ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A HYBRID APPROACH OF ACTIVE LEARNING USING SVM AND N-GRAM TECHNIQUE

Pravesh Kumar Dwivedi*, Mr. Anurag Jain, Mr. Sanjay Pal
* MTech. Scholar in Dept. of CSE, Radharaman Institute of Tech. And Science, Bhopal,India Dept. of CSE, Radharaman Institute of Tech. And Science, Bhopal,India

ABSTRACT
Active learning is an important subject in data mining and machine learning, which has been studied extensively and has a wide range of applications. Learning based on association rules, also called rule based learning, is a technique that uses association rules learning and taking proper decision. Heterogeneous DTC employs a novel data structure, association rule, to compactly store and efficiently retrieve a large number of rules for learning. We consider an active learning scenario in which the supervisor (trainer) can make decisions regarding the possibility to choose new examples for learning. In the classical forms of supervised learning, the training set is chosen according to some known or random given distribution. The supervisor is a passive agent in the sense that he is not able to interact with the training set in order to improve the performances of the learning process. In this paper hybrid approach that use N-gram with SVM has been proposed. Hybrid approach lead formal manner to learn "difficult learning" and "easy learning" related to the training data set.

KEYWORDS: Machine Learning, Active Learning, N-gram , SVM, Classification.

INTRODUCTION
Over the past 50 years, the study of machine learning[1] has passed the efforts of a handful of software engineers exploring whether computers can learn to play, and a field of statistical computing considerations largely ignored, a broad discipline that produced the major theories of statistical calculation of the learning process, designed algorithms commonly used in commercial systems for speech recognition, computer vision, and a variety of other learning tasks, and was divided in a data exploration industry to discover hidden patterns in growing volumes of online data

The successful implementation of a measurement of progress in machine learning is the major real-world applications, such as given below. Although many now take for granted these applications, it is interesting to note that even in 1985, there was almost no commercial applications of machine learning[2].

Currently available commercial systems for speech recognition all use machine learning in one fashion or another to train the system to recognize speech. The reason is simple: the speech recognition accuracy is greater if one trains the system, than if one attempts to program it by hand. In fact, many commercial speech recognition systems involve two distinct learning phases [3]: one before the software is shipped (training the general system in a speaker-independent fashion), and a second phase after the user purchases the software (to achieve greater accuracy by training in a speaker-dependent fashion).

Many current vision systems, facial recognition systems, systems that automatically classify microscopic images of cells, are developed using machine learning again because the resulting systems are more accurate than the programs handmade. An application on a large scale computer vision formed through learning of the machine use by US Mail to automatically classify letters with handwritten addresses. Over 85% of handwritten mail in the United States are automatically sorted using the analysis software Letter formed a very high accuracy by learning the machine on a very large data set.

A variety of government efforts to detect and track disease outbreaks through learning machine[4]. For example, the ROD project involves real-time collection of reports to emergency services through Western Pennsylvania

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

consumption, and use of learning software of the machine to learn the profile of the typical income that can detect scenarios symptoms and their geographical distribution. The current work is to add in a wide range of additional data, such as retail purchases over-the-counter medications to increase the flow of information in the system, which increases the need for automated learning methods given this set of more complex data.

Learning machine methods have been used successfully in a number of robotic systems. For example, several researchers have demonstrated that the use of machine learning to develop control strategies for aerobatic helicopters and stable flying helicopters. The recent contest sponsored Darpa involves autonomous driving robot for more than 100 kilometers in the desert was won by a robot that uses machine learning to refine their ability to detect distant objects (the same formation of coherent self-made data on land originally seen in the distance and you will see later near).

Many data intensive science now make use of learning methods of the machine to help in the process of scientific discovery. Learning the machine is used for learning gene expression in the cell from high-performance data models to discover unusual astronomical objects 2 from massive data collected by the Sloan Sky Survey, and characterize complex patterns of brain activation that indicate different cognitive states of people in fMRI scanners. Learning methods of the machine are transforming the practice of many intensive science of empirical data, and many of these sciences now have workshops on learning the machine as part of his lectures fields.

RELATED WORK

Machine learning [1] is an approach by which, it is possible to recognize the unknown samples with the help of learning from known samples. In current scenario the Artificial Neural Network, Support Vector Machine and Genetic Algorithm are the famous approached used in machine learning, but it also seems to be that they have some merits and demerits as well. In this paper, a novel machine double-layer learning strategy has been proposed by the author. This method integrates the advantages of all above three methods that is ANN/SVM and GA. ANN/ SVM is used to carry out inner layer learning in order to obtain model's inner parameters, and GA is used to implement outer layer learning so as to acquire model's outer parameters. Therefore the new learning method need carry out double layers learning, by comparison to common machine learning, the new method possesses stronger self-adaptive ability, and it can make up the shortcomings of single learning method and fully assure model's generalization ability. In the end, the machine double-layer learning method is applied for non-linear time series forecasting, and examples show the correctness and validity of the new method.

Edward Zimudzi [2]presents an implementation of educational model based on active learning to teach programming techniques on a computer training course that prepares undergraduate students to teach Botswana General Certificate of Studies of computer education. This point of the program is very important for the development of life skills in problem solving and critical thinking; skills that are crucial in computer science graduates career. The issue has always been very difficult to control for IT students from pre-service education, with little programming experience. We offer this active learning approach for the reason that students actively participate in the discussion and the course tutor can easily identify alternative conceptions that students and be able to provide the necessary assistance for Future computer studies teachers. Active learning is a method of constructivist teaching that actively engages students in the learning process. Students learn problem solving through action, a step by step process, always based on what they already know before. Use different techniques methodological interaction, which enhances the understanding of programming concepts to students and the general motivation to learn more. The author has also discussed the role of teachers in active learning approach.

Despite strong evidence of the positive impact of active learning strategies, STEM faculty demonstrate a spectrum of receptiveness to incorporating active learning into their classrooms, and for a variety of reasons, engineering classes continue to be dominated by a passive lecture style. This paper draws on data from a four-year study that investigated the use of five social instruction strategies, including active learning. Twenty-four STEM faculties at 4 institutions were interviewed regarding their understanding of and attitudes toward these strategies. This paper focuses on the results of the active learning component of these interviews. Faculty most often interpreted active learning as what students do and viewed self-motivation as a key component of what students think while active learning. These results, while drawn from a small sample population, can nevertheless make an important contribution to understanding why passive learning remains predominant in the STEM classroom. Cheryl Allendoerfer [3] examines how the findings from this study can inform efforts to promote changes in STEM education that would bring more active learning to the classroom.

In many applications, the use of all relevant data to extract information from multiple sources and to achieve greater accuracy in predicting is desirable. Cooperative learning is observed in human societies and some animals. Good knowledge and the acquisition of information, cooperation in learning multi-agent systems can lead to greater efficiency compared to individual learning. Cooperative learning in multi-agent systems is generally expected to improve the quality and speed of learning. According to the survey of the research is going to focus on high coordinated approach to agents. Multiple data sources can be considered as different points of view regarding the same problem of learning where the dependencies between the views could be taken into complex structures. This leads to problems with learning interesting and challenging machine where the data sources are combined with learning. Deepak Vidhate [4] proposed a framework that includes several data combination of work and related issues such as the transfer of learning, multi-task learning, Multiview learning, and learning under covariate shift. Cooperative learning is an approach in which one or more students team work together to achieve a better understanding of a specific task. The purpose of this article is to use this approach to describe a proposal for the design and construction of a (multi-Learning System) cooperative learning machine containing two or more students of the machine that cooperate together.

Active learning is motivated in many supervised learning tasks unlabeled data may be abundant but labeled examples are costly to obtain. The goal of active learning is to maximize the performance of a learning model using the least significant training data as possible, minimizing the cost of annotation data. So far, there is still very little work on active learning for regression. Wenbin Cai [5] proposed , a new active learning framework is proposed for maximizing planned called regression model change (GMCS), which aims to choose examples that lead to greater variation in the current model. The pattern of change is measured as the difference between the current model parameters and updated after training with all of the extended training parameters. Inspired by stochastic gradient descent (SGD) updated standard, an estimated change as the slope of the loss compared to an exemplary candidate for active learning. In this framework, we derive new active learning algorithms for both linear regression and nonlinear regression to select the most instructive examples. Outbreaks experimental results on benchmark datasets from the UCI machine learning repository has shown that the proposed algorithms are very effective in selecting the most informative and robust for different types of data distributions examples.

ACTIVE LEARNING

Active learning is a well-studied research area, addressing data mining scenarios where a learning algorithm can periodically select new examples to be labeled by a human annotator and add them to the training dataset to improve the learner's performance on new data. Its aim is to maximize the performance of the algorithm and minimize the human labeling effort.

Active learning (AL) [7] is a type of data sampling technique where, instead of selecting a subset of random examples to train a classifier, subsets of the most informative examples are selected. The great challenge here is on how to choose a criterion to select these examples. Baram [8] identifies three types of active learning methods according to their goals. The first, named uncertainty sampling, identifies the examples the learner is less certain about its predictions. The second, which is used in this work, and was named query-by-committee, uses a committee of learners and the disagreement between committee members as a measure of training examples in formativeness. The third selects examples which, when labeled, lead to the greatest reduction in error by minimizing prediction variance.

In active learning, the learning algorithm periodically asks an oracle (e.g., a human annotator) to manually label the examples which he finds most suitable for labeling. Using this approach and an appropriate query strategy, the number of examples that need to be manually labeled is largely decreased. Typically, the active learning algorithm first learns from an initially labeled collection of examples. Based on the initial model and the characteristics of the newly observed unlabeled examples, the algorithm selects new examples for manual labeling. After the labeling is

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

	i	want	to	eat	indian	food
i	5	827	0	9	0	0
want	2	0	608	1	6	6
to	2	0	4	686	2	0
eat	0	0	2	0	16	2
indian	1	0	0	0	0	82
food	2	0	15	0	1	4

finished, the model is updated and the process is repeated for the new incoming examples. This procedure is repeated until some threshold (for example, time limit, labeling quota or target performance) is reached or, in the case of data streams, it continues as long as the application is active and new examples are arriving.

N-GRAMS

N-Grams is an algorithm in order to predict the word with the help of probabilistic approach. In this method the prediction will take place after the observation of N-1 words for next word. So it seems to be that, the calculation of the probability of the next word is strongly connected to calculate the probability of the word sequence.

Simple (Unsmoothed) N-grams

Simple probabilistic model for predicting words can be assigned the same probability to each word. So suppose there are N words in a language, then the probability of a word with another word would be 1 / N. However, this approach ignores the fact that some words are more common than others in tongues.

Markov Assumption

It seems to be that the above idea that some words are more likely to follow the word in certain contexts. Would need to know all the words to the word you are trying to predict, but it would be inefficient to know the whole story, since we can find an infinite sequence of sentences and history, we would have never happened before. Therefore, we will bring the story to a few words.
Table 1:- Smoothing Using N-Gram

Smoothing

As shown in the above table 1, most of the table is filled with zeros. Because your body is limited, most of the sequences of words are assigned a zero probability even they must have a non-zero probability. The MLE estimation gives accurate results when the sequences are common in our training data, but does not give good results in zero sequences of probability and low frequency sequences. Another problem is the perplexity, a measure used for the evaluation of N-grams, does not work when there is zero probability of sequences in our data. Therefore, we will modify MLE to collect a certain probability mass from high-frequency sequences, and distribute it to zero frequency sequences. This change is called smoothing.

BACKOFF

Like discounting algorithms, backoff algorithms are used to solve the problem of zero Frequency N-grams. The intuition behind back off algorithms is look for a (N-1)-gram if there is no N-gram for the specific word sequence. For example, if we do not have an example of particular trigram wn-2 wn-1 wn to compute P(wn | wn-2 wn-1), then we can use the bigram probability P(wn | wn-1). Similarly, if we cannot compute P(wn | wn-1), we can look to the unigram P(wn). In backoff algorithms, if we backoff to lower order N-gram if we have zero count for the higher order N-gram.

SUPPORT VECTOR MACHINE

A support vector machine is a machine learning approach which is also known as support vector networks. This approach is a type of supervised learning approach. This approach analyses the data and also able to recognize the patterns.

SVM have been developed in reverse order to the development of neural networks (RNE). SVM evolved from the theory of implementation and experiences, while the NN followed by more heuristically, applications, and extensive experimentation to theory. Interestingly, the strong theoretical basis for SVM ignored much appreciated at first. Publication of the first work of Vapnik Chervonenkis and colleagues unnoticed until 1992. This was due to the widespread belief in the statistical community and / or machine learning, despite being theoretically attractive, SVM are neither appropriate nor relevant for practical applications. They were taken seriously when excellent results were obtained in the landmarks of practical learning. Today, the SVM shows better results than comparable results (or a) NN and other statistical models on the problems most popular reference.

PROPOSED METHODOLOGY

As the primary purpose of our research is to harvest active learning, our main effort is to develop a hybrid method that use N-gram technique for feature extraction and SVM for classification through Machine learning. Classifier modules in text mining are our main focus. However, we adopted two classifier modules to show the effectiveness of the proposed methods: N-Gram [3,4] and SVM [6]. There are two reasons: (1) we have developed software in our previous work and (2) the two modules are well-known and perform well. Note that the classifiers in the proposed approach are exchangeable. In practice, other classifier models can be inserted or replaced in our framework.

Proposed framework broadly have three main phases tagging, feature vector extraction and labeling.

Tagging

Tagging is use to estimate Likelihood estimation. Proposed methodology use basic idea of the N-Gram [3,4] method is to compute the probability that term t will show up after specific sequence of n terms t1, t2, ... , tn, i.e. P(ti|t1, ... , ti 1).Since there are limited terms in documents and extremely expensive in computing time and capacity, N-Gram often assumes to have Markov property, i.e. P(ti|t1, ... , ti 1) P(ti|ti (n-1), ... , ti-1). That means the probability of observing the i-th term ti in the context history of the continuous preceding i-1 terms can be roughly approximated by the probability of observing ti in the shortened context history of the continuous preceding n-1 words (n-th order Markov property).

Feature Vector

The Vector Space Model (VSM) is a well-known model to convert text terms into a vector of identifiers, i.e. indexed terms. Documents and queries are respectively represented by: dj =(t1,j, t2,j, tt,j) and q =(t1,q, t2,q, tt,q). Each item corresponds to a separate term ti,j. If a term ti,j occurs in the document dj, ti,j is non-zero. All the terms will first filter through a pre-defined dictionary after stemlized, i.e. filtered out stem words. If a term is not in the pre-defined dictionary, the weight for the term will be 0, i.e. ti,j =0. The The algorithm that computes feature weights is very important.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

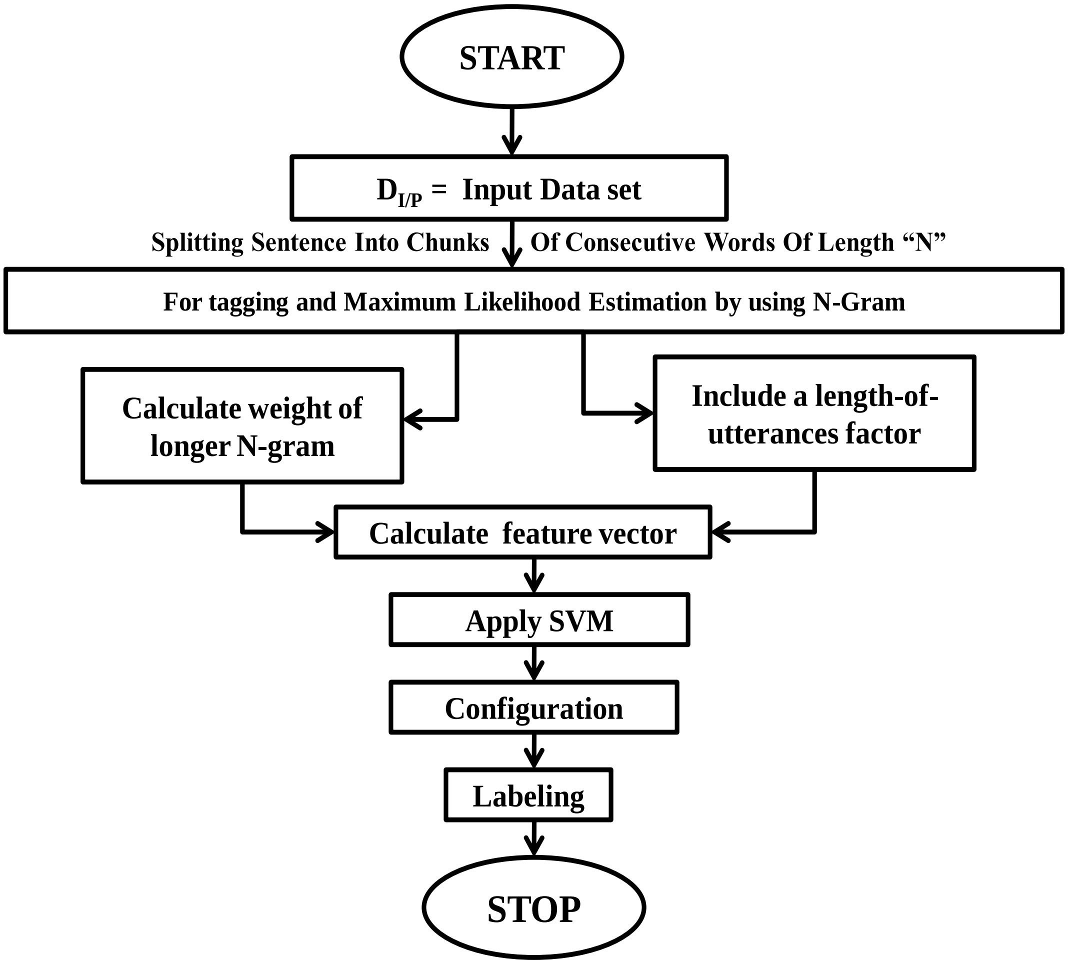

Figure 1:- Proposed Framework for Active Learning through Hybrid Approach

Support Vector Machine
Support vector machine is a supervised learning tool. Therefore, messages in documents will be tagged with classes. We have bipolar feature vector data, marked as {xi, yi}, i =1,2, ... ,n, yi {1, 1}, and xi are every comment in document D. Suppose there is a perfect hyper plane H: x + b = 0 which can correctly partition all the text comments into bipolar categories. There are two hyper planes H1 and H2 which are parallel to H:
W.x+ b= 1
W.x+ b= -1

RESULT ANALYSIS
Our hybrid model for active learning using N-gram and SVM simulated in the Matlab 10b and used wine data set for the justification of classification results. However Wine dataset contain three types of data C type1, C type2 and C type 3.

Performance of hybrid approach for active learning algorithm shows that when we used supervised active learning on wine data set then value of runtime is 2.91722 Sec and F-Measure is 82.8176% as show in table 2.

Table2:- Performance Evaluation of Supervised Active Learning and N-Gram Based SVM Active Learning

		N-Gram Based SVM		Supervised Active Learning	
Sup-port	**Conf-idence**	**F-Measure**	**Time**	**F-Measure**	**Time**
0.4	0.1	90.4812	2.80802	82.8176	2.91722
0.4	0.2	90.5208	2.69882	82.8584	2.68322
0.4	0.3	90.1639	2.40242	82.4914	2.54282
0.4	0.4	90.2036	2.48042	82.5322	2.51162
0.4	0.5	90.2432	2.32441	82.5729	2.19961
0.4	0.6	90.2829	2.16841	82.6137	2.19961
0.4	0.7	90.3226	2.44922	82.6545	2.49602
0.4	0.8	90.3622	2.21521	82.6953	2.24641
0.4	0.9	90.7985	3.12002	82.736	2.30881

Performance of hybrid approach algorithm using N-gram and SVM shows that when we used machine learning technique by using SVM and N-gram on wine data set then value of runtime is 2.80802 Sec and F-Measure is 90.4812%, which is showing that classification rate accuracy increased above 90%.

CONCLUSION

Active learning by using hybrid approach so improved rate of learning in comparison of supervised. In the process of supervised the calculation complexity are increases, the complexity of time are also increases. Our proposed algorithm test wine data set. In this data set the rate of F-measure is up to 92%.We also use another data set (abalone data set) and estimate some little bit difference of rate of leaning is up to 91%. In future we minimize the complexity of time and also increase the rate of learning using Meta heuristic function such as ant colony optimization, power of swarm (pos) and dendrites cell algorithm.

REFERENCE

[1] Guo Chen Rongtao Hou, “A New Machine Double-Layer Learning Method and Its Application in Non-Linear Time Series Forecasting”, Proceedings of the 2007 IEEE, pp 795-799.

[2] Edward Zimudzi, “Active Learning for Problem Solving In Programming in A Computer Studies Method Course”, Academic Research International, Vol 3, No. 2, September 2012 pp 284-292.

[3] Cheryl Allendoerfer, Mee Joo Kim, Rebecca Bates, “Awareness of and Receptiveness to Active Learning Strategies among STEM Faculty”, IEEE Oct 2012, pp 1-6.

[4] Deepak Vidhate, Dr. Parag Kulkarni, “Cooperative Machine Learning with Information Fusion for Dynamic Decision Making in Diagnostic Applications ”, Advances in Mobile Network, Communication and its Applications (MNCAPPS), IEEE Aug 2012, pp 70 – 74

[5] Wenbin Cai, Ya Zhang and Jun Zhou, “Maximizing Expected Model Change for Active Learning in Regression”, 2013 IEEE 13th International Conference on Data Mining, pp 51-60

[6] Mohri M, Rostamizadeh A, Talwalkar A, Foundations of Machine Learning, The MIT Press ISBN: 9780262018258; 2012.

[7] Lai WC, Goh K, Chang EY, On scalability of active learning for formulating query concepts. Workshop on Computer Vision Meets Databases (CVDB) in cooperation with ACM International Conference on Management of Data (SIGMOD); 2004. p. 11–18

[8] Baram, Y., El-Yaniv, R., and Luz, K. 2004. “Online Choice of Active Learning Algorithms”. Journal of Machine Learning Research, 5: 255-291.

[9] Liu, H., Lieberman, H., & Selker, T. “A model of textual affect sensing using real- world knowledge” (pp. 125–132). Proceedings of Intelligent User Interfaces (IUI).200

[10] Cardie, C., Farina, C., Bruce, T., & Wagner, E.“ Using natural language processing to improve eRulemaking”. In Proceedings of the 7th annual international conference on digital government research (pp. 177–178), San Diego.2006