International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 7 (2018) pp. 5384-5392

A Survey on Applications of Data Mining Techniques

B V Chowdary
Assistant Professor,
Department of Computer Science & Engineering
Vignan Institute of Technology & Science, Hyderabad, India.

Dr. Y. Radhika
Professor,
Department of Computer Science & Engineering
GIT, GITAM University, Visakhapatnam, India.

Abstract

In the current era, huge amount of data is being produced by many sources, i.e. science, business, medicine, sports, geography, environment etc. This generated data is in unstructured, massive sized and raw format, thus not much useful. So, the need arises for some techniques with which, the useful data can be extracted. Data mining helps to extract the useful data from big databases. It deals with extraction of an implicit, previously not known and potentially useful information from data. It also requires programs that detect regularities and patterns in the data. In past years, machine learning methodologies have been successfully used for a wide range of real world application scenarios.

This paper presents an extensive literature review of latest advances in data analysis methods that may be applied to fetch useful knowledge of various optimization and classification problems which may be utilized for processing massive sized raw and unstructured data. Knowledge discovery equips the tools to automate the entire operation of data analysis and optimal hypothesis selection. This paper discusses various knowledge learning methods. Finally, significant open problems in this domain are outlined and our further research directions. Along with this, the dimensionality reduction problem has specifically considered and have thrown some light on it in section 5.

Keywords: Machine learning, Data mining, Knowledge discovery, Support Vector Machine, Artificial Neural Networks, Classification.

INTRODUCTION

In the world of Internet of Things, a huge amount of data is generated from several sources. Data may be present in the structured or unstructured form. Data mining [3] deals with the unstructured, erroneous and incomplete form of data [51]. Objective of data mining is to search consistent patterns, systematic relationships between data, validate the finding by applying the detected pattern to new subset of data and predict new findings on new datasets. Machine learning [11][12] is a technical approach for data mining. Data mining domain involves mainly five classes of tasks, which are - *(i) Association Rule Learning (ii) Anomaly detection (iii) Clustering (iv) Classification (v) Regression.*

Data mining, sometimes, can be viewed as a complex task, as the algorithms utilized may possess very complex nature and the data may not always present at single place. It requires being integrated from several heterogeneous data sources. The prime issues are - *mining methodology adopted* and *user interaction, diverse data types issues* and *performance issues.* Knowledge discovery is another vital terminology, used in data mining literature. The processes involved in the knowledge discovery are - Data Cleaning, Data Selection, Data Integration, Data Transformation, Knowledge Presentation and Pattern Evaluation. Data mining [1] can be thought of as the subset of knowledge discovery. Data mining systems may combine techniques [15] from the following sources e.g. Spatial Data Analysis, Image Analysis, Signal Processing, In-formation Retrieval, Pattern Recognition, Computer Graphics, Web Technology [6]. The learning algorithms customized from previous experiences with the probability theory, principles of statistics, combinatorial optimization, reinforcement learning, logic, control theory and search etc. There are multiple applications of machine learning techniques such that forecasting, language processing, vision processing, pattern recognition, game theory, expert systems, robotics etc.

The most common methods are Supervised learning and Unsupervised learning [51][52], shown as fig.1.

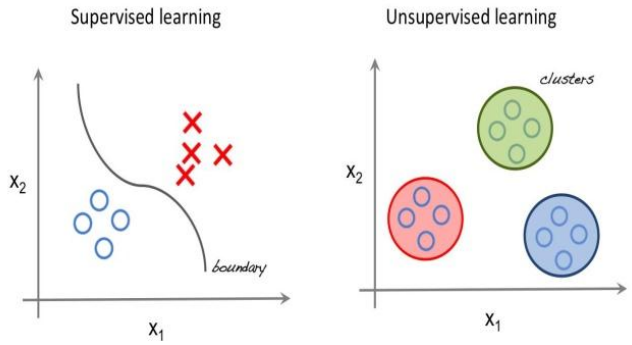

Figure 1. Supervised and Unsupervised learning scenario

Supervised learning deals with make the function learn from the available training part of dataset. A supervised learning algorithm exploits the available training data part and makes an inferred function, which can be then exploited further for mapping new ones. Multiple supervised learning algorithms are available such as ”Support Vector Machines, Neural Networks and Naive Bayes classifiers”.

Unsupervised learning deals with unlabelled data without taking any previously-defined dataset for model training. Unsupervised learning can be thought as a potent tool for look

for patterns and trends and analyzing available data. There are various approaches used by unsupervised learning e.g. "K-means clustering, hierarchical clustering, self-organizing maps" etc.

A. Motivation to the problem

The volume of data today, is raging at an unusual rate as a result of advancements and developments in Web technologies, social media, and mobile devices etc. Traditional strategies are hardly suffering when faced with this massive sized data. Therefore, computationally efficient and practically applicable data cleaning, pre-processing and mining methods are need of the hour.

In present day applications, many nursing homes handle medical information using health care information system; when the system contains enormous instances of data, used to excerpt the secret clue for preparing an imaginative medical analysis. The major objective of this analysis is to design a creative Prediction System that provides analysis of disease using factual body parameters related data. To establish this system, patient's medical conditions such as - heart rate, blood pressure, and cholesterol like several aid aspects are utilized.

B. Organization of the paper

Section 2 introduces the term - data mining, various significant methods for classification are summarized. Different kinds of knowledge learning methods are presented in section 3 .An extensive literature survey is presented in section 4. Some of the open issues in this domain along with our further research directions regulation are given in section 5. Finally, section 6 provides conclusive summary of the paper.

DATA MINING

Data mining is a form of knowledge discovery. In data mining, there are three main approaches - classification, regression and clustering. In these approaches, instances are combined into identified classes [2]. Categorization is useful to examine and study existing sample dataset as well as predicts the expected behaviour of dataset. It consists of two phases. First, learning phase and second is testing of data [3]. In the learning phase, model will analyse the training data then patterns and rules are generated. Later, in testing phase, it will test data then archives the classification pattern accuracy. In classification technique, several machine learning algorithms can be used such that Naive Bayes classifier, Support Vector Machine, Lin-ear Regression, Decision Tree, K-means clustering, Logistic Regression, Artificial Neural Networks etc. Regression is used for mapping. It maps data items into prediction variable, while clustering approach depends on unsupervised learning so that there are no classes exists which are predefined. In this, data will pool as a form of cluster [4].

A. Significant classification and clustering procedures

Some of the significant and practically used classification and clustering procedures are summarized below-

1) *Support vector machine:* SVMs [7] are supervised learning model algorithms that identify data used for classification and regression analysis. It can predict the class of new sample through classifier training. SVM is the approximate implementation of SRM (Structural Risk Minimization) principle. It is first practical implementation in the early 90's. There are two types of SVM technique, i.e." mathematical programming and kernel function". Mathematical programming searches an optimal hyper-plane between different classes data points in high dimensional space. Kernel function, nonlinear classification, completely mapping inputs into high dimensional space.

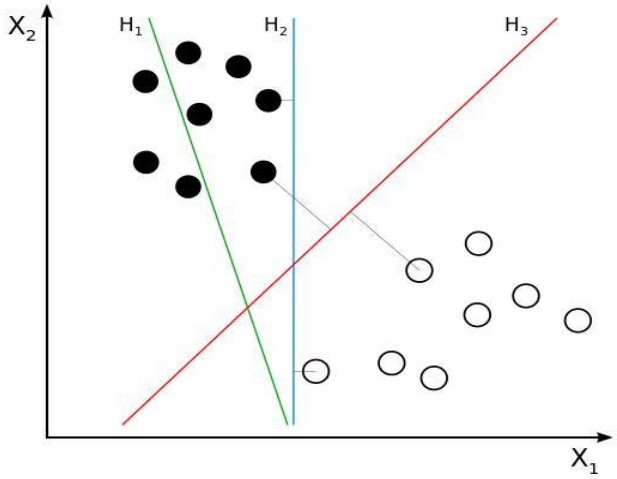

Figure 2. Support vector machine

In fig.2, H_1 hyper-plane is not separating the classes, H_2 hyper-plane is separating the classes but only with a small margin and H_3 is separating the classes with maximum margin.

2) Artificial Neural networks: Artificial Neural networks are registering systems inspired toward biological neural net-works. Artificial Neural Networks [5] are made out of several nodes. Each node indicates as biological neurons of the hu-man brain. Associated by several links, neurons communicate with each other. The nodes will accept input data and perform straightforward operations on the data. The consequence of these operations is passed to different neurons. Each node output is called as node value or activation. A simple ANN structural form is represented as fig.3 –

Each link will have some weight. ANNs are capable of learning, which takes place by modifying weight values. There are two types of ANNs structures - feedforward and feedback.

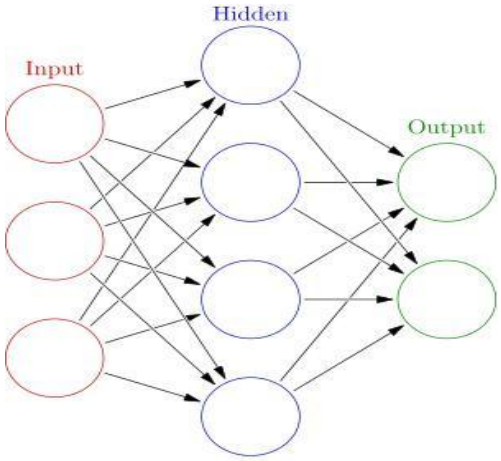

Figure 3. Artificial Neural Networks

3) Decision Tree: A Decision tree [3] could be an structure that comes with a root end, branches, and leaf ends. Root node is that the top node within the tree. In 1980, J. Ross Quinlan proposed a decision tree algorithm named as ID3(Iterative Dichotomiser). Further, He proposed an extension of ID3. This algorithm follows the greedy approach. No backtracking is available in this algorithm. Decision tree example for an instance is shown in fig.4 -

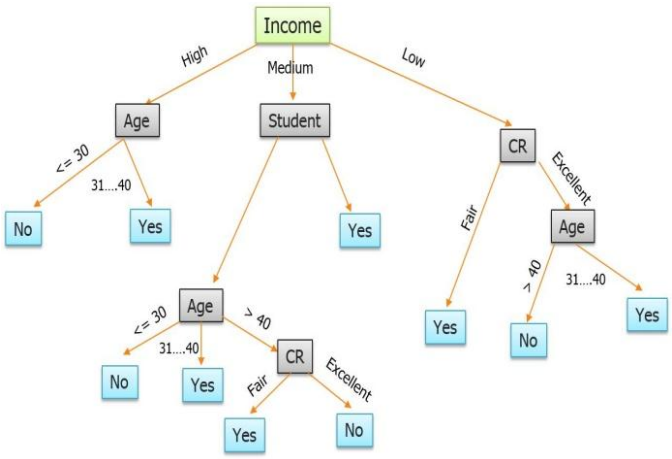

Figure 4. An Example Decision tree

4) K-means clustering: K-means clustering [8] could be a quite unattended learning that is employed for information while not outlined groups or categories (i.e. unlabelled data). Clustering is a key tool for understanding sample data. The objective of this algorithm is to search groups in the data and variable K(number of groups or cluster). In this algorithm, each data point is assigned with one of K groups on iteration basis. Clustering on data point is based on feature similarity. The output of the K-means grouping algorithm are the K clusters centroid and labels for the training data.

The K cluster centroid is used to label new data. In labels for the training data, each data point belongs to a single cluster. The K-means clustering employs iterative refinement to process a final result. The inputs values are K(i.e. number of cluster) and dataset. The dataset indicates a collection of features for every data point. The algorithm begins with introductory estimates for those centroids. The algorithm iterates between two phases such that data assignment and centroid update. It iterates between phase one and two until the point, when a halting criterion is met. The result of this procedure might be a local optimum. K-means clustering algorithm applies in some applications, such as inventory categorisation, behavioural segmentation, detecting bots or anomalies, sorting sensor measurements etc.

B. Interdisciplinary nature of Knowledge Discovery

The data-mining element of KDD presently depends heavily on famed techniques from machine learning, pattern recognition, and statistics to seek out patterns from information within the data-mining step of the KDD method. Algorithms will be scaled to huge information sets and still run with efficiency, however results will be understood and pictured, and the way the man-machine interaction will usefully be sculptured and supported. KDD places viewpoint on finding perceivable patterns which will be understood as helpful or fascinating data. Thus, for instance, neural networks, though a robust modelling tool, are comparatively troublesome to grasp compared to decision trees.

Knowledge discovery from information is essentially a applied statistics endeavour. Most data-analytics algorithmic procedures from" statistics, pattern recognition, and machine learning" taken consideration that information aspects are within the main memory and pay very less attention to however the rule formulates if solely restricted views of the information are potential.

MACHINE LEARNING APPROACHES IN DATA MINING

A proper information illustration may be a basic part of all information discovery processes. Most of the architectures may be classified into 2 broad varieties - "implicit and explicit". An implicit illustration doesn't have a proper sketch and, hence, the associated information can't be transferred unambiguously. Data processing and information discovery may be a immense space of analysis and there's lots of techniques which will turn out implicit and explicit information. Numerous learning ways are summarized as below: -

A. Active learning

It chooses a subset of an unstructured and critical occurrence for purpose of labelling. The active learner [9][20] obtains larger accuracy using reduced number of occurrences.

B. Kernel-based learning

It is proven to be a dominant methodology to efficiently enhance the computational capacity. It is advantageous in

terms of that, both linear as well as non-linear vector kernel functional methods [10] are present to deal with the non-linearity of data in N-dimensional feature space.

C. Transfer learning

It is mainly beneficial in the sense that it can efficiently apply knowledge, which has been learned previously in order to find solution for new problems in fast and effective manner [11].

D. Distributed learning

This kind of learning inhibits the cluster formation, in which one processing thread is assigned to each cluster in order to perform multi-threading in parallel and distributed manner [12].

E. Deep learning

Deep learning [13][14] considers more complicated, compartmented statistical patterns of inputs and manages to be robust for new fields as compare to traditional learning systems. Two deep learning methodologies are -" Deep belief networks (DBNs) and convolutional neural networks (CNNs)".

F. Association rule learning

Association rule learning(ARL) [16] is an approach for locating some attention-grabbing relations between the variables in giant databases.

G. Inductive logic programming

It is an approach to rule learning victimisation logic programming as a regular illustration for input examples, information, and hypotheses [17].

H. Bayesian networks

These are a form of probabilistic graphical model that represents a collection of random variables and corresponding conditional independencies via a directed acyclic graph (DAG) [18].

I. Reinforcement learning

Reinforcement learning [19] is regarded with however associate degree agent have to be compelled to opt for actions in associate degree setting thus on create some notion of long-run reward maximum. These learning algorithms endeavour to seek out a policy that maps states of the globe to the actions the agent have to be compelled to absorb those states. The diagrammatic representation is given in fig.5

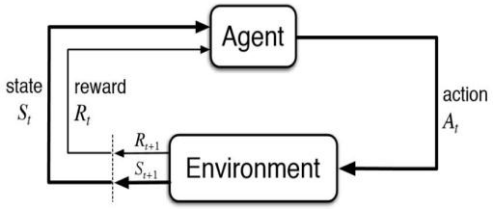

Figure 5. Reinforcement learning framework

J. Similarity and metric learning

Here, the" training machine is equipped pairs of examples that are purportedly similar and pairs of less connected objects. It then needs learning a similarity functional operator (or a distance metric function) that may predict if new objects are similar". It's generally practiced in Recommendation engines [21].

K. Genetic algorithms

A genetic rule is a hunt heuristic that aims simulates the way of action and employs strategies like - "mutation and crossover" to make new genotype within the hope of obtaining smart solutions for a given problem. In ML, genetic algorithms gained some applications within the Eighties and Nineteen Nineties [22][23].

L. Rule-based machine learning

Rule-based learning may be generally thought of as, any machine learning system that acknowledges various rules to cache, manipulate the gained information. The optimally oriented process feature of a rule-based machine learner is that the description and utilization of a group of relative precepts that together define the information gained by the system [24].

LITERATURE REVIEW

This section extensively represents the research work and developments that has taken place in past years.

In 2003, Abonyi et. al. [25] proposed a" supervised fuzzy classifier logic-based pattern recognition technique". In 1992, Boser et. al. [26] presented the" training procedure for optimal boundary classifier". In 2003, Frohlich et. al. [27] presented the" attributes extraction procedure for support vector ma-chines, utilizing genetic algorithms". Joachims et. al. [28] proposed" some significant procedures for text categorization using SVM". Osuna et. al. [29] performed the support vector machine(SVM) concept in the process of face detection. In 1982 (Pawlak, 1982) proposed Rough Set (RS) concept [30][31][32] as a mathematical model to extract and represent knowledge along with to deal with uncertainty. In 2010, U. Stanczyk [33] presented the rough set-based features analysis for ANN classifier.

Geraldin B. Dela Cruz et. al. [34] described about" hybrid data mining method which is based on PCA-GA. Both the

classifiers are used as fitness function in GA and also in data mining classification process in which performance is increased". Junfei Qiu et. al. [35] given the" literature survey of latest technologies in machine learning for big data processing". Chintan Amrit et. al. [36] described" how text mining and analysis is used for identifying and predicting cases of child abuse in the public health institution". Alvaro et. al. [37] proposed" machine learning based method which recommends optimal parameters which is used for task parallelization in big data workloads". Lina Zhou et. al. [38] introduced the" framework of ML on big data along with its opportunities, the framework is centred on ML that follows the phases of pre-processing, learning and evaluation". Filippo Piccinini et. al. [39] described about "Advanced Cell Classifier" for pheno-typic analysis. Boukaye et. al. [40] mainly focused on" remote sensing satellite data processing by using the data mining methods for discovering the risk areas of epidemic disease by correcting the environment and health etc". Huihui Chen et. al. [41] has discovered the "A kernel-based clustering method for gene selection with gene expression data". Valds et. al. [42] applies information and knowledge visualizat-ion with the" virtual reality spaces, neural networks and rough sets".

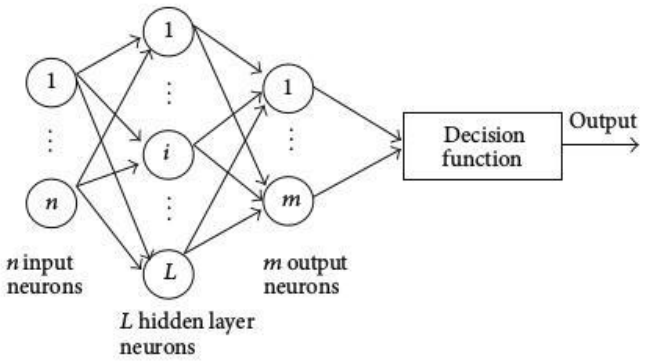

Figure 6. ELM architecture

The extreme learning machine (ELM) was formerly developed by [53][54] and can be categorized as a supervised learning algorithm competent of solving linear and nonlinear classification problems. It is best befitted for higher training samples and conjointly the impact of variety of hidden neurons exploitation totally different ratios of the quantity of options of testing and training information was examined. Fig.6 rep-resents architecture of the ELM.

A. Comparison of some significant machine learning and statistical methods

The comparison of various techniques is presented in Table-1, in which the advantages and drawbacks of some significant data mining methods are given -

Table 1. Comparison table

Advantages and Disadvantages of various techniques		
Name of Technique	**Advantages**	**Drawbacks**
Artificial Neural Nets	*(i) By nature, ANN is self adaptive method.* *(ii) ANN has vast scope in classification process. It can perform multi-class classification.*	*(i) Training phase is time taking, produces less accurate and efficient results in some scenarios. (ii) It is a complex procedure here to select network topology.* *(iii) Every time the results are not unique when run the ANN classifier repeatedly, thus can be considered as big drawback of it.*
Support Vector Machines	*(i) In performing experiments repeatedly on same dataset, SVM by nature always produces same kernel classifier boundary, thus gives advantage over ANN (ii) decrement in the computational complexity of SVM classifier is possible. (iii) Its flexible in the selection of threshold. (iv) Computing error factor is easy.*	*(i) More complexity is involved to understand the structure. (ii) By nature basically SVM is two-class classifier. (iii) Working in higher dimension hyper-planes may give more computational complexity in some cases.*
Fuzzy SVM	*(i) Several stochastic correlations can be identified.* *(ii) Adopting fuzzy logic in SVM, it solves the problem of noise and outliers in kernel functions*	*(i) Requires the pre-knowledge about data sets like - stochastic and probabilistic information.*
Decision Tree	*(i) DT produce many possible outcomes as one can think of due to tree structure.*	*(i) The irrational expectations in decision trees may drive to flaws, errors or missing node value information in resultant tree*

B. Data Mining Developments in Medical Applications

This section throws light on some significant developments done in past in the domain of medical applications. Akay et. al. [43] given the non-invasive detection of coronary artery disease using wavelet-based fuzzy neural networks. Alexopoulos et. al. [44] used inductive machine learning in medical diagnosis of stroke. Dawant et. al. [45] presented a neural network approach to magnetic resonance imaging tissue characterization. Guo et. al. [46] performed the classification of heart sounds in patients with porcine bioprosthetic valves. Lane et. al. [47] applied the neural networks for decision making related to asthma diagnosis and other respiratory disorders. Lim et. al. [48], Strausberg et. al. [49] in their papers, discussed the applications of self-sufficient neural network systems to medical pattern classification tasks. C.V.Subbulakshmi et al.[50] proposed "a Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine classifier". This suggested methodology joins the thought of SRLPSO for optimizing the weights in "ELM neural network". Overview of steps of the proposed approach are as follows –

- " Initialize positions with a set of input weights and hidden biases: W_{11}, W_{12} ...W_{1n}, ..., W_{H1}, $W_{H2,...}$, W_{Hn}, ..., b_1, b_2, …, b_H . These are going to be randomly initialized within the range of [-1,1] on D dimensions within the search space.
- For every member within the cluster, the individual output final weights are computed at ELM as given in below step three.
- Invoke self-regulated learning Particle Swarm Optimization (PSO).
- Further, the fitness, (which is the mean square error (MSE) of each member), is evaluated -

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N} E_i^2$$

In this process, the inputs being taken are - the number of training samples, error of the actual output and target output of the k^{th} output neuron. So, the fitness of each member is adopted as the mean squared error. Then the SRL acceptance depending on the fitness is obtained. Further, update the velocity and position eqns. of each particle as mentioned previously.

- The procedure repeats above steps until certain criteria met, along with hard threshold value as maximum number of iterations". Once it stops, the algorithmic procedure reports values with optimal weights with minimal MSE as its solution.

OPEN ISSUES AND RESEARCH DIRECTIONS

Some of the open issues, problem identification and future research directions are given in below subsections:-

A. Problem Identification

Nowadays, within the Internet-dependent world, a colossal quantity of knowledge is generated from many sources. Of this, a large quantity of knowledge is available in an unstructured format. Analysing the unstructured information exploitation data processing techniques will enable higher decision making. Data processing includes tons of tasks like," document agglomeration, document classification, information summarisation, sentiment analysis, social network analysis, topic detection, online page classification, identification of author, plagiarism detection, spam/malware analysis, patent analysis, monetary decision making" etc. Prime aim of the ma-chine learning algorithms is that they learn from the empirical information and may be utilized in cases that the modelled development are hidden, non-evident, or not nevertheless delineated. Data processing algorithms like support vector machines, rectilinear regression, logistical regression, neural networks, naive bayes and decision trees play a predominant role in this domain.

In recent years, machine learning methodologies have been successfully used for a wide range of real world applications. Machine learning algorithms categorizes the learning task in two types - Supervised learning and Unsupervised learning.

The available unstructured data need to be converted into structured format before starting the data mining process. Predictive analytics and modeling encompasses a variety of statistical techniques from machine learning that analyze the present and historical facts to make the predictions about the future events.

B. Future Research Directions

Our future research directions are outlined as below -

- Processing the bulky sized unstructured and vague data by computing machines is a challenging task. Regarding only digital information, every day, Google processes approx. 24 PB data. Feature choice intends to work out a nominal feature set which might represent a similar information because it was delineated by the initial feature variables.

  New attribute/feature selection algorithms will be proposed as our research work. Further, only the selected relevant features of the data will take part in the classification process. The methodology must also give the guarantee of dimensionality reduction for the given input data.

- Some traditional machine learning techniques perform the feature extraction and classification tasks more accurately only for the complete information systems. Fuzzy logic and Reinforcement learning based optimization ML approaches will be employed, with which may be able to deal with the incomplete and inconsistent information systems, more accurately and efficiently.

- Experiments will be performed using our designed approaches for standard medical datasets from UCI machine learning repository e.g. Wisconsin Breast Cancer, Pima Indians Diabetes, Heart-Statlog, Hepatitis datasets etc. Through simulation, the practicality and computational efficiency of our proposed methods will be analysed.

C. Dimensionality - A challenging problem & overview of approaches to deal it -

The recent outburst of data set size, in amount of records and traits, has triggered the advancement of a number of big data principles as well as parallel data analytics algorithms. At the same time though, it has stirred for practice of data dimensionality reducing procedures. Indeed, more is not always sufficient. Large amounts of data might sometimes produce critical performances in data analytics applications. Currently available techniques to deal with this challenging problem fall into below categories -

- Low Variance Filter
- High Correlation Filter
- Missing Values ratio
- Random Forests / Ensemble Trees
- Principal Component Analysis (PCA)
- Forward Feature selection
- Backward Feature Elimination

Based on our dataset behaviour, an optimal methodology or a hybrid approach (combination of two or more techniques)can be employed in the view of getting reduced dimension in reasonable computational time.

CONCLUSION

Most of the traditional methods are not scalable to manage data with the properties of its massive volume, diverse types, inconsistency, and uncertainty along with incompleteness. This paper is a discussion along with an extensive survey of machine learning techniques for data mining. Various learning methods and some most significant and practically usable procedures for classification and clustering procedures have been discussed. Later in this paper, some open issues, challenges in this domain, problem identification and our future research directions are presented.

REFERENCES

[1] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu and Shuo Feng, A survey of machine learning for big data processing, EURASIP Journal on Advances in Signal Processing. December 2016, 2016:67.

[2] Han, J., Kamber, M., Pei, J. (2006). Data mining: concepts and tech-niques. Morgan kaufmann.

[3] Seema Sharma, Jitendra Agrawal, Shikha Agarwal, Sanjeev Sharma, Machine Learning Techniques for Data Mining: A Survey, 2013 IEEE International Conference on Computational Intelligence and Computing Research.

[4] Merceron A., Yacef K., Educational Data Mining: a Case Study. (2005, May), pp. 467-474.

[5] A. K. Jain, Jainchang Mao, Artificial Neural Networks:Tutorials, Journal of Computers - Special issue: neural computing: companion issue to Spring 1996 IEEE Computational Science & Engineering, Volume 29 Issue 3, March 1996, Page 31-44 .

[6] Eric P. Xing, Qirong Ho, Pengtao Xie, Wei Dai, Strategies and Principles of Distributed Machine Learning on Big Data, Engineering, Volume 2, Issue 2, June, 2016, Pages 179-195.

[7] C. Cortes, V. Vapnik, Support-Vector Networks, Machine Learning, 20, 273-297 (1995).

[8] Kanungo T., Mount D. M., Netanyahu N. S., Piatko, C. D., Silverman, R., Wu A. Y. (2002). An efficient k-means clustering algorithm: Anal-ysis and implementation. IEEE Trans. Pattern Analysis and Machine Intelligence. 24(7): 881-892.

[9] Y Fu, B Li, X Zhu, C Zhang. Active learning without knowing individual instance labels: a pairwise label homogeneity query approach. IEEE Trans Knowl Data Eng 26(4), 808-822 (2014).

[10] M. Chen, S. Mao, Y. Zhang, V.C.M. Leung. Big Data: Related Tech-nologies Challenges and Future Prospects, Springer, Cham, Heidelberg, New York, Dordrecht, London, 2014.

[11] EW Xiang, B Cao, DH Hu, Q Yang. Bridging domains using world-wide knowledge for transfer learning. IEEE Trans Knowl Data Engg. 22(6), 770-783 (2010).

[12] D Peteiro-Barral, B Guijarro-Berdias. A survey of methods for dis-tributed machine learning. Progress in Artificial Intelligence 2(1), 1-11 (2012).

[13] D Yu, L Deng. Deep learning and its applications to signal and information processing. IEEE Signal Proc Mag 28(1), 145-154 (2011).

[14] I Arel, DC Rose, TP Karnowski. Deep machine learning-a new frontier in artificial intelligence research. IEEE Comput Intell Mag 5(4), 13-18 (2010).

[15] D Che, M Safran, Z Peng, From big data to big data mining: challenges, issues, and opportunities, in Proceedings of the 18th International Conference on DASFAA(Wuhan, 2013), pp.1-15.

[16] https://www.users.cs.umn.edu/kumar/dmbook/ch6.pdf

[17] Stephen M., Inductive Logic Programming, New Generation Computing, 8 (1991) pp. 295-318.

[18] Ben-Gal I., Bayesian Networks, Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons (2007).

[19] http://web.stanford.edu/class/cs234/index.html

[20] Yoshua Bengio, Aaron Courville, and Pascal Vincent, Representation Learning: A Review and New Perspectives, International Conference on Learning Representations, 1993.

[21] http://researchers.lille.inria.fr/abellet/talks/metric_ learning_tutorial CIL.pdf.

[22] Goldberg, David E.; Holland, John H. (1988). Genetic algorithms and machine learning. Machine Learning.

[23] Michie, D. Spiegelhalter, D. J. Taylor, C. C. (1994). Machine Learning, Neural and Statistical Classification. Ellis Horwood.

[24] Bassel, George W.; Glaab, Enrico; Marquez, Julietta; Holdsworth, Michael J.; Bacardit, Jaume (2011-09-01). ”Functional Network Con- struction in Arabidopsis Using Rule-Based Machine Learning on Large- Scale Data Sets”. The Plant Cell. 23 (9): 3101-3116. ISSN 1532-298X.

[25] Abonyi, J., Szeifert, F. (2003). Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognition Letters, 24(14), 2195-2207.

[26] Boser, B. E., Guyon, I. M., et al. (1992). A training algorithm for optimal margin classifiers. In Fifth annual workshop on computational learning theory. Pittsburgh: ACM.

[27] Frohlich, H., Chapelle, O., et al. (2003). Feature selection for support vector ma- chines by means of genetic algorithms. In Proceedings of the15th IEEE interna- tional conference on tools with artificial intelligence, Sacramento, CA, USA, pp. 142-148.

[28] Joachims, T., Nedellec, C., et al. (1998). Text categorization with support vector machines: Learning with many relevant. In Proceedings of the 10th European conference on machine learning, pp. 137-142.

[29] Osuna, E., Freund, R., Girosi, F. (1997). Training support vector machines: Application to face detection. In Proceedings of computer vision and pattern recognition, Puerto Rico, pp. 130-136.

[30] Pawlak, Z. (1982). Rough sets. International Journal of Parallel Programming, 11(5), 341-356.

[31] Pawlak, Z. (1996). Why rough sets. In Proceedings of the IEEE international confe-rence on fuzzy system.

[32] Pawlak, Z. (1997). Rough set approach to knowledge-based decision support. European Journal of Operational Research, 99(1), 48-57.

[33] Urszula stanczyk, Rough Set-Based Analysis of Characteristic Features for ANN Classifier, Hybrid Artificial Intelligence Systems, pp.565-572, DOI: 10.1007/978-3-642-13769-3 69, June 2010.

[34] Geraldin B. Dela Cruz, Bobby D. Gerardo, Bartolome T. Tanguilig III, Agricultural Crops Classification Models Based on PCA GA Im- plementation in Data Mining, International Journal of Modeling and Optimization, Vol.4, No. 5, 2014.

[35] Junfei Qiu, Qihui Wu, Guoru Ding, Yuhua Xu and Shuo Feng, A survey of machine learning for big data processing EURASIP Journal on Advances in Signal Processing (2016) 2016:67.

[36] ChintanAmrit, TimPaauw, RobinAly, MihaLavric , Identifying child abuse through text mining and machine learning, Expert Systems with Applications Volume.88, 1 December 2017, Pages 402-418.

[37] Lvaro Brandn Hernndez, Mara S.Perez, SmratiGupta, VictorMunts- Mulero, Using machine learning to optimize parallelism in big data applications Using machine learning to optimize parallelism in big data applications, Future Generation Computer Systems Available online 17 July 2017.

[38] Lina Zhou, Shimei Pan, Jianwu Wang, Athanasios V. Vasilakos, Machine learning on big data: Opportunities and challenges, Neurocomput- ing,Volume 237, 10 May 2017, Pages 350-361.

[39] FilippoPiccinini, TamasBalassa, Abel-Szkalisity, CsabaMolnar, KrisztinaBuzas, Ad- vanced Cell Classifier: User-Friendly Machine Learning-Based Software for Disco-vering Phenotypes in High-Content Imaging Data, Volume 4, Issue 6, 28 June 2017, Pages 651-655.

[40] Boukaye BoubacarTraore, BernardKamsu-Foguem, FanaTangara, Data mining techniques on satellite images for discovery of risk areas, Expert Systems with Applications Volume 72, 15 April 2017, Pages 443-456.

[41] HuihuiChen, YusenZhang, IvanGutman, A kernel-based clustering method for gene selection with gene expression data, Journal of Biomedical Informatics, Volume 62, August 2016, Pages 12-20.

[42] Valds, J. J., Romero, E., & Barton, A. J. (2012). Data and knowledge visualization with virtual reality spaces, neural networks and rough sets: Application to cancer and geophysical prospecting data. Expert Systems with Applications, 39, 13193-13201.

[43] Akay, M., Welkowitz, W., and Kostis, J.B. (1994). ”Noninvasive detection of coronary artery disease using wavelet-based fuzzy neural networks”. IEEE Engineering in Medicine and Biology, 761-764.

[44] Alexopoulos, E., Dounias, G.D., and Vemmos, K. (1999). ”Medical diagnosis of stroke using inductive machine learning”. In Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course in Artificial Intelligence-ACAI99, Chania, Greece, 20-23.

[45] Dawant, B.M., Ozkan, M., Sprenkels, H., Aramata, H., Kawamura, K.,and Margolin, R.A. (1990). ”A neural network approach to magnetic resonance imaging tissue character-rization”. Communication, Control, and Signal Processing, Arikan E.,(ed.), 2, 1803-1809, Bilkent University, Ankara, Turkey. Elsevier, Amsterdam.

[46] Guo, Z., Durant, L.G., Lee, H.C., Allard, L., Grenier, M.C., and Stein, P.D. (1994). ”Artificial neural networks in computer-assisted classifi- cation of heart sounds in patients with porcine bioprosthetic valves”. Medical, Biological Engineering and Computing, 32, 311-316.

[47] Lane, V.P., Lane, D., and Littlejohns, P. (1996). ”Neural networks for decision making related to asthma diagnosis and other respiratory disorders”. In Proceedings of Healthcare Computing, Harrogate, UK, 85-3.

[48] Lim, C.P., Harrison, R.F., and Kennedy, R.L. (1997). ”Application of autonomous neural network systems to medical pattern classification tasks”. Artificial Intelligence in Medicine, 11, 215-239.

[49] Strausberg, J. and Person, M. (1999). ”A process model of diagnostic reasoning in medicine”. International Journal of Medical Informatics, 54, 9-3.

[50] C. V. Subbulakshmi and S. N. Deepa, Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle Swarm Optimization with Extreme Learning Machine Classifier, Hindawi Publishing Corporation, Scientific World Journal, Volume 2015, Article ID 418060, 12 pages, http://dx.doi.org/10.1155 /2015/418060.

[51] Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar (2012) Foundations of Machine Learning, The MIT Press ISBN: 9780262018258.

[52] M.R. Smith and T. Martinez (2011). ”Improving Classification Accuracy by Identifying and Removing Instances that Should Be Misclassified”. Proceedings of International Joint Conference on Neural Networks (IJCNN 2011). pp. 2690-2697.

[53] G.-B. Huang and L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing, vol. 71, no. 16-18, pp.3460-3468, (2008).

[54] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing, vol. 70, no. 1-3, pp. 489-501,2006.