enhance the performance of deep learning. On the other hand, the magnificence imbalance trouble is some other undertaking. The magnificence follows a noticeably-skewed distribution in real existence, representing maximum facts samples belong to few categories. For example, the dataset of floor defects is typically too small and steeply-priced to collect. The ratio of correct to horrific elements is fairly imbalanced ranging from nine:1 to even much less than 1,000,000. Thus, it's far tough to use widespread type techniques to differentiating precise elements from scraps. Appropriate measures along with elegance re-sampling, cost-touchy training, and integration of boot strapping may be essential for deep studying model to address magnificence imbalance troubles

4.1.4 Incremental learning

The deep getting to know algorithms aren't basically constructed to analyze incrementally and are consequently vulnerable to the facts speed troubles. For a new trouble setup, deep gaining knowledge of might also want to rebuild the version from scratch and the existing knowledge may be difficult to make use of. Additionally, the facts within the new event ualities is likewise an problem. It is essential to permit deep studying with incremental gaining knowledge of talents. Transfer studying targets to extract the knowledge from one supply task after which applies the found out know-how to a extraordinary but related undertaking .It could rent the pre-trained deep learning model from a relevant task for version initialization and first-class-tuning to enable understanding reuse and updating as transferred deep learning. Some preceding works focusing on transferred function extraction/dimensionality reduction were done. A most imply discrepancy (MMD) degree evaluating the discrepancy among supply and goal domains is introduced into the target characteristic of deep neural networks .Thus, transferred deep learning is meaningful and promising for clever manufacturing to permit expertise updating and intelligence upgrading.

5. CONCLUSION

Deep getting to know provides superior analytics and gives remarkable potentials to smart manufacturing inside the age of massive information. By unlocking the unparalleled amount of facts into actionable an in sightful facts, deep gaining knowledge of gives choice-makers new visibility into their operations, as well as actual-time overall performance measures and costs. To facilitate advanced analytics, a comprehensive assessment of deep gaining knowledge of strategies is offered with the programs to smart production. Four usual deep mastering models along with Convolutional Neural Network, Restricted Boltzmann Machine, Auto Encoder, and Recurrent Neural Network are mentioned in element. The emerging studies effort of deep gaining knowledge of in applications of manufacturing is likewise summarized. Despite of the promising results pronounced up to now, there are nevertheless a few barriers and substantial demanding situations for in addition exploration. As the evolution of computing sources (e.G., cloud computing, fog computing, etc.), computational intelligence including deep gaining knowledge of can be push into cloud, allowing extra convenient and on-call for computing offerings for smart manufacturing.

6. ACKNOWLEDGEMENTS

Manju Susan Thomas deeply acknowledge the encouragement received from the management of SIAS, Vazhayoor.

7. REFERENCES

1. Putnik G, Sluga A, ElMaraghy H, Teti R, Koren Y, Tolio T, et al. Scalability in manufacturing systems design and operation: state-of-the-art and future developments roadmap. CIRP Ann Manuf Technol 2013;62(2):751–74.
2. Lee YT, Kumaraguru S, Jain S, Hatim Q, Robinson S, Helu M, et al. A classification scheme for smart manufacturing systems' performance metrics. Smart Sustain Manuf Syst 2017;1(1):52–74.
3. Hu T, Li P, Zhang C, Liu R. Design and application of a real-time industrial Ethernet protocol under Linux using RTAI. Int J Comput Integr Manuf 2013;26(5):429–39.
4. Ye Y, Hu T, Zhang C, Luo W. Design and development of a CNC machining process knowledge base using cloud technology. Int J Adv Manuf Technol 2016:1–13.
5. Tao F, Qi Q. New IT driven service-oriented smart manufacturing: framework and characteristics. IEEE Trans Syst Man Cybern Syst 2017;99:1–11.
6. Ang J, Goh C, Saldivar A, Li Y. Energy-efficient through-life smart design, manufacturing and operation of ships in an industry 4.0 environment. Energies 2017;10(5):610.
7. Huang Z, Hu T, Peng C, Hou M, Zhang C. Research and development of industrial real-time Ethernet performance testing system used for CNC system. Int J Adv Manuf Technol 2016;83(5–8):1199–207.
8. Lalanda P, Morand D, Chollet S. Autonomic mediation middleware for smart

manufacturing. IEEE Internet Comput 2017;21(1):32–9.

9. Smart Manufacturing Coalition. Manufacturing growth continues despite uncertain economy, according to ASQ outlook survey; 2013. https:// smartmanufacturingcoalition.org/sites/default/files/12.16.13 manufacturing outlook survey.pdf. [Accessed 10 Sepember 2017].

10. Wang L, Törngren M, Onori M. Current status and advancement of cyber-physical systems in manufacturing. J Manuf Syst 2015;37:517–27.

11. Wang P, Gao RX, Fan Z. Cloud computing for cloud manufacturing: benefits and limitations. J Manuf Sci Eng 2015;137:1–10.

12. Lu Y, Xu X, Xu J. Development of a hybrid manufacturing cloud. J Manuf Syst 2014;33(4):551–66.

13. Wu D, Rosen DW, Schaefer D. Cloud-based design and manufacturing: status and promise. Comput Aided Des 2015;59:1–14.

14. Choudhary AK, Harding JA, Tiwari MK. Data mining in manufacturing: a review based on the kind of knowledge. J Intell Manuf 2009;20(5):501–21.

15. Lade P, Ghosh R, Srinivasan S. Manufacturing analytics and industrial internet of things. IEEE Intell Syst 2017;32(3):74–9.

A Study on Big Data Analytics: Technologies & Tools.

Mr.Chandrakant R. Gujar
Assistant Professor, MCA
W.K.B.S. Mandal's
Dr. Suryakanta R. Ajmera
MCA College for
Women,Deopur Dhule, MS
India

Abstract: Data is a very valuable asset in the world today. The economics of data are based on the idea that data value can be extracted through the use of analytics. Through Big data and analytics are still in their initial growth stage, their importance cannot be undervalued. As big data starts to expand and grow, Importance of big data analytics will continue to grow in everyday lives, both personal and business. In addition, the size and volume of data is increasing every single day, making it important to address the manner in which big data is addressed every day. A huge repository of terabytes of data is generated every day from modern information systems and digital technologies such as Internet of Things and cloud computing. Analysis of these massive data requires a lot of efforts at multiple levels to extract knowledge for decision making. Therefore, big data analysis is a current area of research and development. The basic objective of this paper is to explore the potential impact of big data technologies and challenges associated with it. This paper provides a platform to explore big data at numerous stages. Additionally, it opens a new horizon for researchers to develop the solution, based on the challenges and open research issues.

Keywords: Big Data, Big Data technologies, Big Data Challenges, Hadoop, Big Data analytic.

Introduction

What is Data?

The quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.[1]

What is Big Data?

Big Data is also **data** but with a **huge size**. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.[1]

Examples Of Big Data

Following are some the examples [1] of Big Data-

1) The New York Stock Exchange generates about *one terabyte* of new trade data per day.
2) Social Media: -The statistic shows that *500+terabytes* of new data get ingested into the databases of social media site Facebook, every day. This data is mainly generated in terms of photo and video uploads, message exchanges, putting comments etc.
3) A single Jet engine can generate *10+terabytes* of data in *30 minutes* of flight time. With many thousand flights per day, generation of data reaches up to many *Petabytes.*

What Comes Under Big Data?

Big data involves the data produced by different devices and applications. Given below are some of the fields that come under the umbrella of Big Data [2].

- **Black Box Data** − It is a component of helicopter, airplanes, and jets, etc. It captures voices of the flight crew, recordings of microphones and earphones, and the performance information of the aircraft.
- **Social Media Data** − Social media such as Facebook and Twitter hold information and the views posted by millions of people across the globe.
- **Stock Exchange Data** − The stock exchange data holds information about the 'buy' and 'sell' decisions made on a share of different companies made by the customers.
- **Power Grid Data** − The power grid data holds information consumed by a particular node with respect to a base station.
- **Transport Data** − Transport data includes model, capacity, distance and availability of a vehicle.
- **Search Engine Data** − Search engines retrieve lots of data from different databases.

Types of Big Data

- **Structured data** − Relational data.
- **Semi Structured data** − XML data.
- **Unstructured data** − Word, PDF, Text, Media Logs.

Benefits of Big Data

- Using the information kept in the social network like Facebook, the marketing agencies are learning about the response for their campaigns, promotions, and other advertising mediums.
- Using the information in the social media like preferences and product perception of their consumers, product companies and retail organizations are planning their production.
- Using the data regarding the previous medical history of patients, hospitals are providing better and quick service.

Big Data Technologies

Big data technologies are important in providing more accurate analysis, which may lead to more concrete decision-making resulting in greater operational efficiencies, cost reductions, and reduced risks for the business.

There are various technologies in the market from different vendors including Amazon, IBM, Microsoft, etc., to handle big data. While looking into the technologies that handle big data, we examine the following two classes of technology −

a) Operational Big Data

This includes systems like MongoDB that provide operational capabilities for real-time, interactive workloads where data is primarily captured and stored.

NoSQL Big Data systems are designed to take advantage of new cloud computing architectures that have emerged over the past decade to allow massive computations to be run inexpensively and efficiently. This makes operational big data workloads much easier to manage, cheaper, and faster to implement.

Some **NoSQL** systems can provide insights into patterns and trends based on real-time data with minimal coding and without the need for data scientists and additional infrastructure.

b) Analytical Big Data

These includes systems like Massively Parallel Processing (MPP) database systems and **MapReduce** that provide analytical capabilities for retrospective and complex analysis that may touch most or all of the data.

MapReduce provides a new method of analyzing data that is complementary to the capabilities provided by SQL, and a system based on Map Reduce that can be scaled up from single servers to thousands of high and low end machines.

Big Data Challenges

The major challenges associated with big data are Capturing data, Curation, Storage, Searching ,Sharing, Transfer, Analysis & Presentation, To fulfill the above challenges, organizations normally take the help of enterprise servers.

Traditional Approach

In this approach, an enterprise will have a computer to store and process big data. For storage purpose, the programmers will take the help of their choice of database vendors such as Oracle, IBM, etc. In this approach, the user interacts with the application, which in turn handles the part of data storage and analysis.

Google's Solution

Google solved this problem using an algorithm called MapReduce. This algorithm divides the task into small parts and assigns them to many computers, and collects the results from them which when integrated, form the result dataset

Hadoop

Apache Hadoop is an open source software framework used to develop data processing applications which are executed in a distributed computing environment.

Applications built using HADOOP are run on large data sets distributed across clusters of commodity computers. Commodity computers are cheap and widely available. These are mainly useful for achieving greater computational power at low cost.

Similar to data residing in a local file system of a personal computer system, in Hadoop, data resides in a distributed file system which is called as a Hadoop Distributed File system. The processing model is based on 'Data Locality' concept wherein computational logic is sent to cluster nodes (server) containing data. This computational logic is nothing, but a compiled version of a program written in a high-level language such as Java. Such a program, processes data stored in Hadoop HDFS [3]

What is Big Data Analytics?

Big data analytics refers to the strategy of analyzing large volumes of data, or big data. This big data is gathered from a wide variety of sources, including social networks, videos, digital images, sensors, and sales transaction records. The aim in analyzing all this data is to uncover patterns and connections that might otherwise be invisible, and that might provide valuable insights about the users who created it. Through this insight, businesses may be able to gain an edge over their rivals and make superior business decisions.

Big Data Analytics Tools

Big Data Analytics software is widely used in providing meaningful analysis of a large set of data. This software helps in finding current market trends, customer preferences, and other information.[3], There are following latest tools available in market year 2020

Azure HDInsight is a Spark and Hadoop service in the cloud. It provides big data cloud offerings in two categories, Standard and Premium. It provides an enterprise-scale cluster for the organization to run their big data workloads.

Skytree is a big data analytics tool that empowers data scientists to build more accurate models faster. It offers accurate predictive machine learning models that are easy to use.

Talend is a big data tool that simplifies and automates big data integration. Its graphical wizard generates native code. It also allows big data integration, master data management and checks data quality.

Splice Machine is a big data analytic tool. Their architecture is portable across public clouds such as AWS, Azure, and Google.

Apache Spark is a powerful open source big data analytics tool. It offers over 80 high-level operators that make it easy to build parallel apps. It is used at a wide range of organizations to process large datasets.

Plotly is an analytics tool that lets users create charts and dashboards to share online

Apache SAMOA is a big data analytics tool. It enables development of new ML algorithms. It provides a collection of distributed algorithms for common data mining and machine learning tasks

Lumify is a big data fusion, analysis, and visualization platform. It helps users to discover connections and explore relationships in their data via a suite of analytic options.

Elasticsearch is a JSON-based Big data search and analytics engine. It is a distributed, RESTful search and analytics engine for solving numbers of use cases. It offers horizontal scalability, maximum reliability, and easy management.

R is a language for statistical computing and graphics. It also used for big data analysis. It provides a wide variety of statistical tests

ACKNOWLEDGMENTS

We are thankful to our President West Khandesh Bhagini Seva Mandal's Mr. Atulji R. Ajmera , Mrs.MeghaV.Kamerkar OS ,Director Dr. Rajeev B.Kharat, and all teaching staff /office staff of Dr. Suryakanta R. Ajmera MCA College for Women Deopur Dhule MS, India for motivation and encouragement which helped us to complete this research paper.

REFERENCES

[1] https://www.guru99.com/what-is-big-data.html#1

[2] https://www.tutorialspoint.com

[3] https://www.guru99.com/big-data-analytics-tools.html

Memory Effect on Decay Property of Solutions to Plate Equations

Yongqin Liu
School of Mathematics and Physics
North China Electric Power University
Beijing 102206, China

Xuehui Wu
School of Mathematics and Physics
North China Electric Power University
Beijing 102206, China

Abstract: In this paper we focus on the initial-value problem of linear plate equations with memory in multi-dimensions, the decay structure of which is of regularity-loss property. We obtain fundamental solutions by using Fourier transform and Laplace transform. By virtue of the point-wise estimate of solutions in the Fourier space, we gain estimates and properties of solution operators, by utilizing which decay estimates of solutions to the linear problem are obtained and the decay rate can be as large as desired if the initial data are sufficiently smooth.

Keywords: plate equation; memory; decay estimates; regularity-loss type; initial-value problem

1. INTRODUCTION

In this paper we consider the initial-value problem of the following linear plate equation with memory term in multi-dimensional space R^n ($n \ge 1$):

$$u_{tt} + (1+\Delta^2)u - g * u = 0, \qquad (1.1)$$

with the initial data

$$u(x,0) = u_0(x), u_t(x,0) = u_1(x). \qquad (1.2)$$

Here $u = u(x,t)$ is the unknown function of $x = (x_1,...,x_n) \in R^n$ and $t > 0$, and represents the transversal displacement of the plate at the point x and the time t. The term $-g * u = -\int_0^t g(t-\tau)u(\tau)d\tau$ accords with the memory term which reflects that the stress at an instant relies on the whole history of the strains the material has suffered, and g satisfies:

Assumption[A]:

a) $g \in C^2(R^+) \cap W^{2,1}(R^+)$,

b) $g(s) > 0, -C_0 g(s) \le g'(s) \le -C_1 g(s),$
$|g''(s)| \le C_2 g(s), \forall s \in R^+,$

c) $1 - \int_0^t g(s)ds \ge C_3, \forall t \in R^+,$

where $C_i (i = 1,2,3)$ are positive constants.

In [8], Liu and S. Kawashima learned the decay property of a semi-linear plate equation with memory-type dissipation, whose linear part is given by:

$$u_{tt} + \Delta^2 u + u + g * \Delta u = 0, \qquad (1.3)$$

here the dissipation is given by the memory term $g * \Delta u$. In that paper, the authors obtained the global existence and the optimal decay estimates of solutions by introducing a set of time-weighted Sobolev spaces and using the contraction mapping theorem. They also showed that the dissipative structure is characterized by the function

$$\rho_1(\xi) = \frac{|\xi|^2}{1+|\xi|^4},$$

here $\rho_1(\xi)$ is introduced in the point-wise estimate in the Fourier space of solutions to the corresponding linear problem. $\rho_1(\xi)$ decides that the energy restricted in the either lower-frequency or higher-frequency domains decays polynomially and the decay property is of regularity-loss type

In [13], Liu and W. Wang studied the point-wise estimate of solutions to a dissipative wave equation

$$u_{tt} - \Delta u + u_t = 0, \qquad (1.4)$$

and they showed that the dissipative structure in (1.4) is characterized by the function $\rho_2(\xi) = \frac{|\xi|^2}{1+|\xi|^2}$.

This $\rho_2(\xi)$ determines that the energy restricted in the lower-frequency domain decays polynomially and exponentially in the higher-frequency domain and the decay property is of standard type instead of regularity-loss type. For more studies of such decay structure, we refer to [5, 6, 7, 15, 16, 17].

To have a better comparison of the dissipative structures, we study the equation (1.1).

Same as the above memory plate equation (1.3), the plate equation (1.1) is also of regularity-loss property. The decay structure of the regularity-loss type is characterized by the property $\rho(\xi) = \frac{1}{1+|\xi|^4}$, where $\rho(\xi)$ is introduced in the point-wise estimate in the Fourier space (3.1) of solutions to the linear problem. It is obvious that the decay structure is very weak in the higher-frequency domain since $\rho(\xi) \to 0$ as $\xi \to \infty$. In fact $\rho(\xi)$ determines that the energy restricted in the lower-frequency domain decays exponentially and polynomially in the higher-frequency domain and the decay property is of regularity-loss type. There is one point worthy to be mentioned. The solutions in [8], [13] and this paper all decay polynomially. However, the dissipative structures are different. The decay rates in [8] and [13] are fixed, while the decay rate in this paper can be as large as desired if only the initial data are sufficiently smooth. For more studies on aspects of dissipation of plate equations, we refer to [1, 2, 9, 10]. Also, as for the study of decay properties for hyperbolic systems of memory-type dissipation, we refer to [3, 4, 11, 12, 14].

The prime objective of this paper is to study the decay estimates of solution to the initial-value problem (1.1)-(1.2). For our problem, due to the existence of memory term, it is a difficult task to obtain precisely the solution operator or its Fourier transform. While, by using Fourier transform and Laplace transform, we obtain the solution u to the linear problem (1.1)-(1.2) given by (2.4) and the solution operators $G(t)*$ and $H(t)*$. Furthermore, by employing the energy method in the Fourier space, we gain the point-wise estimate in the Fourier space of solutions to the problem (1.1)-(1.2). Appealing to this point-wise estimate, we obtain the point-wise of solution operators and their properties. Therefore, the decay estimates of solutions to (1.1)-(1.2) are achieved.

The contents of the paper are as follows. Solution formula are obtained in section 2. In section 3, we obtain the estimates and properties of solutions operators, which is based on the point-wise estimate in the Fourier space of solutions to the linear problem. In the last section, we prove the decay estimates of solutions to the linear problem by virtue of the properties of solution operators.

Before the end of this section, we give some notations to be used below. Let the Fourier transform of f indicated as $\mathrm{F}[f]$:

$$\mathrm{F}[f](\xi)=\hat{f}(\xi):=\frac{1}{(2\pi)^{\frac{n}{2}}}\int_{R^n}e^{-ix\cdot\xi}f(x)dx,$$

and we denote its inverse transform as F^{-1}.

Let the Laplace transform of f indicated as $\mathrm{L}[f]$:

$$\mathrm{L}[f](\lambda):=\int_0^\infty e^{-\lambda t}f(t)dt,$$

and we denote its inverse transform as L^{-1}.

$L^p=L^p(R^n)(1\le p\le\infty)$ is the usual Lebesgue space with the norm $\|\cdot\|_{L^p}$.

Z_+ denotes the totality of all the non-negative integers.

$W^{m,p}(R^n)$, $m\in Z_+$, $p\in[1,\infty]$ denote the usual Sobolev space with its norm

$$\|f\|_{W^{m,p}}:=(\sum_{k=0}^{m}\|\partial_x^k f\|_{L^p}^p)^{\frac{1}{p}}.$$

In particular, we use $W^{m,2}=H^m$. Here, for a nonnegative integer k , ∂_x^k denotes the totality or each of all the k -th order derivatives with resect to $x\in R^n$. Also, $C^k(I;H^m(R^n))$ denotes the space of k -times continuously differentiable functions on the interval I with values in the Sobolve space $H^m=H^m(R^n)$.

Finally, in this paper, we denote every positive constant by the same symbol C or c without confusion. $[\cdot]$ is Gauss' symbol.

2. Solution formula

In this section, our purpose is to obtain the solution formula of the problems (1.1)-(1.2). Suppose $G(x,t)$ and $H(x,t)$ are solutions to the following problems,

$$\begin{cases} G_{tt}+(1+\Delta^2)G-g*G=0,\\ G(x,0)=\delta(x),\\ G_t(x,0)=0.\end{cases}\qquad(2.1)$$

$$\begin{cases} H_{tt}+(1+\Delta^2)H-g*H=0,\\ H(x,0)=0,\\ H_t(x,0)=\delta(x).\end{cases}\qquad(2.2)$$

Apply Fourier transform and Laplace transform to (2.1) and （2.2）, then we have formally that

$$\hat{G}(\xi,t)=C\mathrm{L}^{-1}[\frac{\lambda}{1+|\xi|^4+\lambda^2-\mathrm{L}[g](\lambda)}](\xi,t),$$

$$\hat{H}(\xi,t)=C\mathrm{L}^{-1}[\frac{1}{1+|\xi|^4+\lambda^2-\mathrm{L}[g](\lambda)}](\xi,t).$$

here C is the constant determined by the initial data in (2.1) and (2.2).

Now we just compute $\hat{G}(\xi,t)$, similarly we could get $\hat{H}(\xi,t)$. First, apply Fourier transform to (2.1), we can obtain the following equation:

$$\begin{cases} \hat{G}_{tt}+(1+|\xi|^4)\hat{G}-g*\hat{G}=0,\\ \hat{G}(\xi,0)=\hat{\delta}(\xi),\\ \hat{G}_t(\xi,0)=0.\end{cases}$$

then apply Laplace transform to above equation, we can get

$$\int_0^\infty \hat{G}_{tt}e^{-\lambda t}dt+(1+|\xi|^4)\int_0^\infty\hat{G}e^{-\lambda t}dt -\int_0^\infty(g*\hat{G})e^{-\lambda t}dt=0,$$

by computing, we have that

$$-\lambda C+(1+|\xi|^4+\lambda^2)\mathrm{L}[\hat{G}](\lambda) -\mathrm{L}[g](\lambda)\cdot\mathrm{L}[\hat{G}](\lambda)=0,$$

so

$$\mathrm{L}[\hat{G}](\lambda)=\frac{C\lambda}{1+|\ \xi|^4+\lambda^2-\mathrm{L}[g](\lambda)},$$

finally, we have formally that

$$\hat{G}(\xi,t)=C\mathrm{L}^{-1}[\frac{\lambda}{1+|\ \xi|^4+\lambda^2-\mathrm{L}[g](\lambda)}](\xi,t).$$

Similarly,

$$\hat{H}(\xi,t)=C\mathrm{L}^{-1}[\frac{1}{1+|\ \xi|^4+\lambda^2-\mathrm{L}[g](\lambda)}](\xi,t),$$

here C is a constant determined by the initial data in (2.1).

Lemma2.1

$\hat{G}(\xi,t)$ and $\hat{H}(\xi,t)$ exist.

Proof.

We only prove $\hat{G}(\xi,t)$ exists; similarly we could prove $\hat{H}(\xi,t)$ exists.

Denote $F(\lambda):=\lambda^2+1+|\xi|^4-\mathrm{L}[g](\lambda)$.

To prove $\mathrm{L}^{-1}[\frac{\lambda}{F(\lambda)}]$ exists, we need to consider the zero points of $F(\lambda)$. Denote $\lambda=\sigma+i\nu$, $\sigma>-C_1$, C_1 is same as that in Assumption [A] b), then $\mathrm{L}[g](\lambda)$ exists. Assume that $\lambda_1=\sigma_1+i\nu_1$ is a zero point of $F(\lambda)$ and $\sigma_1>-C_1$, then σ_1 and ν_1 satisfy

$$\begin{cases}\mathrm{Re}F(\lambda_1) & =\sigma_1^2-\nu_1^2+1+|\xi|^4- \\ & \quad\int_0^\infty \cos(\nu_1 t)e^{-\sigma_1 t}g(t)dt=0, \\ \mathrm{Im}F(\lambda_1) & =\int_0^\infty \sin(\nu_1 t)e^{-\sigma_1 t}g(t)dt \\ & \quad +2\sigma_1\nu_1=0.\end{cases} \qquad (2.3)$$

We claim that $\sigma_1\le 0$. Now we prove the claim by contradiction.

Assume that $\sigma_1>0$. If $\nu_1=0$, then in view of $\int_0^\infty g(t)dt<1$, we obtain that

$$\mathrm{Re}F(\lambda_1)=\sigma_1^2+1+|\xi|^4-\int_0^\infty e^{-\sigma_1 t}g(t)dt>0,$$

it yields contradiction with $(2.3)_1$.

If $\nu_1\neq 0$, then we have that

$$\mathrm{Im}F(\lambda_1)=\nu_1\Big(2\sigma_1+\int_0^\infty\frac{\sin(\nu_1 t)}{\nu_1}e^{-\sigma_1 t}g(t)dt\Big).$$

Next we prove that $\int_0^\infty\frac{\sin(\nu_1 t)}{\nu_1}e^{-\sigma_1 t}g(t)dt>0.$

Denote $a_m=\int_0^{\frac{2m\pi}{|\nu_1|}}\frac{\sin|\nu_1 t|}{|\nu_1|}e^{-\sigma_1 t}g(t)dt$, and we will prove $\{a_m\}_{m=1}^\infty$ is a convergent sequence. By direct computation, we have that

$a_1=$

$$\int_0^{\frac{\pi}{|\nu_1|}}\frac{\sin|\nu_1 t|}{|\nu_1|}\Big(e^{-\sigma_1 t}g(t)-e^{-\sigma_1(t+\frac{\pi}{|\nu_1|})}g(t+\frac{\pi}{|\nu_1|})\Big)dt.$$

Since $\partial_t\big(e^{-\sigma_1 t}g(t)\big)<0$, we have that

$$0<a_1<\int_0^{\frac{\pi}{|\nu_1|}}te^{-\sigma_1 t}g(t)dt.$$

Similarly,

$$a_{m+1}-a_m=\int_{\frac{2m\pi}{|\nu_1|}}^{\frac{2m\pi+\pi}{|\nu_1|}}\frac{\sin|\nu_1 t|}{|\nu_1|}$$

$$\Big(e^{-\sigma_1 t}g(t)-e^{-\sigma_1(t+\frac{\pi}{|\nu_1|})}g(t+\frac{\pi}{|\nu_1|})\Big)dt,$$

so we have that

$$0<a_{m+1}-a_m<\int_{\frac{2m\pi}{|\nu_1|}}^{\frac{2m\pi+\pi}{|\nu_1|}}te^{-\sigma_1 t}g(t)dt.$$

It yields that

$$0<a_m<\int_0^{\frac{2m\pi}{|\nu_1|}}te^{-\sigma_1 t}g(t)dt\le\frac{g(0)}{(\sigma_1+C_1)^2},$$

so $\{a_m\}_{m=1}^\infty$ is a bounded and monotonic increasing sequence. Since $a_1>0$, $a(\lambda_1):=\lim_{m\to\infty}a_m>0$. Thus we proved that $\int_0^\infty\frac{\sin(\nu_1 t)}{\nu_1}e^{-\sigma_1 t}g(t)dt>0.$ Also, because

$\sigma_1>0$ and $\nu_1\neq 0$,

it results that $\mathrm{Im}F(\lambda_1)\neq 0$. This contradicts with $(2.3)_2$. Thus by contradiction we proved the claim $\sigma_1\le 0$.

Combining the two cases, we know that $\frac{\lambda}{F(\lambda)}$ is analytic in $\{\lambda\in£;\mathrm{Re}(\lambda)>0\}$. Take $\lambda=\sigma+i\nu$, $\sigma>\max\{\mathrm{Re}\lambda_s\}$, here $\{\lambda_s\}$ is the set of all the singular points of $F(\lambda)$, then by standard calculation we can prove that $\mathrm{L}^{-1}\left[\frac{\lambda}{F(\lambda)}\right](t)$ converges.

The constant C in the expression of $\hat{G}(\xi,t)$ and σ are determined by the initial data of $G(x,t)$. So far we complete the proof. W

In consideration of Lemma 2.1 and Duhamel principle, the solution to the problem (1.1)-(1.2) can be expressed as following:

$$u(t) = G(t)*u_0 + H(t)*u_1. \quad (2.4)$$

3. Decay properties of solution operators

In this section, we think of a way to obtain the next decay estimates of the solution operators $G(t)*$ and $H(t)*$ arising in the solution formula (2.4).

Proposition 3.1

Let k and l be integers, $\varphi \in H^{s+1}(R^n)$, $\psi \in H^{s-1}(R^n)$, then the next estimates hold:

1) $$\|\partial_x^k G(t)*\varphi\|_{L^2} \le Ce^{-Ct}\|\varphi\|_{L^2} + C(1+t)^{-\frac{l}{4}}\|\partial_x^{l+k}\varphi\|_{L^2},$$

for $k \ge 0$, $l \ge 0$, $l+k \le s+1$.

2) $$\|\partial_x^k G_t(t)*\varphi\|_{L^2} \le Ce^{-Ct}\|\varphi\|_{L^2} + C(1+t)^{-\frac{l}{4}}\|\partial_x^{l+k+2}\varphi\|_{L^2},$$

for $k \ge 0$, $l \ge 0$, $l+k \le s-1$.

3) $$\|\partial_x^k H(t)*\psi\|_{L^2} \le Ce^{-Ct}\|\psi\|_{L^2} + C(1+t)^{-\frac{l}{4}-\frac{1}{2}}\|\partial_x^{l+k}\psi\|_{L^2},$$

for $k \ge 0$, $l+2 \ge 0$, $0 \le l+k \le s-1$.

4) $$\|\partial_x^k H_t(t)*\psi\|_{L^2} \le Ce^{-Ct}\|\psi\|_{L^2} + C(1+t)^{-\frac{l}{4}}\|\partial_x^{l+k}\psi\|_{L^2},$$

for $k \ge 0$, $l \ge 0$, $l+k \le s-1$.

To testify Proposition 3.1, the most important step is to gain the point-wise estimates of the fundamental solutions in the Fourier space. In fact we can obtain this by using the following point-wise estimate of solutions to the linear problem (1.1)-(1.2).

Lemma 3.2

Assume u is the solution of (1.1)-(1.2), then it satisfies the following point-wise estimate in the Fourier space:

$$|\hat{u}_t(\xi,t)|^2 + (1+|\xi|^4)|\hat{u}(\xi,t)|^2 + (g\text{W}\hat{u})(\xi,t) \le Ce^{-C\rho(\xi)t}\left(|\hat{u}_1(\xi)|^2 + (1+|\xi|^4)|\hat{u}_0(\xi)|^2\right),$$

here $\rho(\xi) = \dfrac{1}{1+|\xi|^4}$.

To prove Lemma 3.2, we denote some notations. For any real or complex-valued function $f(t)$, we define

$$(g*f)(t) := \int_0^t g(t-\tau)f(\tau)d\tau,$$

$$(g\Diamond f)(t) := \int_0^t g(t-\tau)(f(\tau)-f(t))d\tau,$$

$$(g\text{W}f)(t) := \int_0^t g(t-\tau)|f(t)-f(\tau)|^2 d\tau.$$

We have the following lemma by direct calculation, which is useful in obtaining our point-wise estimate of solution in the Fourier space.

Lemma 3.2

For any function $k \in C(R)$, and any $\phi \in W^{1,2}(0,T)$, it holds that

1) $(k*\phi)(t) = (k\Diamond\phi)(t) + \int_0^t k(\tau)d\tau\phi(t)$,

2) $$\begin{aligned}\mathrm{Re}\{(k*\phi)(t)\bar{\phi}_t(t)\} &= -\frac{1}{2}k(t)|\phi(t)|^2 + \frac{1}{2}(k'\text{W}\phi)(t) \\ &\quad -\frac{1}{2}\frac{d}{dt}\{(k\text{W}\phi)(t) - (\int_0^t k(\tau)d\tau)|\phi(t)|^2\},\end{aligned}$$

3) $|(k\Diamond\phi)|^2 \le (\int_0^t |k(\tau)|d\tau)(|k|\text{W}\phi)(t)$.

Next we will obtain the point-wise estimates in the Fourier space of solutions to the problem (1.1)-(1.2).

Proof of Lemma 3.2.

Step1: By using Fourier transform to (1.1) we get the following equality:

$$\hat{u}_{tt} + (1+|\xi|^4)\hat{u} - g*\hat{u} = 0. \quad (3.2)$$

Multiplying (3.2) by $\bar{\hat{u}}_t$ we obtain the next equality by taking the real part,

$$Re\left\{\bar{\hat{u}}_t\left(\hat{u}_{tt} + (1+|\xi|^4)\hat{u} - g*\hat{u}\right)\right\} = 0.$$

It yields that

$$\{\frac{1}{2}|\hat{u}_t|^2 + \frac{1}{2}(1+|\xi|^4)|\hat{u}|^2\}_t - Re\{g*\hat{u}\bar{\hat{u}}_t\} = 0. \quad (3.3)$$

To the term $Re\{g*\hat{u}\bar{\hat{u}}_t\}$ in (3.3) apply 2) in Lemma 3.3 we have that

$$Re\{g*\hat{u}\bar{\hat{u}}_t\} = -\frac{1}{2}g(t)|\hat{u}|^2 + \frac{1}{2}(g'\text{W}\hat{u})(t) - \frac{1}{2}\frac{d}{dt}\left\{(g\text{W}\hat{u})(t) - \int_0^t g(\tau)d\tau|\hat{u}|^2\right\}.$$

We denote

$$E_1(\xi,t) = |\hat{u}_t|^2 + (1+|\xi|^4)|\hat{u}|^2 + g\text{W}\hat{u} - (\int_0^t g(s)ds)|\hat{u}|^2,$$

$$F_1(\xi,t) = g|\hat{u}|^2 - g'\text{W}\hat{u},$$

then we have that

$$\frac{\partial}{\partial t}E_1(\xi,t) + F_1(\xi,t) = 0. \quad (3.4)$$

Step 2: Multiplying (3.2) by $\{-(g*\bar{\hat{u}})_t\}$ and we obtain the next equality by taking the real part,

$$Re\left\{-(g*\bar{\hat{u}})_t\right\}\left\{\hat{u}_{tt}+(1+|\xi|^4)\hat{u}-g*\hat{u}\right\}=0.$$

It results that

$$\{\frac{1}{2}|g*\hat{u}|^2\}_t-Re\{(g*\bar{\hat{u}})_t\hat{u}_{tt}\}$$
$$-Re\{(1+|\xi|^4)\hat{u}(g*\bar{\hat{u}})_t\}=0. \quad (3.5)$$

Due to $(g*\bar{\hat{u}})_t=g(0)\bar{\hat{u}}+g'*\bar{\hat{u}}$, the second term in (3.5) yields that

$$\begin{aligned}&-Re\{\hat{u}_{tt}(g*\bar{\hat{u}})_t\}\\&=-Re\{\hat{u}_t(g*\bar{\hat{u}})_t\}_t+Re\{\hat{u}_t(g*\bar{\hat{u}})_{tt}\}\\&=-Re\{\hat{u}_t(g*\bar{\hat{u}})_t\}_t+Re\left\{\hat{u}_t\left(g(0)\bar{\hat{u}}_t+(g'*\bar{\hat{u}})_t\right)\right\}\\&=-Re\{\hat{u}_t(g*\bar{\hat{u}})_t\}_t+Re\{g(0)|\hat{u}_t|^2+\hat{u}_t(g'*\bar{\hat{u}})_t\}.\end{aligned}$$

We denote

$$E_2(\xi,t)=\frac{1}{2}|g*\hat{u}|^2-Re\{\hat{u}_t(g*\bar{\hat{u}})_t\},$$
$$F_2(\xi,t)=g(0)|\hat{u}_t|^2,$$
$$R_2(\xi,t)=Re\{-\hat{u}_t(g'*\bar{\hat{u}})_t+(1+|\xi|^4)\hat{u}(g*\bar{\hat{u}})_t\},$$

then obtain that

$$\frac{\partial}{\partial t}E_2(\xi,t)+F_2(\xi,t)=R_2(\xi,t). \quad (3.6)$$

Step 3: Multiplying (3.2) by $\bar{\hat{u}}$ and we obtain the next equality by taking the real part,

$$Re\{\bar{\hat{u}}(\hat{u}_{tt}+(1+|\xi|^4)\hat{u}-g*\hat{u})\}=0.$$

It yields that

$$Re\left\{\hat{u}_t\bar{\hat{u}}\right\}_t-|\hat{u}_t|^2+(1+|\xi|^4)|\hat{u}|^2-$$
$$Re\left\{g*\hat{u}\bar{\hat{u}}\right\}=0. \quad (3.7)$$

Due to 1) in Lemma 3.3 , we obtain that

$$Re\{g*\hat{u}\bar{\hat{u}}\}=(\int_0^t g(s)ds)|\hat{u}|^2+Re\{g\Diamond\hat{u}\bar{\hat{u}}\}.$$

We denote

$$E_3(\xi,t)=Re\{\hat{u}_t\bar{\hat{u}}\},$$
$$F_3(\xi,t)=(1+|\xi|^4)|\hat{u}|^2-(\int_0^t g(s)ds)|\hat{u}|^2,$$
$$R_3(\xi,t)=|\hat{u}_t|^2+Re\{g\Diamond\hat{u}\bar{\hat{u}}\},$$

then (3.7) yields that

$$\frac{\partial}{\partial t}E_3(\xi,t)+F_3(\xi,t)=R_3(\xi,t). \quad (3.8)$$

Define $\rho(\xi)=\dfrac{1}{1+|\xi|^4}$, and denote

$$E(\xi,t)=E_1(\xi,t)+\rho(\xi)(\alpha E_2(\xi,t)+\beta E_3(\xi,t)),$$
$$F(\xi,t)=F_1(\xi,t)+\rho(\xi)(\alpha F_2(\xi,t)+\beta F_3(\xi,t)),$$
$$R(\xi,t)=\rho(\xi)(\alpha R_2(\xi,t)+\beta R_3(\xi,t)),$$

where α,β are positive constants, then (3.4), (3.6) and (3.8) yield that

$$\frac{\partial}{\partial t}E(\xi,t)+F(\xi,t)=R(\xi,t). \quad (3.9)$$

We introduce Lyapunov functionals:

$$E_0(\xi,t)=|\hat{u}_t|^2+(1+|\xi|^4)|\hat{u}|^2+gW\hat{u},$$
$$F_0(\xi,t)=gW\hat{u}+g|\hat{u}|^2.$$

We know that there exist some positive constants c_i (i=1,2,3) from the definitions of $E_1(\xi,t)$ and $F_1(\xi,t)$, such that the following inequalities hold:

$$\begin{cases}c_1E_0(\xi,t)\le E_1(\xi,t)\le c_2E_0(\xi,t),\\ F_1(\xi,t)\ge c_3F_0(\xi,t).\end{cases} \quad (3.10)$$

On the other hand,

$$|E_2(\xi,t)|\le C(|\hat{u}_t|^2+|\hat{u}|^2+gW\hat{u}),$$
$$|E_3(\xi,t)|\le C(|\hat{u}_t|^2+|\hat{u}|^2),$$
$$|\rho(\xi)(\alpha E_2(\xi,t)+\beta E_3(\xi,t))|$$
$$\le c_4(\alpha+\beta)E_0(\xi,t).$$

Choose α,β properly small such that

$c_4(\alpha+\beta)\le min(\frac{c_1}{2},\frac{c_2}{2})$,

from (3.10) we have that

$$\frac{c_1}{2}E_0(\xi,t)\le E(\xi,t)\le\frac{3c_2}{2}E_0(\xi,t). \quad (3.11)$$

By virtue of (3.10) and noticing that $0\le\int_0^t g(s)ds\le 1$, it is not hard to prove that

$$F(\xi,t)\ge c_3F_0(\xi,t)+$$
$$\rho(\xi)\left\{\alpha g(0)|\hat{u}_t|^2+\frac{\beta}{2}(1+|\xi|^4)|\hat{u}|^2\right\}. \quad (3.12)$$

By virtue of Lemma 3.3, we have that

$$|R_2(\xi,t)|\le\varepsilon|\hat{u}_t|^2+\delta(1+|\xi|^4)|\hat{u}|^2+$$
$$C_{\varepsilon,\delta}(1+|\xi|^4)F_0(\xi,t),$$

and

$$|R_3(\xi,t)|\le|\hat{u}_t|^2+\gamma|\hat{u}|^2+C_\gamma gW\hat{u},$$

where $\varepsilon,\delta,\gamma$ are positive constants. Then it is easy to get that

$$\begin{aligned}|R(\xi,t)|\le&\ \rho(\xi)\{(\alpha\varepsilon+\beta)|\hat{u}_t|^2\\&+(\alpha\delta+\beta\gamma)(1+|\xi|^4)|\hat{u}|^2\\&+\alpha C_{\varepsilon,\delta}(1+|\xi|^4)F_0(\xi,t)+\beta C_\gamma gW\hat{u}\}\\\le&\ (\alpha\varepsilon+\beta)\rho(\xi)|\hat{u}_t|^2+(\alpha\delta+\beta\gamma)|\hat{u}|^2\\&+(\alpha+\beta)C_{\varepsilon,\delta,\gamma}F_0(\xi,t).\end{aligned}$$

.We claim that there exist $\gamma,\varepsilon,\delta,\alpha,\beta$ such that

$$|R(\xi,t)|\le\frac{1}{2}F(\xi,t). \quad (3.13)$$

First choose $\gamma=\frac{1}{8}$, $\varepsilon=\frac{1}{4}g(0)$, $\delta=\frac{1}{32}g(0)$, $\beta=\frac{1}{4}\alpha g(0)$, then the next three inequalities hold:

$$\frac{1}{4}\beta\geq\alpha\delta+\beta\gamma, \qquad \frac{1}{2}\alpha g(0)\geq\alpha\varepsilon+\beta,$$

$$\frac{1}{2}c_3\geq(\alpha+\beta)C_{\varepsilon,\delta,\gamma}.$$

So as to prove (3.13) (here (3.10) is also considered), it suffices to choose α suitably small such that

$$(1+\frac{1}{4}g(0))\alpha\leq min\{\frac{c_3}{2C_{\varepsilon,\delta,\gamma}},\frac{c_1}{2c_4},\frac{c_2}{2c_4}\}.$$

Due to (3.13) and (3.9), we get that

$$\frac{\partial}{\partial t}E(\xi,t)+\frac{1}{2}F(\xi,t)\leq 0. \quad (3.14)$$

On the other hand, due to (3.11) and (3.12) we obtain that

$$F(\xi,t)>c\rho(\xi)E(\xi,t). \quad (3.15)$$

From (3.14) and (3.15), we have that

$$E(\xi,t)\leq e^{-C\rho(\xi)t}E(\xi,0). \quad (3.16)$$

By virtue of (3.11) and (3.16), we have that

$$|\hat{u}_t|^2+(1+|\xi|^4)|\hat{u}|^2+g\square\hat{u}$$
$$\leq Ce^{-C\rho(\xi)t}\left(|\hat{u}_1(\xi)|^2+(1+|\xi|^4)|\hat{u}_0(\xi)|^2\right),$$

so, we obtain the point-wise estimates of solutions to (1.1)-(1.2) in the Fourier space. $\square$

As a simple corollary of Lemma 3.2, we have the following point-wise estimates of the fundamental solutions $G(x,t)$ and $H(x,t)$ in the Fourier space.

Lemma 3.4
$G(x,t)$ and $H(x,t)$ satisfy

1). $|\hat{G}(\xi,t)|\leq Ce^{-C\rho(\xi)t}$;

2). $|\hat{G}_t(\xi,t)|\leq Ce^{-C\rho(\xi)t}(1+|\xi|^4)^{\frac{1}{2}}$;

3). $|\hat{H}(\xi,t)|\leq Ce^{-C\rho(\xi)t}(1+|\xi|^4)^{-\frac{1}{2}}$;

4). $|\hat{H}_t(\xi,t)|\leq Ce^{-C\rho(\xi)t}$,

where $\rho(\xi)=\frac{1}{1+|\xi|^4}$.

Proof.
Putting (2.4) with $u_1=0$ in (3.1), it results that

$$|\hat{G}_t(\xi,t)|^2+(1+|\xi|^4)|\hat{G}(\xi,t)|^2 \leq Ce^{-C\rho(\xi)t}(1+|\xi|^4),$$

it yields 1) and 2).

Putting (2.4) with $u_0=0$ in (3.1), it results that

$$|\hat{H}_t(\xi,t)|^2+(1+|\xi|^4)|\hat{H}(\xi,t)|^2\leq Ce^{-C\rho(\xi)t},$$

it yields 3) and 4). $\square$

Next we use Lemma 3.4 to prove Proposition 3.1

Proof of Proposition 3.1: With a view of 1) in Lemma 3.4 , we have that

$$\|\partial_x^k G(t)*\varphi\|_{L^2}^2 \leq C\int_{R^n}|\xi|^{2k}e^{-C\rho(\xi)t}|\hat{\varphi}(\xi)|^2d\xi$$
$$\leq C\int_{\{\xi:|\xi|\leq 1\}}|\xi|^{2k}e^{-\frac{C}{2}t}|\hat{\varphi}|^2d\xi$$
$$+C\int_{\{\xi:|\xi|\geq 1\}}|\xi|^{2k}e^{-\frac{Ct}{2|\xi|^4}}|\hat{\varphi}(\xi)|^2d\xi$$
$$\leq Ce^{-Ct}\|\varphi\|_{L^2}^2+C(1+t)^{-\frac{l}{2}}\|\partial_x^{k+l}\varphi\|_{L^2}^2,$$

here $k\geq 0$, $l\geq 0$, $l+k\leq s+1$. Thus 1) is proved.
By virtue of 2), 3) and 4) in Lemma (3.4) , 2), 3) and 4) in Proposition 3.1 can be similarly proved. $\square$

All material on each page should fit within a rectangle of 18 x 23.5 cm (7" x 9.25"), centered on the page, beginning 2.54 cm (1") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75"). The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

4. Decay estimates for linear problem

In this section, we study the decay estimates of solutions of the linear problem (1.1)-(1.2).

Theorem 4.1.
Let $s\geq 1$ be an integer. Suppose that $u_0\in H^{s+1}(R^n)$ and $u_1\in H^{s-1}(R^n)$, and set $I_0:=\|u_0\|_{H^{s+1}}+\|u_1\|_{H^{s-1}}$. Then the solution u of the problem (1.1)-(1.2) given by (2.4) satisfies that

$$u\in C^0([0,\infty);H^{s+1}(R^n))\cap C^1([0,\infty);H^{s-1}(R^n)),$$

and the following energy estimate:

$$\|u_t(t)\|_{H^{s-1}}^2+\|u(t)\|_{H^{s+1}}^2$$
$$+\int_0^t(\|u_t(\tau)\|_{H^{s-3}}^2+\|u(\tau)\|_{H^{s-1}}^2)d\tau\leq CI_0^2.$$

Proof. We have obtained the solution u of (1.1)-(1.2) given by (2.4) and proved that it satisfies the point-wise estimates (3.1) in the Fourier space. Due to (3.14) and (3.15) we obtain that $\frac{\partial}{\partial t}E(\xi,t)+C\rho(\xi)E(\xi,t)\leq 0$. Integrate the above inequality with respect to t and use the inequality (3.11), so we have that

$$E_0(\xi,t)+\int_0^t\rho(\xi,t)E_0(\xi,\tau)d\tau\leq CE_0(\xi,0). \quad (4.1)$$

Multiply (4.1) by $(1+|\xi|^2)^{s-1}$ and integrate the resulting inequality with respect to $\xi\in R^n$, then we obtain that

$$\int_{R^n}(1+|\xi|^2)^{s-1}E_0(\xi,t)d\xi$$
$$+\int_{R^n}(1+|\xi|^2)^{s-1}\int_0^t \rho(\xi,t)E_0(\xi,\tau)d\tau d\xi$$
$$\le C\int_{R^n}(1+|\xi|^2)^{s-1}E_0(\xi,0)d\xi,$$

it yields that

$$\|u_t(t)\|^2_{H^{s-1}}+\|u(t)\|^2_{H^{s+1}}$$
$$+\int_0^t(\|u_t(\tau)\|^2_{H^{s-3}}+\|u(\tau)\|^2_{H^{s-1}})d\tau \le CI_0^2. \quad (4.2)$$

(4.2) guarantees the regularity of the solution (2.4). So far we complete the proof of Theorem 4.1. W

By using Proposition 3.1 we obtain the following decay estimates of u given by (2.4), if initial data $u_0 \in H^{s+1}(R^n)$ and $u_1 \in H^{s-1}(R^n)$.

Theorem 4.2.

With the same conditions as Theorem 4.1, if $d \in Z_+$, then u given by (2.4) satisfies the following decay estimate:

$$\|\partial_x^k u(t)\|_{H^{s+1-k-d}} \le CI_0(1+t)^{-\frac{d}{4}}, \quad (4.3)$$

here $k \ge 0$, $d \ge 0$, $k+d \le s+1$;

if $d \notin Z_+$, the following decay estimate holds:

$$\|\partial_x^k u(t)\|_{H^{s-k-[d]}} \le CI_0(1+t)^{-\frac{d}{4}}, \quad (4.4)$$

for $k \ge 0$, $d \ge 0$, $k+[d] \le s$.

Proof.

Assume $k \ge 0$, $m \ge 0$ are integers. By using (2.4) and applying 1) and 3) in Proposition 3.1, we have that

$$\|\partial_x^{k+m}u(t)\|_{L^2}$$
$$\le \|\partial_x^{k+m}G(t)*u_0\|_{L^2}+\|\partial_x^{k+m}H(t)*u_1\|_{L^2}$$
$$\le Ce^{-C(t+1)}\|u_0\|_{L^2}+C(1+t)^{-\frac{l_1}{4}}\|\partial_x^{k+m+l_1}u_0\|_{L^2}$$
$$+Ce^{-C(t+1)}\|u_1\|_{L^2}+C(t+1)^{-\frac{l_2}{4}-\frac{1}{2}}\|\partial_x^{k+m+l_2}u_1\|_{L^2}$$
$$\le Ce^{-C(1+t)}\|(u_0,u_1)\|_{L^2}$$
$$+C(1+t)^{-\frac{l_1}{4}}\|\partial_x^{k+m+l_1}u_0\|_{L^2}$$
$$+C(1+t)^{-\frac{l_2}{4}-\frac{1}{2}}\|\partial_x^{k+m+l_2}u_1\|_{L^2}, \quad (4.5)$$

here $l_1 \ge 0$, $l_2 \ge -2,$ $k+m+l_1 \le s+1$, $0 \le k+m+l_2 \le s-1$.

Choose the minimal integers l_1 and l_2 satisfying

$$\frac{l_1}{4} \ge \frac{d}{4}, \frac{l_2}{4}+\frac{1}{2} \ge \frac{d}{4},$$

i.e.

$$l_1 = \begin{cases} d, & d \in Z_+; \\ [d]+1, & d \notin Z_+, \end{cases} \quad l_2 = l_1 - 2.$$

At the same time, the next inequality holds:

$$e^{-C(1+t)} \le C(1+t)^{-\frac{d}{4}}.$$

So if $d \in Z_+$, we obtain from (4.5) that

$$\|\partial_x^{k+m}u(t)\|_{L^2} \le CI_0(1+t)^{-\frac{d}{4}},$$

for $0 \le m \le s+1-k-d$. Take sum with $0 \le m \le s+1-k-d$, we get (4.3).

If $d \notin Z_+$, we obtain from (4.5) that

$$\|\partial_x^{k+m}u(t)\|_{L^2} \le CI_0(1+t)^{-\frac{d}{4}},$$

for $0 \le m \le s-k-[d]$. Take sum with $0 \le m \le s-k-[d]$, then we get (4.4). W

Remark 1. With the same conditions as Theorem 4.1, through the similar proof to Theorem 4.2 we have the following estimates:

if $d \in Z_+$, $\|\partial_x^k u_t(t)\|_{H^{s-1-k-d}} \le CI_0(1+t)^{-\frac{d}{4}}$,

for $k \ge 0$, $d \ge 0$, $k+d \le s-1$;

if $d \notin Z_+$, $\|\partial_x^k u_t(t)\|_{H^{s-2-k-[d]}} \le CI_0(1+t)^{-\frac{d}{4}}$,

for $k \ge 0$, $d \ge 0$, $k+[d] \le s-2$.

Remark 2. In the special case $d=k$ in Theorem 4.2, we obtain the following estimate:

$$\|\partial_x^k u(t)\|_{H^{s+1-2k}} \le CI_0(1+t)^{-\frac{k}{4}},$$

here $k \ge 0$, $2k \le s+1$.

Remark 3. The estimates in Theorem 4.1 and Theorem 4.2 indicate that the decay structure of solutions to the linear problem (4.1)-(4.2) is of regularity-loss type. To have $\frac{d}{4}$-order decay, we have to lose d -order regularity.

Remark 4. The condition c) in Assumption [A] plays a key role to obtain the dissipative structure in this paper. If c) is weakened to $1-\int_0^t g(s)ds \ge 0$, then the dissipative structure would be totally different. Take memory kernels $g_1(t)=a\mu e^{-\mu t}$ and $g_2(t)=\mu e^{-\mu t}$ ($a<1$ and μ are constants) as examples, by direct calculation we can see the difference between the two dissipative structures, which in some way reflects the optimality of the dissipative structure in this paper.

5. ACKNOWLEDGMENTS
This work is partially supported by the Fundamental Research Funds for the Central Universities (Grant No. 2018MS052).

6. REFERENCES
[1] R.C. Char ão, E. Bisognin, V. Bisognin and A.F. Pazoto, Asymptotic behavior for a dissipative plate equation in ¡ ^{N} with periodic coefficients, Electronic J. Differential Equations, 46 (2008), 23 pp.

[2] C.R. da Luz and R.C. Char ão, Asymptotic properties for a semi-linear plate equation in unbounded domains, J. Hyperbolic Differential Equations, 6 (2009), 269-294.

[3] P. M. N. Dharmawardane, J. E. Mu$\tilde{\rm n}$oz Rivera and S. Kawashima, Decay property for second order hyperbolic systems of viscoelastic materials, J. Math. Anal. Appl., 366 (2010), 621-635.

[4] M. Fabrizio and B. Lazzari, On the existence and the asymptotic stability of solutions for linear viscoelastic solids, Arch. Rational Mech. Anal., 116 (1991), 139-152.

[5] R.Ikehata, A remark on a critical exponent for the semilinear dissipative wave equation in the one dimensional half space, Differ. Integral Equ. 16 (2003), 727-736.

[6] R. Ikehata, K. Nishihara and H.-J. Zhao, Global asymptotics of solutions to the Cauchy problem for the damped wave equation with Absorption, J. Differential Equations, 226 (2006), 1--29.

[7] Y. Liu, The Pointwise Estimates of Solutions for Semilinear Dissipative Wave Equation, Commun. Pur. Appl. Anal., 12 (2013), 237-252.

[8] Y. Liu and S. Kawashima, Decay property for a plate equation with memory-type dissipation, Kinet. Relat. Mod., 4 (2011), 531-547.

[9] Y. Liu and S. Kawashima, Global existence and asymptotic behavior of solutions for quasi-linear dissipative plate equation, Discrete Contin. Dyn. Syst., 29 (2011), 1113-1139.

[10] Y. Liu and S. Kawashima, Global existence and decay of solutions for a quasi-linear dissipative plate equation, J. Hyperbol. Differ. Eq., 8 (2011), 591-614.

[11] Y. Liu and S. Kawashima, Decay property for the Timoshenko system with memory-type dissipation, Math. Models Meth. Appl. Sci., 22 (2012), 1-19.

[12] Y. Liu and S. Kawashima, Global existence and asymptotic decay of solutions to the nonlinear Timoshenko system with memory, Nonlinear Analysis-TMA, 84 (2013), 1-17.

[13] Y. Liu and W. Wang, The pointwise estimates of solutions for dissipative wave equation in multi-dimensions, Discrete Contin. Dyn. Syst., 20 (2008), 1013-1028.

[14] J.E. Mu ñoz Rivera, M.G. Naso and F.M. Vegni, Asymptotic behavior of the energy for a class of weakly dissipative second-order systems with memory, J. Math. Anal. Appl., 286 (2003), 692-704.

[15] K. Nishihara, $L^p - L^q$ estimates of solutions to the damped wave equation in 3-dimensional space and their application, Math. Z., 244 (2003), 631-649.

[16] K. Nishihara and H.-J. Zhao, Decay properties of solutions to the Cauchy problem for the damped wave equation with absorption, J. Math. Anal. Appl., 313 (2006), 598--610.

[17] W. Wang and T. Yang, The pointwise estimates of solutions for Euler equations with damping in multi-dimensions, J. Differential Equations, 173 (2001), 410-450.

Smart Helmet on IOT Technology for Safety and Accident Detection

Tejashwini R
UG Student
Nagarjuna College of Engineering and Technology
Bangalore, India – 562110

Sharanya S
UG Student
Nagarjuna College of Engineering and Technology
Bangalore, India – 562110

Swathi S
Associate Professor
Nagarjuna College of Engineering and Technology
Bangalore, India – 562110

ABSTRACT: IoT made the present generation to connect with the different network of devices for exchanging of the data. Nowadays it had made compulsion on wearing the helmet while riding. In our paper we introduce a helmet which is made smart using the latest IOT technologies. The papers' main objective is to build a safety system that is integrated with the smart helmet and intelligent bike which reduce the chances of two-wheeler accident and drunk drive cases. The pressure sensor check whether the person is wearing helmet or not. Alcohol sensors (Gas sensor) which is installed at the bottom of the helmet detect the alcoholic content in riders' breath. If the rider is not wearing the helmet or if there is any alcohol content found in rider's breath, the bike remains off. The bike will start if and only if the rider wears the helmet provided with no alcoholic content present. When the rider crashes, helmet hits the ground, sensors detect the motion of helmet and reports the occurrence of an accident. It sends information of corresponding location and message to the registered number through GPS & GSM module respectively. And along with it we are implementing cooling system inside the helmet which would help the rider to stay cool during the climatic changes which in turn reduces the irritation that is created from the helmet.

Keywords: Pressure Sensor, Gas Sensor, GPS & GSM module.

1. INTRODUCTION

A Smart Helmet is a protective headgear used by the rider which makes riding safer than before. The idea of inventing the helmet is to do good deeds towards the society. Accidents of two wheelers are increasing drastically day by day which leads to the loss of many lives in the country. The intention of our project is to build a safety measure for the rider which is integrated with the smart helmet and intelligent bike to reduce the accidents two-wheeler.The idea of our project concentrate on four major areas. [1] Position of the helmet. The ignition of the bike will remain off until the rider wears the helmet which is monitored by the pressure sensor that is present inside the helmet. [2] The level of alcohol in the riders breath. If the rider is alcoholic, the bike will still remain off Gas sensor ie., MQ3 alcohol sensor helps us to detect the alcohol content. [3] Accident detection. If accident occurs suddenly, there would be no person to help the victim. In such situation its very much necessary to inform to ambulance or family members. So we create a module which sends immediate message to the registered number with the current location stating that he/she met with an accident at so an so place through GSM and GPS technology. [4] Managing the temperature inside the helmet which continuously monitors the body temperature and tries to maintain the constant temperature which allows the rider to stay cool all time rather than getting irritated due to increase in humidity.

The paper comprises with normal test case for checking the various conditions of occurrence which would be solved using new emerging IOT technologies.

2. LITERATURE SURVEY

[1]. Mohd Khairul Amri Kamarudin has established, "Smart Helmet with Sensors for Accident Prevention. This paper provides an intelligent system for two wheeler accident prevention and detection for human life safety. The prevention part involves, Smart Helmet, which automatically checks whether the person is wearing the helmet and has non-alcoholic breath while driving.

[2]. Vijay J, Saritha B, Priyadharshini B,Deepeka S and Laxmi R (2011) has established, "Drunken Drive Protection System". International Journal of Scientific & Engineering Research. This system efficiently checks the wearing of helmet and drunken driving. By implementing this system a

safe two wheeler journey is possible which would decrease the head injuries during accidents and also reduce the accidents due to drunken driving. An intelligent system has been embedded in the helmet itself.

[3]. Harish Chandra Mohanta, Rajat Kumar Mahapatra and Jyotirmayee Muduli(2014), “ Anti-Theft Mechanism System with Accidental Avoidance and Cabin Safety System for Automobiles”. An anti-theft system is any device or methid used to prevent or detect the unauthorized appropriation of items considered valuable. Theft is one of the most common and oldest criminal behaviours.

[4]. Sudarsan K and Kumaraguru Diderot P (2014), “Helmet for Road Hazard Warning with Wireless Bike Authentication and Traffic Adaptive Mp3 Playback”. Helmet for road hazard warning is designed with wireless bike authentication and traffic adaptive mp3 playback. The main aim of this project is to encourage people to wear helmet and prevent road accidents, which is achieved. Thus road accidents can be prevented to some extent and safety of bike riders is ensured.

[5]. Safety measures for “Two wheelers by Smart Helmet and Four wheelers by Vehicular Communication”, The small voltage of ignition of the two wheelers is grounded. In normal condition when the helmet is wearied the pressure is senses pressure and the RF transmitter radiates the FM modulated signal.

[6]. Nitin Agarwal Anshul Kumar Singh, Pushpendra Pratap Singh, Rajesh Sahani, “SMART HELMET”, International Research Journal of Engineering and Technology, volume 2, issue 2, May 2015, “ Next generation motor cycle helmet with sound control and 360 degree vision that will transform your ride. The cross helmet X1 is a revolutionary motor cycle helmet that will transform your ride.

[7]. D Kumar, S Gupta, S.Kumar, s.Srivastava “Accident detection and reporting system using GPS and GSM module”, It aims at finding the occurrence of any accident and reporting the location of the accident to the previously coded numbers so that immediate help can be provided by ambulance or the relative concerned.. Jennifer William, Kaustubh Padwal, Nexon Samuel, Akshay Bawkar Smita Rukhande, “Intelligent Helmet”, The intelligent helmet band is an idea which makes motor cycle driving safer than before. This is implemented using GSM and GPS technology. Limit switch is placed in the helmet which will detect whether the rider has worn the helmet or not. If not the bike will not start.

3. EXISTING SYSTEM

In existing system, if the person met with an accident we may not ensure the fast first aid treatment; the person may die due to late medication. By using this proposed system, it sends an automatic alert message to the authorized person or ambulance in case of an accident or any emergency situations. The alert message body contains the place and time of the consequences to speed up the first aid service to the victim.

The other existing system is to control the speed in which the biker is going in. The helmet is fixed with all the components and sensors that read the speed of the bike and accordingly instruct the rider to reduce or increase the speed based on the obstacles ahead the bike. It also contains wireless communication which enables the rider to listen to music while he is riding and it also helps in finding the best through the Google map facilities.

Likewise there is also a device which is used to detect the position of helmet ie whether the helmet is worn or not. Alcohol sensing is also available. The drawbacks of existing system are listed below.

- Checking whether the person worn helmet or not is done at initial stage only. If incase helmet is removed later their will not effect on it.
- Same follows in case of alcohol detection, alcohol content is tested only during the starting of the bike.
- Maintenance of temperature inside the helmet is not done.

4. PROPOSED SYSTEM

The helmet checks if the rider is alcoholic even while he is driving. If the rider drunk and then tries to ride the bike the ignition of the bike is avoided and hence not letting the rider to start the bike. In this system we use Node MCU Microcontroller interfaced with alcohol sensor and it is used to monitor user’s breath continuously and sends signals to microcontroller. The microcontroller on encountering alcohol signal from the gas sensor and send the data to motor using RF transmitter and we connect a RF receiver to motor driver which stops to demonstrate as engine locking. The system need push button to start engine. If the alcohol is detected the system locks the engine.

We are also implementing the temperature monitoring system that is fixed inside the helmet which allows the rider to stay cool all the time whenever he wears the helmet to avoid the itching sensation while riding the bike. Advantages of proposed system are.

- User friendly.
- Through GSM module we send message along with current location of the victims so that they can be easily found in case of emergency.
- Through GPS we can even find our bike if its stolen.

5. Proposed Framework

The flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task.Here in our project we consider 4 conditions first and foremost checking whether the helmet is worn. Second the determination of the alcohol content in riders breath. Thirdly checking whether the accident occurred or not and finally checking whether the rider has fallen or not. Based on these conditions the further steps are followed.

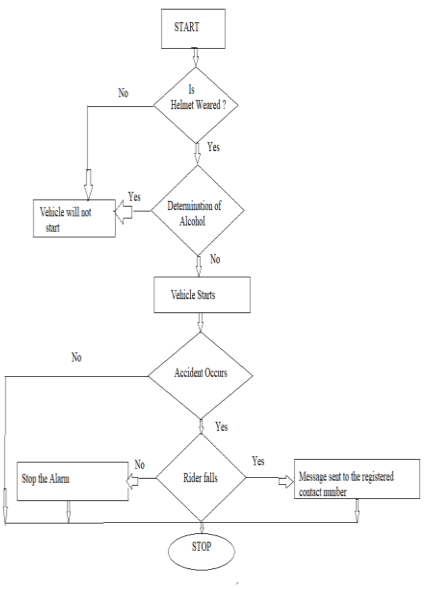

Fig:5.1 Flowchart of the project

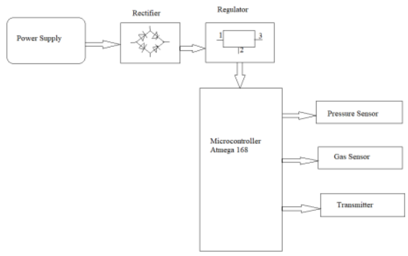

Fig 5.2. Bike unit

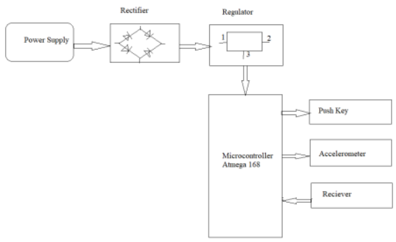

Fig 5.3 Helmet unit

The proposed block diagram shows how the model is implemented. Initially the wearing of helmet by the rider is taken into consideration. Further this input is taken as input for preprocessing unit using algorithm and extracts the features for better accuracy

6. Specifications

A. Pressure Sensor

To measure any gases or liquids we use a device called as pressure sensor. Pressure is generally referred as the force which is required to stop the fluid that's being expanding. The SI unit of pressure sensor is force per unit area. It normally act as a transducer, that generates a signal as a symbol of the pressure imposed. Pressure sensors are used in various fields for controlling and monitoring numerous application. These are normally used in the systems to measure variables like water level, speed, gas/fluid flow and altitude. In our project it is used to detect weather the rider wear the helmet or not. If the rider refuses to wear the helmet bike will not start and intimates the rider to wear the helmet. The pressure sensor is connected with the Arduino in the helmet unit. This sensor works by bending the sensor itself. As the sensors are being flexed or bent, the resistance across the sensor increase. The greater the angle of bending, the greater the resistance. This can be tested with multi-meter.

B. MQ-3/ Gas Sensor

Alcohol sensor module – MQ3. This module is made using alcohol gas sensor MQ3. It is a low cost semiconductor sensor which can detect the presence of alcohol gases at concentrations from 0.05mg/L to 10mg/L. When a drunk person breathes near the alcohol sensor it detects the ethanol in his breathe and provides an output based on alcohol concentration, if there is more alcohol concentration more LED's would lit. This sensor provides an analog resistive output based on alcohol concentration. Normally alcoholic limit in our blood is 0.08 in case of alcoholic person it would increase to 0.065 after one hour. It usually take five hours twenty minute to our body to get completely metabolize alcohol and eliminate from the body. And hence till five hours the bike will in OFF state.

C. Accelerometer ADXL335

The ADXL335 is a small, thin, low power, complete 3-axis accelerometer with signal conditioned voltage outputs. The sensor can measure the static acceleration of gravity in tilt sensing applications as well as dynamic acceleration resulting from motion, shock, or vibrations. The acceleration of the bike is measured by this device. The variations and frequency change in the accelerometer gives the bandwidth of the accelerometer.

D. Node MCU

Node MCU is an open-source firmware and development kit that helps us to prototype or build IoT product. It is low cost and WiFi module chip that can be configured to connect to IoT and similar technology. This module can be used to send the message when the person met with an accident with WiFi connected to it.

7. RESULTS

When the pressure sensor becomes flexible, change in resistance occurs and a constant voltage is produced. This voltage is connected to analog pin of arduino. When alcohol is detected, its digital output is connected to the digital input pin of arduino. Here we are using ADXL 335 as a tilt sensor. It has three axes. It is connected to a reset pin. This pin would be set in x-direction and it is according to the variation in tilt to the y and z direction that occurrence of accident is detected. GPS is connected to Arduino digital pin and this arduino is connected to Node MCU and it is from this module that the message concerning the occurrence of accident is sent to a predefined number and corresponding location is also specified. An encoder IC is used in the transmitter section. Its function is to encode all the sensor's output and this signal is transmitted by the RF unit except GPS module. The receiver section consists of relay, ignition, power supply and a decoder IC. The transmitted signals are received by using RF receiver and these signals are decoded using HT12D decoder IC. This decoded output is connected to ignition. An npn transistor is given to the ignition to provide voltage to the base and the emitter is grounded. This base voltage makes the transistor on. Relay which is used acts as a switch and it makes the ignition.

8. CONCLUSION

The two-wheeler safety system developed with smart helmet and intelligent bike system is reliable and aims to help in the prevention, detection and reporting of accidents hence reducing the probability of the drunk drive cases. It also has several advantages compared to the previous systems. Our proposed system gives the primary importance of preventing the accidents and ensures safety for a greater extent in two wheelers. Nowadays, most accident cases occur due to motor bike. The severities of those accidents are increased because of the absence of helmet or by the usage of alcoholic drinks. By implementing this system, a safe two wheeler journey is possible which would decrease the head injuries throughout accidents caused due to the absence of helmet and additionally reduce the accident rate due to drunken driving. A GSM modem is used in this system that will send a message to the predefined numbers that are programmed using microcontroller in case of any accident.

REFERENCES

[1]. Smart Helmet with Sensors for Accident Prevention Mohd Khairul Afiq Mohd Rasli, Nina Korlina Madzhi, Juliana Johari Faculty of Electrical Engineering University Tecnology MARA40450 Shah Alam Selangor, MALAYSIAjulia893@salam.uitm.edu.my)

[2]. Vijay J, Saritha B, Priyadharshini B,Deepeka S and Laxmi R (2011), "Drunken Drive Protection System", International Journal of Scientific & Engineering Research, Vol. 2, No. 12, ISSN: 2229-5518.

[3]. Harish Chandra Mohanta, Rajat Kumar Mahapatra and Jyotirmayee Muduli(2014)", Anti-Theft Mechanism System

with Accidental Avoidance and Cabin Safety System for Automobiles", International Refereed Journal of Engineering and Science (IRJES), Vol. 3, No. 4, pp. 56.

[4]. Sudarsan K and Kumaraguru Diderot P (2014), "Helmet for Road Hazard Warning with Wireless Bike Authentication and Traffic Adaptive Mp3 Playback", International Journal of Science andResearch (IJSR), Vol. 3, No. 3, ISSN (Online): 2319-7064.

[5]. Safety measures for "Two wheelers by Smart Helmet and Four wheelers by Vehicular Communication" Manjesh N 1, Prof. Sudarshan raju C H 2 M Tech, ECEDSCE, JNTUA, Hindupur Email: manjesh405@gmail.com HOD & Asst. Prof. BIT-IT, Hindupur International Journalof Engineering Research and Applications (IJERA) ISSN: 2248-9622 NATIONAL CONFERENCE on Developments, Advances & Trends in Engineering Sciences (NCDATES09th & 10th January 2015) .

[6]. Nitin Agarwal, Anshul Kumar Singh, Pushpendra Pratap Singh, Rajesh Sahani, "SMART HELMET", International Research Journal of Engineering and Technology, volume 2, issue 2, May 2015

[7]. D Kumar, S Gupta, S.Kumar, s.Srivastava "Accident detection and reporting system using GPS and GSM module" May 2015.

International Journal of Computer Applications Technology and Research
Volume 9–Issue 03, 115-124, 2019, ISSN:-2319–8656

Model for Intrusion Detection Based on Hybrid Feature Selection Techniques

Joseph Mbugua Chahira
Department of Information and Computer Science
Garissa University, Kenya

Abstract
In order to safeguard their critical systems against network intrusions, organisations deploys multiple Network Intrusion Detection System (NIDS) to detect malicious packets embedded in network traffic based on anomaly and misuse detection approaches. The existing NIDS deal with a huge amount of data that contains null values, incomplete information, and irrelevant features that affect the detection rate of the IDS, consumes high amount of system resources, and slowdown the training and testing process of the IDS. In this paper, a new feature selection model is proposed based on hybrid feature selection techniques (information gain, correlation, chi squere and gain ratio) and Principal Component Analysis (PCA) for feature reduction. This study employed data mining and machine learning techniques on NSL KDD dataset in order to explore significant features in detecting network intrusions. The experimental results showed that the proposed model improves the detection rates and also speed up the detection process.
Key words cyber attacks, Intrusion detection, feature selection, data mining.

Introduction

Network Intrusion Detection System (IDS) [1] monitors the use of computers and networks over which they communicate , searching for unauthorised use, anamolous behaviour, and attempt to deny users, machines or portions of networks access to the services. Although the intrusion detection systems are increasingly deployed in the computer network, they deal with a huge amount of data that contains null values, incomplete information, and irrelevant features. The analysis of the large quantities of data can be tedious, time-consuming and error-prone. Data mining and machine learning[2] provides tools to select best relevance features subset which improves detection accuracy and removes distractions.

Feature selection problem can be characterised in the context of machine learning [3][4], [5]. Assume that T = D(F,C) is a training dataset with m instances and n features, where D = o_1, o_2, . . . , om and F = f_1, f_2, . . . , fn are the sets of instances and features. C = c_1, c_2, . . . , ck refers to the set of class labels. For each instance oj ∈ D, it can be denoted as a value vector of features, i.e., $o_j = (v_{j1}, v_{j2}, \ldots, v_{jn})$, v_{ji} is the value of o_j corresponding to the feature f_i.. Therefore, feature selection plays an important role in alert correlation through reduction in the amount of data needed to achieve learning, improved predictive accuracy, learned knowledge that is more compact and easily understood and reduced execution time.

The existing feature selection techniques in machine learning can be broadly cassified into two categories i.e wrappers and filters. Wrappers selection techniques evaluate the worth of features using the learning algorithm applied to the data while filters evaluate the worth of features by using heuristics based on general characteristics of the data. Feature selection algorithms can be further differentiated by the exact nature of their evaluation function, and by how the space of feature subsets is explored. Wrappers often give better results in terms of the final predictive accuracy of a learning algorithm than filters because feature selection is optimized for the particular learning algorithm used. However, since a learning algorithm is employed to evaluate each and every set of features considered, wrappers are prohibitively expensive to run, and can be intractable for large databases containing many features. Furthermore, since the feature selection process is tightly coupled with a learning algorithm, wrappers are less general than filters and must be re-run when switching from one learning algorithm to another.

The advantages of filter approaches in feature selection outweigh their disadvantages. Flters execute many times faster as compaired to wrappers and therefore applicable in databases with a large number of features [6]. They do not require re-execution for different learning algorithms and can provide an intelligent starting feature subset for a wrapper incase improved accuracy for a particular learning algorithm is required[7]. Filter algorithms also exhibited a number of drawbacks. Some algorithms do not handle noise in data, and others require that the level of noise be roughly specified by the user a-priori [3], [7]. In some cases, a subset of features is not selected explicitly; instead, features are ranked with the final choice left to the user. In other cases, the user must specify how many features are required, or must manually set a threshold by which feature selection terminates. Some algorithms require data to be transformed in a way that actually increases the initial number of features. This last case can result in a dramatic increase in the size of the search space[3].

The rest of the paper is organized as follows: Section II presents some related researches on intrusion detection which cover the feature selection and data mining. Section III briefly describes the KDD dataset used in this research. Section IV explains the details of the dataset pre-processing phase of the proposed model. The proposed model is presented in Section V. Finally, the experimental results and analysis are presented in Section 6 followed by some conclusions in the final section.

RELATED WORK

Recent study indicates that machine learning algorithms can be adversely affected by irrelevant and redundant training information [8]. The simple nearest neighbour algorithm is sensitive to irrelevant attributes, its sample complexity (number of training examples needed to reach a given accuracy level) grows exponentially with the number of irrelevant attributes[9][10]. Sample complexity for decision tree algorithms can grow exponentially on some concepts (such as parity) as well. The naive Bayes classifier can be adversely affected by redundant attributes due to its assumption that attributes are independent given the class [11] . Decision tree [12], [13] algorithms such as C4.5 overfit training data, resulting in large trees. In many cases, removing irrelevant and redundant information can result in C4.5 producing smaller trees.

As a result, most researchers combines the feature selection and classification algorithms to improve the detection accuracy and make intelligent decisions in determining intrusions. Siraj et al. [16] proposed new, automated and intelligent hybrid clustering model called Improved Unit Range and Principal Component Analysis with Expectation Maximization (IPCA-EM) to aggregate similar alerts as well as to filter the low quality alerts. Panda et al. [2] proposed a hybrid intelligent approach using combination of classifiers in order to make the decision intelligently, so that the overall performance of the resultant model is enhanced. These two models use hybrid classifiers to make intelligent decisions and the filtering process is applied after adding supervised or unsupervised learning techniques to obtain the final decision. Agarwal et al. [47] proposed hybrid approach for anomaly intrusion detection system based on combination of both entropy of important network features and support vector machine.

Madbouly et al.(2014), proposed a relevant feature selection model that selects a set of relevant features to be used in designing a lightweight, efficient, and reliable intrusion detection system. Although, the model achieved good overall detection result; detection results for PROBE, U2R, R2L attack types were low.

Lin et al. (2015) studied the importance of feature representation method on classification process. They proposed cluster centre and nearest neighbour (CANN) approach as a novel feature representation approach. In their approach, they measured and summed two distances. The first distance measured the distance between each data sample and its cluster centre. The second distance measured the distance between the data and its nearest neighbour in the same cluster. They used this new one-dimensional distance to represent each data sample for intrusion detection by a k-nearest neighbour (k-NN) classifier. The proposed approach provided high performance in terms of classification accuracy, detection rates, and false alarms. In addition, it provided high computational efficiency for the time of classifier training and testing

Zhao et al. (2015) proposed a new model based on immune algorithm (IA) and BPNN. The new developed method is used to improve the detection rate of new intruders in coal mine disaster warning internet of things. IA was used to preprocess network data, extract key features and reduce dimensions of network data by feature analysis. BPNN is adopted to classify the processed data to detect intruders. Experiments' results showed the feasibility and effectiveness of the proposed algorithm with a detection rate above 97%.

Methodology

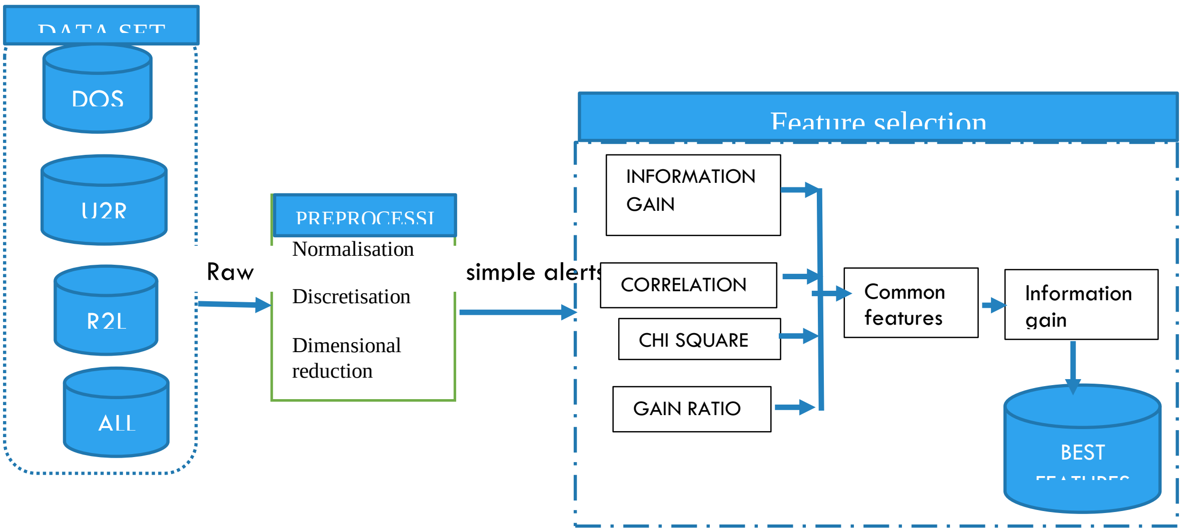

Figure 1 best feature selecion process

The proposed model has four phases as shown in figure 1 above:

Phase 1 data pre-processing

Phase 2 Dimension Reduction.

Phase 3 best feature selection.

Phase 4 Evaluation.

Data Preprocessing

To make efficient use of the available dataset for analysis the data preprocessing is required to provide solutions to Clean the data to remove noise and duplicate information and then deal with any incomplete or missing data an efficient algorithm based on normalization and discretization techniques. Data normalizaion is a process of scaling the value of each feature into a well-proportioned range, so that the bias in favor of features with greater values is eliminated from the dataset [14]. Every attribute within each record is scaled by the respective maximum value and falls into the same range of [0-1]. Normalization follows equation 1,

$$\text{Normalized}(x_i) = \frac{(X_i - X_{min})}{(X_{Max} - X_{min})} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

where X_{min} is the minimum value for variable X, X_{max} is the maximum value for variable X. For a specific symbolic feature, we assigned a discrete integer to each value and then used equation 1 to normalize it.

Discretization transforms continuous valued attributes to nominal [15][16]. The main benefit is that some classifiers can only take nominal attributes as input, not numeric attributes and also some classifiers can only take numeric attributes and hence can achieve improved accuracy if the data is discretized prior to learning.

Recently, organizations use cooperative NIDSs to provide a better detection and global view of intrusion activities. This contributes to the diversity of output formats. In order to correlate alerts such diversified formats have to be converted into a unified standard representation. Intrusion Detection Messaging Format (IDMEF) define a common data formats and exchange procedures for shairing important information to intrusion detection and response system.

Dimension Reduction Using Principal Component Analysis (PCA)

Data reduction algorithms reduce massive data-set to a manageable size without significant loss of information represented by the original data [22]. PCA is a useful technique for dimension reduction and multivariate analysis where the extracted components are statistically orthogonal to each other [13]. This enables speedup of training and robust convergence and hence can be applied in the intrusion alerts dataset to find the principal components of the alerts, i.e., the attributes vector that can describe the alerts exactly and sufficiently, but not redundantly. Mathematically, will establish the principal components of the distribution of the alerts, or the eigenvectors of the covariance matrix of the set of the alerts [23], [24],[21].

The PCA algorithm consists of 5 main steps :

1. Deduct the mean: deduct the mean from each of the data dimensions.

2. Calculate the covariance matrix: Cmxn = (Ci,j, Ci,j = cov(Dimi,Dimj) ....(1) Where Cmxn is a matrix in which all entry is the result of computing the covariance between two distinct dimensions.

3. Compute the eigenvectors and eigenvalues of the covariance matrix.

4. Select components and form a feature vector: once FeatureVector = (eig1, eig2, ..., eigN) ....(2)

5. Derive the new data set. Consider the transpose of the FeatureVector and multiply it on the left of the main data set, transposed:

FinalData = RowFeatureVector x RowDataAdjusted ... (3)

where RowFeatureVector is the matrix with the eigenvectors in the columns transposed and RowDataAdjusted is the mean-adjusted data.

Feature Selection Techniques

The feature selection techniques help to identify some of the important attributes in a data set, thus reducing the memory requirement, increase the speed of execution and improves the classification accuracy[17]. The purpose of this work is to find out which data feature selection algorithm gives better results with decision trees classifiers. Several feature subset selection techniques have been used in data mining.

i. Correlation based feature selection (CFS)

CFS is considered as one of the simplest yet effective feature selection method which is based on the assumption that features are conditionally independent given the class, where feature subsets are evaluated according to a correlation based heuristic evaluation function.[18]. A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other. The major advantage of CFS, it is a filter algorithm, which makes it much faster compared to a wrapper selection method since it does not need to invoke the learning algorithms [19],[20].

$$\rho(X,Y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2\right]^{\frac{1}{2}}}$$ ....(2)

Pearson's correlation coefficient (2), where all variables have been standardized shows that the correlation between a composite and an outside variable is a function of the number of component variables in the composite and the magnitude of the inter-correlations among them, together with the magnitude of the correlations between the components and the outside variable.

ii. Information Gain

Information gain is used as a measure for evaluating the worth of an attribute based on the concept of entropy (1), the higher the entropy the more the information content. Entropy can be viewed as a measure of uncertainty of the system. The largest mutual information between each feature and a class label within a certain group is then selected (2). The performance evaluation results show that better classification performance can be attained from such selected features [21],[19].

$$-\sum_i P(c_i)\log_2 P(c_i).$$ .... (3)

$$IG(A) = I(D) - \sum_{j=1}^{p} \frac{|D_j|}{|D|} I(D_j^A)$$ ....(4)

Algorithm 1: Feature selection according to information gains

Input: A training dataset T = D(F,C), number of features to be selected L

Output: Selected features S

1. Initialize relative parameters: F← fi, i =1, 2, ...n, C←'class labels', S =? ;
2. for each feature fi ∈ F do
   a. Calculate its information gain IG(fi) ;
   b. insert fi into S in descending order with regard to IG(fi) ;
3. Retain first L feature in S, and delete the others ;
4 Return Selected features: S.

iii. Chi-square

Chi-square [19] test is commonly used method, which evaluates features individually by measuring chi- square statistic with respect to the classes. The statistic is

$$\sum_{i=1}^{k} \sum_{i=1}^{k} \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \qquad \text{....................5}$$

Where,

k = No. of attributes,

n = No. of classes,

Aij = number of instances with value i for attribute and j for the class,

Eij = the expected No. of instances for Aij.

The larger value of the χ^2, indicates highly predictive to the class.

Data Set

The experiments will be conducted on MIT Lincoln's Lab's DARPA 2000 Scenario Specific NSL-KDD, 2014 which contains simulated attack scenarios in a protected environment an off-site server. KDD"99 testing set includes 37 attack types that are included in the testing set.

The NSL-KDD dataset has the following advantages over the original KDD dataset [4], [25].

i. It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.

ii. The number of selected records from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification rates of distinct machine learning methods vary in a wider range, which makes it more efficient to have an accurate evaluation of different learning techniques.

iii. The numbers of records in the train and test sets are reasonable, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research works will be consistent and comparable.

The simulated attacks in the NSL-KDD dataset fall in one of the following four categories[9], [24], [26], [27].

i. Denial of service attack (Dos), where attempts are to shut down, suspend services of a network resource remotely making it unavailable to its intended users by overloading the server with too many requests to be handled. e.g. syn flooding. Relevant features includes source bytes and percentage of packets with errors. Examples of attacks includes back,land, neptune, pod, smurf, teardrop

ii. Probe attacks, where the hacker scans the networkof computers or DNS server to find valid IP, active ports, host operating system and known vulnerabilities with the aim discover useful information. Relevant features includes duration of connection and source bytes. Examples includes Ipsweep, nmap, portsweep, satan

iii. Remote-to-Local (R2L) attacks, where an attacker who does not have an account with the machine tries to gain local access to unauthorized information through sending packets to the victim machine exfiltrates files from the machine or modifies in transist to the machine. Relevant features includes number of file creations and number of shell prompts invoked. Attacks in this category includes ftp_write, guess_passwd, imap,multihop, phf, spy, warezclient, warezmaster

iv. User-to-Root (U2R) attacks, where an attacker gains root access to the system using his normal user account to exploit vulnerabilities. Relevant features includes Network level features – duration of connection and service requested and host level features - number of failed login attempts. Attacks includes buffer_overflow, loadmodule, perl,rootkit

Using this format each generated alert is characterised by a set of attributes.

i. Basic attributes. that represent an alert and they are in IDMEF format. Examples of these attributes include timestamp, signature identifier, messages associated with alerts, protocol, IP source and IP destination addresses, source port and destination address, Time to live and identification field.

ii. Content features: The features of suspicious behavior in the data portion should be captured in order to detect attacks. E.g. number of failed login attempts. Those features are called content features. The R2L and U2R attacks normally don't appear in intrusion frequent sequential patterns, as they have been embedded in the data portions of packets and only request a single connection. While the DoS and Probing attacks involve many connections to hosts and show the attribute of intrusion frequent sequential patterns.

iii. Time-based traffic features: Only the connections in the past two seconds are examined, which have the same destination host/service as the current connection, and of which the statistics related to protocol behavior, service, etc. are calculated.

iv. Connection-based traffic features: Some slow probing attacks scan the hosts/service at an internal much longer than two seconds, e.g. once in every minute, which cannot be

v. detected by the time-based traffic features, as it only examines the connections in the past 2 seconds. In such case, the features of same destination host/service connections can be re-calculated at an interval of every 100 connections rather than a time window

Table 1: Description of basic features, content features, traffic features and host-based features.

vi.	Feature	Description	Type
1	Duration	Length of the connection.	Basic Features
2	protocol type	Connection protocol (e.g. tcp, udp)	Basic Features
3	Service	Destination service (e.g. telnet, ftp)	Basic Features
4	Flag.	Normal or error status of the connection	Basic Features
5	source byte	Bytes sent from source to destination	Basic Features
6	destination bytes	Bytes sent from destination to source	Basic Features
7	Land	1 - Connection is from/to the same host/port; 0 – otherwise.	Basic Features
8	Wrong_Fragment	Number of “wrong” fragments	Basic Features
9	Urgent	Number of urgent packets	Content Features
10	Hot	number of “hot indicators”.	Content Features
12	num_failed_logins	number of failed login attempts	Content Features
13	logged_in	1 - successfully logged in; 0 - otherwise	Content Features
14	num_compromise	number of “compromised” conditions.	Content Features
15	root_shell	number of “compromised” conditions.	Content Features
16	su_attempted	1 - root shell is obtained; 0 – otherwise.	Content Features
17	num_root	number of “root” accesses.	Content Features
18	num_file_creations	number file creation operations	Content Features
19	num_shells	number of shell prompts	Content Features
20	Num_access_files	number of operations on access control files	Content Features
21	Num_outbound_cmds	number of outbound commands in a ftp session	Content Features
22	is_hot_login	1 - the login belongs to the “hot” list; 0 – otherwise.	Content Features
23	is_guest_login	1 - the login is a “guest”login; 0 - otherwise	Content Features
24	Count	number of connections to the same host as the current connection in the past 2 seconds	Time-based Traffic Feature
25	srv_count	number of connections to the same service as the current connection in the past 2 seconds	Time-based Traffic Features
26	serror_rate	% of connections that have “SYN” error	Time-based Traffic Features
27	rerror_rate	% of connections that have “REJ” errors	Time-based Traffic
28	same srv rate	% of connections to the same service	Time-based Traffic
29	diff srv rate	% of connections to different services	Time-based Traffic
30	srv_serror_rate	% of connections that have “SYN” errors	Time-basedTraffic
31	srv_rerror_rate	% of connections that have “REJ” errors	Time-based Traffic
32	srv_diff_host_rate	% of connections to different hosts	Time-based Traffic
33	Dst_host_count	count of connections having the same destination host	Host-based Traffic Feature
34	dst_host_srv_count	count of connections having the same destination host and using the same service	Host-based Traffic Feature

35	dst_host_same_srv_ rate	% of connections having the same destination host and using the same service	Host-based Traffic Feature
36	dst_host_diff_srv_rate	% of different services on the current host	Host-based Traffic Feature
37	dst_host_same_src_port_rat	% of connections to the current host having the same src port	Host-based Traffic Feature
38	Dst_host_ srv_diff_host_rate	% of connections to the same service coming from different hosts	Host-based Traffic Feature
39	Dst_host_srv_rerror_rate	% of connections to the current host and specified service that have an S0 error Cont	Host-based Traffic Feature
40	dst_host_serror_rate	% of connections to the current host that have an S0 error	Host-based Traffic Feature
41	dst_host_srv_serror_rate	% of connections to the current host and specified service that have an S0 error	Host-based Traffic Feature

Experimental Setup Results And Discussion

Experiment Setup

Several data mining techniques which includes data cleaning and pre-processing, clustering, classification, regression, visualization and feature selection have been implemented in WEKA (Waikato Environment for Knowledge Analysis) [28]. Weka also offers some functionality that other tools do not, such as the ability to run up to six classifiers on all datasets, handling multi-class datasets which other tools continue to struggle with tools.

In the experiment, we apply full dataset as training set and 10-fold cross validation for the testing purposes. The available dataset is randomly subdivided into 10 equal disjoint subsets and one of them is used as the test set and the remaining sets are used for building the classifier. In this process, the test subset is used to calculate the output accuracy while the N_1 subset is used as a test subset and to find the accuracy for each subset. The process is repeated until each subset is used as test set once and to compute the output accuracy of each subset. The final accuracy of the system is computed based on the accuracy of the entire 10 disjoint subsets.

All experiments are performed using Windows platform with the following configuration Intel Core-i5 processor, 2.5GHz speed, and 8GB RAM.

For our experiment, we selected attribute set based on the repetition of attribute from four scheme. Existing FS that are employed in experiments are Correlation Feature Selection (CFS) based evaluator with Best-first searching method, Gain Ratio (GR) Attributes based Evaluator with Ranker searching method, Information Gain (IG) based Attributes Evaluator with ranker searching method, and Chi Squared Eval and Ranker searching method we obtained-.

Table 2: The most important features to distinguish between normal network traffic and cyber-attacks.

Feature selection techniques	No of features	Selected Features
Original Dataset	41	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35, 36,37,38,39,40,41
Information gain, ranker	10	2, 40,3,41,27,26,30,31,32,35
CFS , best first,	8	2,3,9,23,26,27,34,41
Gain ratio and ranker	9	9,23,41,22,36,3,27,35,2
Chi Squared Eval + Ranker	9	2,40,3,41,26,27,30,31,32
Proposed	11	2,3,4,26,27,36,39,41

Table 3: The most important features to distinguish between normal network traffic and DoS attacks.

Feature selection techniques	No of features	Selected Features
Original Dataset	41	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35, 36,37,38,39,40,41
Information gain, ranker	10	2, 40,3,41,27,26,30,31,32,35

CFS , best first,	8	2,3,9,23,26,27,34,41
Gain ratio and ranker	10	9,23,41,22,36,3,27,35,2,26
Chi Squared Eval + Ranker	10	2,40,3,41,26,27,30,31,32,20
Proposed	9	2,3,9,26,41.4,27

Table 4: The most important features to distinguish between normal network traffic and Probing attacks.

Feature selection techniques	No of features	Selected Features
Original Dataset	41	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35, 36,37,38,39,40,41
Information gain, ranker	10	2, 40,3,30,,34,9,33,32,31,38
CFS , best first,	9	2,3,9,24,26,30,34,38,40
Gain ratio and ranker	10	25,9,24,3,2,41,38,40,34,26,
Chi Squared Eval + Ranker	10	2,40,3,33,34,30,32,38,31,37
Proposed	7	2,3,9,30,34,38,40

Table 5: The most important features to distinguish between normal network traffic and R2L attacks

Feature selection techniques	No of features	Selected Features
Original Dataset	41	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35, 36,37,38,39,40,41
Information gain, ranker	10	1,2, 40,3,30,7,33,40,21,20,34,11
CFS , best first,	5	1,2,7,8,33
Gain ratio and ranker	10	1,8,7,19,2,3,33,40,21,20,34,11
Chi Squared Eval + Ranker	10	1,2,7,3,40,33,19,34,30,29,21
Proposed	9	1,2,7,33,3,40,34,30,21

Table 6: The most important features to distinguish between normal network traffic and U2R attacks.

Feature selection techniques	No of features	Selected Features
Original Dataset	41	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,31,33,34,35, 36,37,38,39,40,41
Information gain, ranker	10	11,40,3,30,10,29,14,1,33,21
CFS , best first,	4	6,11,29,30
Gain ratio and ranker	10	6,11,10,14,13,3,29,30,1,33
Chi Squared Eval + Ranker	10	6,11,3,10,14,40,30,29,31,1
Proposed	7	6,11,29,30,3,10,14

Table 7: The best set of relevant features

ALL	8	2,3,4,26,27,36,39,41
DOS	9	2,3,9,26,41,4,26,27,41
PROBE	7	2,3,9,30,34,38,40
R2L	8	1,2,7,33,3,40,34,30,21
U2R	7	6,11,29,30,3,10,14
Best Feature	12	1,2,3,9,26,27,29,30,34,36,39,40

Results Analysis and Discussion

As indicated by Table 7, the 41-features were reduced to 12-features. Features (6,11,29,30,3,10,14) are relevant for U2R, Features (1,2,7,33,3,40,34,30,21) are relevant for R2. Features (2,3,9,30,34,38,40) are relevant for PROBE class. Features (2,3,9,26,41,4,26,27,41) were selected as relevant DOS.

The category with least relevance features for detecting cyber-attacks is the Content Based Features. This results are biased because of the distribution of train and test datasets where the U2R and R2L attacks is less frequent than the most important features for DoS and Probing attacks. This implies that it is not necessary to analyse contents in the network traffic packages to detect cyber-attacks. The aspect of privacy and integrity for employees is then protected. A content feature can be the text in an email, and to store and analyse this kind of content violates the employees' integrity.

The Basic Features are most important to analyse to distinguish between normal network traffic and cyber-attacks.These features describes the number of seconds for the connection, the protocol used for the connection, the network service on the destination, normal or error status of the connection and the number of data bytes sent between source and destination computer. The results indicate the significance of analysing basic network traffic features to detect cyber-attacks.

The most important attributes to detect DoS attacks includes Src_bytes, Diff_srv_rate, Service, Dst_bytes and Flag. The content features are the least important category of features to detect a DoS attack. The reason for this is that the DoS attacks mostly consist of either no content or filled with a large amount of useless information.

The Host-based features like dst_host_srv_count, dst_host_serror_rate is important to detect probing attacks. These attacks takes longer time and in different ports and also seek known vulnerabilities.

For R2L attacks, the most important features duration, Src_bytes, Dst_bytes and srv_count. The duration represents the number of seconds for the connection and several R2L attacks have a duration which is much larger than a normal connections. The time-based feature, srv_count, which represent the number of connections have a low value compared to normal network traffic. During a R2L attack, the attacker tries to gain access to a local user account with a specific service in connections longer than 2 seconds.

To detect U2R attacks the most important feature includes Service, num_failed_logins, root_shell since the U2R attacks involves the use of specific services for remote access, often in combination with a file transfer service. Compared to the other attack categories, the content features are very important to detect U2R attacks. The content features are created by analysing the content in a network connection. The importance of content features to detect U2R attacks are as a result of the remote users actions that can only be noticed when analysing the content in the connection packages.

Conclusions

The most important features to detect cyber-attacks are basic features such as source byte, destination byte, the used service, a flag to indicate the status of the connection. Moreover time-based traffic features is important to analyse and detect cyber-attacks, such as information about the percentage of connections in the past 2 seconds with a different service than current connection. To detect R2L and U2R attacks it is important to study content features.

Refferences

[1] M. M. Siraj, H. Hussein, T. Albasheer, and M. M. Din, "Towards Predictive Real-time Multi-sensors Intrusion Alert Correlation Framework," *Indian J. Sci. Technol. ISSN*, vol. 8, no. 12, pp. 974–6846, 2015.

[2] M. C. Belavagi and B. Muniyal, "Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection," *Procedia Comput. Sci.*, vol. 89, pp. 117–123, 2016.

[3] J. Song, "Feature Selection for Intrusion Detection System Jingping Song Declaration and Statement," p. 132, 2016.

[4] N. A. Biswas, F. M. Shah, W. M. Tammi, and S. Chakraborty, "FP-ANK: An improvised intrusion detection system with hybridization of neural network and K-means clustering over feature selection by PCA," *2015 18th Int. Conf. Comput. Inf. Technol. ICCIT 2015*, pp. 317–322, 2016.

[5] J. H. Assi and A. T. Sadiq, "NSL-KDD dataset Classification Using Five Classification Methods and Three Feature Selection Strategies," vol. 7, no. 1, pp. 15–28, 2017.

[6] M. Othman and T. Maklumat, "Mobile Computing and Communications: An Introduction," *Malaysian J. Comput. …*, vol. 12, no. 2, pp. 71–78, 1999.

[7] K. Kumar, "Network Intrusion Detection with Feature Selection Techniques using Machine-Learning Algorithms," vol. 150, no. 12, pp. 1–13, 2016.

[8] N. A. Noureldien and I. M. Yousif, "Accuracy of Machine Learning Algorithms in Detecting DoS Attacks Types," vol. 6, no. 4, pp. 89–92, 2016.

[9] A. Thesis, "Using Support Vector Machines in Anomaly Intrusion Detection by," 2015.

[10] P. Verma, "Performance of Detection Attack using IDS Technique," vol. 4, no. 3, pp. 624–629, 2016.

[11] J. Juanchaiyaphum, N. Arch-int, and S. Arch-int, "A Novel Lightweight Hybrid Intrusion Detection Method Using a Combination of Data Mining Techniques," *Int. J. Secur. its Appl.*, vol. 9, no. 4, pp. 91–106, 2015.

[12] P. Manandhar, "A Practical Approach to Anomaly - based Intrusion Detection System by Outlier Mining in Network Traffic By," 2014.

[13] A. I. Madbouly, A. M. Gody, and T. M. Barakat, "Relevant Feature Selection Model Using Data Mining for Intrusion Detection System," *Int. J. Eng. Trends Technol.*, vol. 9, no. 10, pp. 501–512, 2014.

[14] M. A. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and U. T. Nagar, "A Novel Feature Selection Approach for Intrusion Detection Data Classification," *2014 IEEE 13th Int. Conf. Trust. Secur. Priv. Comput. Commun.*, pp. 82–89, 2014.

[15] D. a. M. S. Revathi, "A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection," *Int. J. Eng. Res. Technol.*, vol. 2, no. 12, pp. 1848–1853, 2013.

[16] S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014*, no. December, pp. 1348–1353, 2014.

[17] Z. Dewa and L. A. Maglaras, "Data Mining and Intrusion Detection Systems," *Int. J. Adv. Comput. Sci. Appl.*, vol. 1, no. 1, p. 1:7, 2016.

[18] Y. Wahba, E. ElSalamouny, and G. ElTaweel, "Improving the Performance of Multi-class Intrusion Detection Systems using Feature Reduction," *Ijcsi*, vol. 12, no. 3, pp. 255–262, 2015.

[19] V. Barot, S. Singh Chauhan, and B. Patel, "Feature Selection for Modeling Intrusion Detection," *Int. J. Comput. Netw. Inf. Secur.*, vol. 6, no. 7, pp. 56–62, 2014.

[20] M. B. Shahbaz, X. Wang, A. Behnad, and J. Samarabandu, "On Efficiency Enhancement of the Correlation-based Feature Selection for Intrusion Detection Systems," 2016.

[21] A. AliShah, M. Sikander Hayat Khiyal, and M. Daud Awan, "Analysis of Machine Learning Techniques for Intrusion Detection System: A Review," *Int. J. Comput. Appl.*, vol. 119, no. 3, pp. 19–29, 2015.

[22] I. Syarif, A. Prugel-Bennett, and G. Wills, "Unsupervised clustering approach for network anomaly detection," *Networked Digit. Technol.*, vol. 293, 2012.

[23] S. Mallissery, S. Kolekar, and R. Ganiga, "Accuracy Analysis of Machine Learning Algorithms for Intrusion Detection System using NSL-KDD Dataset," vol. 4, no. 1, 2014.

[24] N. Shahadat, I. Hossain, A. Rohman, and N. Matin, "Experimental Analysis of Data Mining Application for Intrusion Detection with Feature reduction," pp. 209–216, 2017.

[25] L. Dhanabal and S. P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 446–452, 2015.

[26] A. Jain and J. L. Rana, "Classifier Selection Models for Intrusion Detection System (Ids)," *Informatics Eng. an Int. J.*, vol. 4, no. 1, pp. 1–11, 2016.

[27] M. R. Parsaei, S. M. Rostami, and R. Javidan, "A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset," vol. 7, no. 6, pp. 20–25, 2016.

[28] M. Govindarajan and R. Chandrasekaran, "Intrusion Detection using an Ensemble of Classification Methods," *Proc. World Congr. Eng. Comput. Sci.*, vol. I, no. October, 2012.