International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 4 No. 2 Oct. 2013, pp. 264-273
© 2013 Innovative Space of Scientific Research Journals
http://www.issr-journals.org/ijias/

Machine Learning based Question Classification Methods in the Question Answering Systems

Farhad Soleimanian Gharehchopogh[1] and Yaghoub Lotfi[2]

[1]Department of Computer Engineering,
Hacettepe University,
Beytepe, Ankara, Turkey

[2]Department of Computer Engineering,
Science and Research Branch, Islamic Azad University,
West Azerbaijan, Iran

ABSTRACT: The Question Answering Systems (QASs) use method of information retrieval and Information extraction to retrieves documents that contain special answers to the question. One of the existence problems is finding the desired information from this very high variety. For this reason, it is necessary to find ways for organizing, classification and retrieving of information. Question classification plays an important role in providing a correct answer on QASs because giving a bunch of formulated questions to provide the correct answer from among the many documents will be highly effective. The aim of classification is selecting suitable label for questions based on the expected response. In this paper, we investigate the effect of automatically classifying questions on machine learning algorithms. In this paper, we will explain different types of algorithms and compare and evaluate them and next we will investigate the existence algorithms' weakness and advantage in question classification. As a result, in the past most classification was done based on sets of words that many studies show that to maximize the efficiency of the classification of algorithms we require semantics and in the questions we should looking for feature that be close to the meaning of questions. A great deal of research proposed to analysis and to classify emotions and to extract knowledge from them and to classify them using semantic and linguistic knowledge but it still requires a lot of research and development.

KEYWORDS: Support vector machine, classification, Question Answering Systems, machine learning, information retrieval.

1 INTRODUCTION

In the present era, that is called the information age; there is a lot of information available to users, and there is the vast expanse of the information on the World Wide Web (WWW) that to this vast volume add lots of users from all over the world every day. One of the existence problems is finding the desired information from this very high variety. For this reason, it is necessary to find ways for organizing, classification and retrieving of information. Another method is the use of systems for retrieving information which provide users with necessary documentation [1], [2]. Using a common information retrieval system, users are confronted with a large number of documents that the reading of all the documents to find the information is very time consuming and frustrating.

For example, when we present the question in search engines, information retrieval systems usually retrieve documents that contain the keywords, while the user desired to know the real answer to his question. For example, if you write a search engine "Who was the first Iranian woman to go into space?" The real answer of the user is "Ms. Anousheh Ansari," but retrieval system retrieves many documents that include the words "first", "Iran" and "space" [1], [3], [4]. Text retrieval conference (TREC), holding worldwide aims at comparing the information of retrieval systems by business groups and academic. Participating systems perform the same queries and retrieve relevant documents. Results are evaluated manually in a separate QAS and the assessment is carried in the so-called separate QAS. The Text Retrieval Conference has started with this purpose in [1], [4]. If search engines were able to receive the user ' question in the form of a question in natural language, understood with minimum redundancy and maximum precision, and respond to the massive volume of retrieved documents, we were not faced with the problems in the retrieval systems. In this regard question classification, which is categorized by putting questions in a sense, plays a key role in this system [3], [4].

Whatever questions be classified more accurately, QAS' comprehension of the question will be more understandable and thus achieving the correct answer will be easier. For example, the question "Who was the first Iranian woman to go into space?" If the system detects that the application is a name 1 –the search range is reduced due to this knowledge. 2 - Retrieved documents are also greatly reduced. Question Classification plays an important role in the performance of systems in most categories of QASs [5], [6].

This paper has been organized as follows: next section is introduction question classification concepts and its methods. Section 3 and section 4 discuss about question answering systems and machine learning and its techniques, respectively. Section 5 is a discussion and comparison of presented techniques such as decision tree, artificial neural network, support vector machine ant etc. In the section 6, we present the results of this research.

2 Question Classification

Nowadays, the problem of classification is one of the issues and many of them can be solved under this classification. One of the issues raised by the issue of classification is machine learning. Different ways has been proposed to solve classification problems in machine learning. In these methods, algorithm learns to predict expected response of users based on packing. Questions classification is a Boolean value (true or false) as ordered pairs $< q_j, c_i > \in Q \ \times C$, where Q is a set of questions and $C = \{c_1, c_2, \dots, c_{|c|}\}$ is a pre-defined set. If question $\mathrm{q_j}$ belongs to group $\mathrm{c_i}$ it is dedicated to the Boolean value true (True) $< \mathrm{q_j}, \mathrm{c_i} >$. Otherwise it is dedicated to false Boolean (False) [3].

In other words questions classification, placing those classified items in several semantic categories, regarding the possible answer [1] is in fact, the prediction of expected response according to the type of user s' response [7]. If the algorithm is able to correctly identify the type of responding meaning, it will be reach accurately to the correct answer question classification can be classified into two main groups; one based on rules and features, and the other based on machine learning [7].

2.1 Question Classification based on Rules

This kind of question classification induces semantic and syntactic rules write the rules manually and give them to algorithm based on expected language. For example, the questions start with terms like 'who' or 'whom' replaced in 'human' group. question classification apply based on these rules and regarding the rules that written manually, and regarding the syntax and features of language, but these process need cost, extra time and hardworking which shows systems based on this method are not free from error [8].

2.2 Question classification based machine learning

Since presenting all syntactic and semantic rules of a language to algorithm is a cumbersome task, for this reason different types of algorithms are made that can receive different examples and have the learning ability and can preview the user' expected response easily. Using machine learning, we can generate systems that includes thousands features of questions and do classification those questions automatically. This action increases the productivity rate of QAS that will be discussed in the following briefly [4], [7].

3 Question Answering Systems

These systems are the systems that let users to say their answers in natural language and reach to their expected answers with minimum redundancy. Understanding natural language for computers is overwhelming, for this reason, in this field studies were carried and will be carried that became the reason for the emergence of methods and inventions in this area. The heart of every QAS is its question classification algorithm. The higher the precision of the taxonomy, the higher the performance of the system will be. The machine learning algorithms are of this type that shows the best operation in the field of question taxonomy which we will review the main algorithms in the following section [3].

4 MACHINE LEARNING ALGORITHMS

Machine learning is one of the major branches of artificial intelligence research. The aim of the research in this field is access to learning techniques that allow the simulation of human intelligent as an intellectual behavior by computers [9]. The main aim in machine learning is the improving of machine that its ability to do that in the machine is constructed. This manner of correcting or improving called machine learning that is based on evaluation and testing and also based on the rate of correctly doing that act, for example the amount of winners in chess. Machine learning using information theory of the mathematical models can be used for getting results. Through program writing you can tell the machine what to do. Through several showed examples, we will get the machine to learn. Machine will be able to learn by experience environments, when the agent is not known for certain machine and examples of instruction are not available; it can be trained through observation [10], [11].

Algorithms have the ability to learn in three ways: 1 - supervised learning 2 – un-supervised learning 3 - semi-supervised learning. In this paper, we propose the supervised learning algorithms. Many machine learning algorithms based on the observation have been constructed so far that following we will point to the most important of these algorithms.

4.1 K-NEAREST NEIGHBOR

K- Nearest neighbor algorithm is a simple type of supervised learning algorithm. Its approach is in the case that when a new training data enters system, it recalls a group of training data previously categorized in the system and categorized them based on new training sample [12]. One drawback of this method is its high computational cost for categorizing each new training data when entering to the system. Another drawback is that the new model will be considered the same based on the characteristics of new training scheme that are assumed the same and are recovered the same.

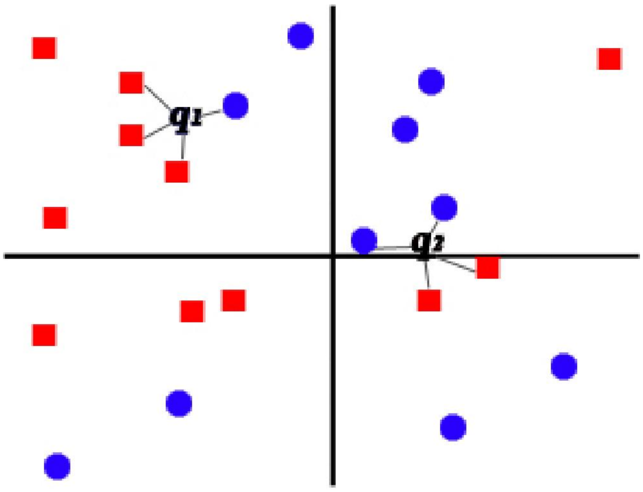

Fig. 1. K-Nearest Neighbor algorithm

In this algorithm, we assume all training data are n-dimensional space into Rn. Nearest neighbor distance between two points computed using the standard rules of the distance of two points [12], [13], [14].

$$d(x_i, y_i) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \quad (1)$$

In this algorithm calculations are different, given that space is considered to be two-dimensional or multi-dimensional training data. This algorithm will be categorized new training sample based on the prior training data. Different version of this algorithm is presented. One of the main problems of this method is that the distance of training data is considered the same based on their characteristics [15]. In the case that these extra and non- relevance characteristics be excessive, it becomes the main reason for algorithm' astray and as a result to algorithm' performance reduction. This algorithm is known as pattern recognition algorithm and it has many applications in this area. In this area, we have considered the assumption that the features of each training data should be similar [12], [15].

4.2 DECISION TREE

Decision tree reconstruct training data in the form of a tree using well-defined query is True / False. In a decision tree structures, leaves represent those of groups and the edges represent a set of features that these features leads to another special features in the groups. Decision trees are well organized in the case that with easily placing new data in roots and executing query model until reaching to special leave, we can classify new data that is the main aim of algorithm.

Decision tree is a good classification with several impressive advantages. The main advantage of decision trees is that it is simple and comprehensible for non-expert users. In addition, repetition of a mathematical algorithm can be easily explained and reviewed, and it can be revealed a comprehensive point of view containing useful information from the logic of the scheme. One of the major drawbacks of this decision tree method is that if training data have many features it wouldn't give good performances [16], [17]. As we know, when new data enter system decision tree algorithm with placing new data in the root restructure decision tree and tree replaces new decision with previous tree.

One of the major drawbacks of decision tree method is that maybe new tree is not better than the previous tree and it intensively reduces the performance of algorithm tree, and in addition to this issue, tree' structure will be very complex with increasing input [18]. One of the applications of recent decision tree is in the advertising of dynamic web pages. There is a variety of versions of decision tree algorithms such as ID3, ID4-hat, ID5, C4.5, and CART which the practical application of decision tree algorithms in the classification of answers will be addressed in the next section of [14].

4.3 SPARSE NETWORK OF WINNOWS (SNoW)

SNoW is a machine learning algorithm that is used for question classification according to the user expected response. This algorithm includes a number of features of the questions that are used in order to be classified as multi- class. The algorithm framework comprised of dispersed framework of linear algorithm containing one or some pre-defined feature space or feature spaces added step by step [8].

Table 1. Li & Roth's two-layer taxonomy for question classification

Coarse	Fine
ABBR	abbreviation, expansion
DESC	definition, description, manner, reason
ENTY	animal, body, color, creation, currency, disease/medical, event, food, instrument, language, letter, other, plant, product, religion, sport, substance, symbol, technique, term, vehicle, word
HUM	description, group, individual, title
LOC	city, country, mountain, other, state
NUM	code, count, date, distance, money, order, other, percent, period, speed, temperature, size, weight

Classification can be categorized into course and fine in the algorithm and it will be done in two steps. In the first step, the question will belong to one of the sixth coarse groups. Then with attention to coarse group, each coarse fine will belong to its sub-category [4], [8]. One advantage of this method is its simplicity and its low confusion at the moment of categorization. Input algorithm is a set of features that constructed semi-automatically. A part of features is automatically made with human intervention and with words meanings. The feature space was comprised of 200,000 features. Full description and results of the software and algorithm are available in [8].

4.4 NAIVE BAYES

Naïve Bayes (NB) classifier is a simple algorithm, which works based on probability and inference. We can make good decisions in this algorithm regarding the probability distribution and the shown data optimal decisions. Simple Bayes algorithm can be used to check the operation of the algorithm that have done previously and are likely to pay attention to algorithms which use probabilities. Bayes' theorem is based on Byes hypothesis provided that we are searching ways to find the probabilistic group for new training sample of input entering system. This is not possible without primary information. Bayesian model is the direct method for finding probabilities in the abstract space [12]. Given that the probabilities are precise, this algorithm operates with a minimal amount of training data to learn well and it will have the necessary parameters for a given rating category. It will calculate just the variables of each class rather than the entire set with taking into account the covariance matrix of the independent variables [12], [14].

Regardless of the simplicity of Bayesian algorithm, it can well perform in too complex categories and show too high performance in spite of what we expect. Software-based classification constructed based on simple Bayesian algorithms have shown an amazing performance in most cases. The main shortcoming of simple Bayesian algorithm is its relatively low efficiency compared with other so far made algorithms such as SVM algorithm which in this paper we will briefly take a look at it. The tests show that the performance of SVM algorithm is much better than that of the simple Bayesian algorithm. During the past years a lot of research has been done to solve performance problems and to increase its productivity. In some studies simple Bayesian algorithm has been combined with other classification algorithms in order to increase efficiency [12], [14]. The results were great with the combination of this algorithm with SVM algorithm [19], [20]. Of important applications of NB algorithms can be mentioned to filtering of spam in e-mails and categorizing of contents on web pages. Based on results, the performance of NB algorithm is high in these types of operations. Also setting NB algorithms on textual and numerical data is simple in comparison with other algorithms. But its performance is low in taxonomy texts, especially in texts close to natural language. One of the good advantages of NB algorithm is its short computational time for learning operates [14].

4.5 Support Vector Machine

A SVM is a subcategory of supervised machine learning algorithms used for classification data. In this way that machine learns to categorize input in pre-defined entries. This is a new technique and is shown a better performance than existing algorithms for classification of data such as neural networks. SVM works to the best possible for multi-dimensional data. Also it handles very well in classification questions automatically because it is non-linear and multi-dimensional [4].

Assume a set of training data (D), with n members with the assumption that those are linearly separable:

$$D = \{(x_i, c_i) | x_i \in R^p, c_i \in \{-1,1\}\}_{i\,=\,1}^{\,n} \qquad (2)$$

There are two training data set, classes Ci=1 and Ci=-1. The aim is to find the best hyper plan separation for these two training data sets. The equation is in the following case. Look at formula (3).

$$\text{w.x-b=0} \qquad (3)$$

Where w is the vector normal to the sheet, and b is the intercept for the hyper plane separation. For each training data xi if $w_i.\mathrm{x_i} - \mathrm{b} \geq 1$ then training given to the class ci=1 and if $w_i.\mathrm{x_i} - \mathrm{b} \leq -1$, then the data can be related to the class ci= -1. The distance of hyper plane from the source is $\frac{b}{\|w\|}$. In the general case, the aim is to find the minimum value of b and w in a manner that it can properly classify the training data and maximizes hyper plane layout and margin.in order to an optimal batch scheduling done, we use the variableξ_i. Then the equation (4) can be written.

$$c_i(w.x_i - b) \geq 1 - \xi_i, \qquad 1 \leq i \leq n \qquad (4)$$

In Classification of questions in optimal face the following relationship approaches to about the lowest level categories: $\min[\langle w.w\rangle + c\sum_i \xi_i]$ that in this relation the parameter c control for optimization that shows the number of accepted training errors. The main idea of this method is if the training samples be linearly separated, we will gain hyper plane with maximum margins that separates training samples in groups. If training samples are not linearly separated, we write them with more spaces and dimensions so that it can categorize training samples linearly in new space. The purpose in SVM is finding optimized hyper plane [4], [12]. SVM training classified sample into two groups of positive training samples and negative ones. The aim is to find the best separating hyper plane from these two samples. The closest training data to the hyper plane separator is called support vector [21]. The idea of SVM is to draw the two borders parallel to hyper plane separator and to apart two planes in a size that they press training data. The best hyper plane separator is those that have the further distance with training data. If the hyper plane is selected correctly, the classification will be applied in the least amount of error.

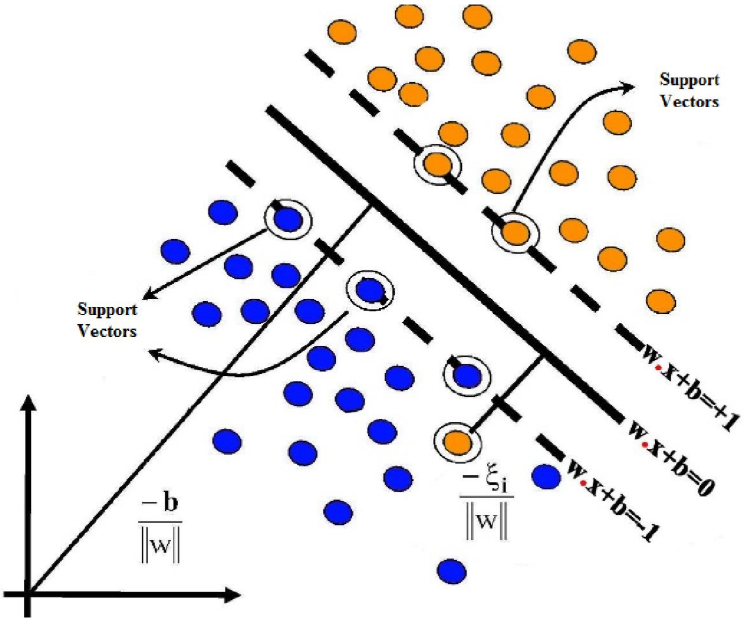

Fig. 2. SVM Algorithm

In SVM training is a relatively simple and efficient for multi-dimensional training examples and it shows a lot of features. Of its disadvantages are the complexity of the algorithm and its difficulty in understanding for non-expert users. The algorithm also handles the action with the high cost of computing at time and algorithms consume a lot of system memory too [14]. Among algorithms those supervised with SVM have shown excellent performance [22], [23]. In one conducted research, much has been done to evaluate the performance of the SVM algorithm using twenty training data set that was close to natural language texts using UCI. Also in order to achieve optimal SVM algorithm, much research has been done [24].

4.6 Artificial Neural Network

From general point of view, this system is composed of many interconnected units that each unit receives some input and computes output. Sometimes, an input is considered to be the output of a single unit. These units are known as neurons. Neurons can be used for storage in Artificial Neural Network (ANN). ANN methods are the comprehension of biological neural networks, but they cannot have many of the biological complexities [25]. According to conducted studies, biological network modeling is very complex and far from availability. But work on algorithms that simulates the behavior of biological networks for capturing better machine learning is more reasonable. A different version of the ANN for the classification of data is presented. But many researchers use single-layer Perceptron algorithm with an input layer and output layer since they are easy to use. In fact a perceptron is a network that neural networks made regarding their structure [12].

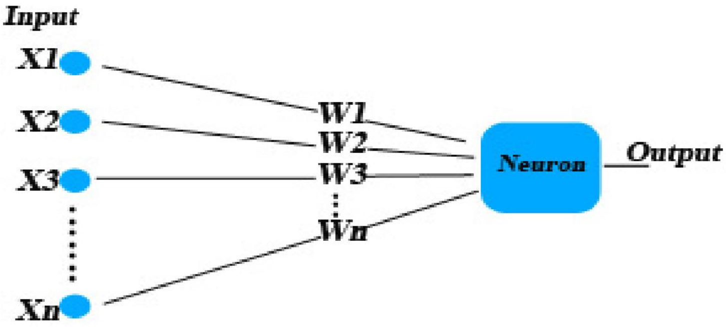

Fig. 3. ANN Algorithm

After receiving a real-valued vector, perceptron computes a linear combination of the set. If this amount was more than the determined amount, it is 1; otherwise outputs will be export the amount -1.

$$O(x_1, \dots, x_n) = \begin{cases} 1 & w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n > 0 \\ -1 & Otherwise \end{cases} \tag{5}$$

Perceptron implement multi-layer perceptron with a single-layer or with a hidden layer or with an output layer that are more complex than one-single perceptron layer. One of the main advantages of the algorithms implemented regarding ANN is that they show good performance due to their multidimensional nature and much communication among elements on classification of data with a lot of features [12]. Their major drawbacks are high costs, high CPU use and high physical memory. Also understanding ANN is a problem for many users.

5 Discussion and Comparison

Using a decision tree, where the leaves represent categories and the edges represent a set of features, we can consider each entry as a new training sample and place the expected questions into its category with putting it in the root of the tree and reconstruction of decision tree. The act of ordering questions will be done with above-mentioned procedures. In an evaluation [12], DT algorithm with algorithm' software C4.5 that is used in placing of this algorithm performs better than the NN and NB learning algorithm but it shows poor performance versus Snow and SVM algorithm [4]. A conducted reassessment shows that the performance of DT with the feature set 0.5 is also higher than Snow [3]. KNN algorithm is one of the simplest machines learning algorithms and it is used more in classification of documents. Different versions of this algorithm are presented that are interesting [14]. According to studies, in evaluating it with other learning algorithms regarding the questions categorization its performance was poor and it had little use. When it is compared with the algorithm like NB, Snow, SVM, it shows efficiency of lower than 67% with experiments on the same dataset [4], [5].

SNoW classification algorithms will be implemented in two stages, first the coarse group of questions is selected and then according to the characteristics of the questions, they will be awarded to the desired component and subcategory. This algorithm has a good accuracy and it shows better performance than the algorithm like NN, NB and DT, but less than SVM [5]. It can cover lots of questions' features due to being multi-class. NB algorithm works regarding the probability inference to the case that when a new question is entering the system, the question dedicated to the "most probable" groups according to the probable answer. In a paper, Mr. Lee and Zhang [4] deal with the study of the categories of questions with the same training and the same tests on a number of learning algorithms. They compared and paid attention to so-called algorithms' performance in the test that the results demonstrated algorithm NB show fairly good performance in comparison with algorithms like NN [5], [14]. Also studies have shown that the algorithm is tagged after training with 5500 questions and showed approximately 83.2% of performance in categorization of the total group [15].

SVM algorithm was one of the most popular and most used algorithms for question classification in recent years. The idea of SVM is in the case that with creating hyper plan between training examples with maximum margin and with the assumption that they are linearly separated, they can be classified. After Zhang and the other researchers' successful studies [4] that studied SVM and the other 4 algorithms, this algorithm showed good performance in comparison with the others. Also in a study conducted to compare the optimized version of the algorithm of KNN and SVM with an improved version of the algorithm NB, it concluded that SVM will be far better. But they achieved the significance and important result that the SVM algorithm will not alone work best, but if the pre-processing step be performed by KNN algorithm, classification by SVM will give better results [32]. The SVM algorithm is combined with the genetic algorithm that in this method the pre-processing procedure and selection of the items and features were conducted by genetic algorithms and after questions were classified by SVM algorithm which has shown satisfactory results [33]. Also in a similar research, the SVM algorithm is used for question classification in Chinese language that the algorithm provided dataset after training. After that the algorithm is tested and outcomes observed we can claim with dare that SVM has done well in this area and has shown good performances [34]. ANN algorithm is a simulation of the biological behavior of neural networks in a way that it formed of a number of connected units that connects the input group to the output and works in harmony. We used neurons for storage in the ANNs [11], [12]. In studies on spam sorting algorithms ANN show low efficiency than learning algorithm such as SVM [35]. In [36] ANN algorithm used to categorize online test questions according to their difficulty level. The results show that the algorithm efficiency is near to 78%.

The main Problem of instance-based machines learning is computational time and access to required features. Whatever these two categories improve the accuracy and efficiency of algorithm' classification increases [14]. Classification of items has many applications and covers an extensive range of issues which a great deal of research has been done but it seems that we're still at the beginning and presented algorithms have many difficulties and shortcomings and there is a need to be developed and debugged. For example, question taxonomy can be used in the area of e-government. as we observe that public services and government agencies daily are added and each organization has diverse sectors that each will also have their own laws and sometimes people are in trouble when they don't know to which part they must provide their questions. If each organization will have an online question and answer system and each user without aware of the various parts of the system can present their questions and the system intelligently present each question to its related sections, at that time we will find the true meaning of e-government. In the past most classification was done based on sets of words that many studies show that to maximize the efficiency of the classification of algorithms we require semantics and in the questions we should looking for feature that be close to the meaning of questions. If we be successful in this regard it will increase the efficiency of the algorithm. A great deal of research proposed to analysis and to classify emotions and to extract knowledge from them and to classify them using semantic and linguistic knowledge but it still requires a lot of research and development.

6 CONCLUSION

Literature demonstrates that the algorithms of the Support Vector Machine (SVM) have positive effect in a text when it is compared with other algorithms. The main Problem of instance -based machines learning is computational time and access to required features. Whatever these two categories improve the accuracy and efficiency of algorithm' classification increases. Classification of items has many applications and covers an extensive range of issues which a great deal of research has been done but it seems that we're still at the beginning and presented algorithms have many difficulties and shortcomings and there is a need to be developed and debugged. For example, question taxonomy can be used in the area of e-government. as we observe that Public services and government agencies daily are added and each organization has diverse sectors that each will also have their own laws and sometimes people are in trouble when they don't know to which part they must provide their questions. If each organization will have an online question and answer system and each user without aware of the various parts of the system can present their questions and the system intelligently present each question to its related sections, at that time we will find the true meaning of e-government. In the past most classification was done based on sets of words that many studies show that to maximize the efficiency of the classification of algorithms we require semantics and in the questions we should looking for feature that be close to the meaning of questions. If we be successful in this regard it will increase the efficiency of the algorithm. A great deal of research proposed to analysis and to classify emotions and to extract knowledge from them and to classify them using semantic and linguistic knowledge but it still requires a lot of research and development.

REFERENCES

[1] Xu-Dong Lin, Hong Peng, Bo Liu, "Support Vector Machines for Text Categorization in Chinese Question Classification," College of Computer Science and Engineering*, South China University of Technology, International Conference on Web Intelligence (WI 2006 Main Conference Proceedings),* IEEE, 2006.

[2] Marcin Skowron, Kenji Araki, "Evaluation of the New Feature Types for Question Classification with Support Vector Machines," Graduate School of Information Science and Technology Hokkaido University, Sapporo, 060-8628, Japan, *International Symposium on Communication and Information Technology (ISCIT)*, 2004.

[3] Hakan Sundblad, *Question Classification in Question Answering Systems*, Thesis No. 1320 ISSN 0280-7971, Department of Computer and Information Science Linkopings University, Linkoping, 2007.

[4] Dell Zhang, Wee Sun Lee, *Question Classification using Support Vector Machines*, National University of Singapore, Singapore-MIT Alliance, Toronto, Canada, 28-August 1, 2003.

[5] Ali Harb, Michel Beigbeder, Jean-Jacques, *Evaluation of Question Classification Systems Using Differing Features*, Institute of Electrical and Electronics Engineers, 2009.

[6] Wenting Tan, Jianrong Cao, Hongyan Li, "Algorithm of Shot Detection based on SVM with Modified Kernel Function," Shan Dong Jianzhu University, Jinan 250101, China, *International Conference on Artificial Intelligence and Computational Intelligence, IEEE*, 2009.

[7] Min-Yuh Day, Chorng-Shyong Ong, *Question Classification in English-Chinese Cross-Language Question Answering: An Integrated Genetic Algorithm and Machine Learning Approach*, Institute of Information Science, Academia Sinica, Taiwan, Department of Information Management, National Taiwan University, Taiwan, IEEE, 2007.

[8] X. Li and D. Roth, "Learning question classifiers," *In Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002)*, pages 556–562.

[9] Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization," *ACM Computing Surveys*, Vol. 34, No. 1, March 2002, pp. 1–47.

[10] Kiri L. Wagstaff, "Machine Learning that Matters," *In Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK, 012,* California Institute of Technology, 2012.

[11] Alex Smola and S.V.N. Vishwanathan, *Introduction to Machine Learning*, Press Syndicate of the University of Cambridge, 252 page, 2010.

[12] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, Khairullah khan, "A Review of Machine Learning Algorithms for Text-Documents Classification," *Journal of Advances in Information Technology*, Vol. 1, No. 1, February 2010.

[13] Bang, S. L., Yang, J. D., & Yang, H. J., "Hierarchical document categorization with k-NN and concept-based thesauri," *Information Processing and Management*, pp. 397–406, 2006.

[14] S. B. Kotsiantis, I. D. Zaharakis, P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Springer Science and Business Media B.V.,* 2007.

[15] P´adraig Cunningham and Sarah Jane Delany, *k-Nearest Neighbour Classifiers*, Technical Report UCD-CSI-2007-4 March 27, 2007.

[16] Shweta C. Dharmadhikari, Maya Ingle, Parag Kulkarni, "Empirical Studies on Machine Learning Based Text Classification Algorithms," *Advanced Computing: An International Journal (ACIJ),* Vol. 2, No. 6, November 2011.

[17] Kim, J., Lee, B., Shaw, M., Chang, H., Nelson, W, "Application of Decision-Tree Induction Techniques to Personalized Advertisements on Internet Storefronts", *International Journal of Electronic Commerce* 5(3) pp. 45-62, 2001.

[18] Russell Greiner, Jonathan Schaffer, *AIxploratorium – Decision Trees*, Department of Computing Science, University of Alberta, Edmonton, ABT6G2H1, Canada, 2001.

[19] Dino Isa, Lam Hong lee, V. P Kallimani, R. RajKumar, "Text Documents Preprocessing with the Bahes Formula for Classification using the Support vector machine," *IEEE, Traction of Knowledge and Data Engineering*, Vol. 20, No. 9 pp. 1264-1272, 2008.

[20] Dino Isa, V. P Kallimani Lam Hong lee, "Using Self Organizing Map for Clustering of Text Documents", *Elsevier, Expert System with Applications,* 2008.

[21] Muhammad Arifur Rahman, Vitalie Scurtu, "Performance Maximization for Question Classification by Subset Tree Kernel using Support Vector Machines," University of Trento, Trento, Italy, University, Computer and Information Technology (ICCIT), IEEE, 2008.

[22] Shi-jin Wang, Avin Mathew, Yan Chen, Li-feng Xi, Lin Ma, Jay Lee, "Empirical analysis of support vector machine ensemble classifiers," *Expert Systems with Applications*, pp. 6466–6476, 2009.

[23] Chung-Hong Lee a, Hsin-Chang Yang, "Construction of supervised and unsupervised learning systems for multilingual text categorization," *Expert Systems with Applications*, pp. 2400–410, 2009.

[24] Zi-Qiang Wang, Xia Sun, De-Xian Zhang, Xin Li,An "Optimal Svm-Based Text Classification Algorithm," *Fifth International Conference on Machine Learning and Cybernetics*, Dalian, 2006.

[25] Guoqiang Peter Zhang, "Neural Networks for Classification: A Survey, IEEE Transactions on Systems, Man, and Cybernetics—Part C," *Applications and Reviews*, Vol. 30, No. 4, 2000. IEEE.

[26] Trappey, A. J. C., Hsu, F.-C., Trappey, C. V., & Lin, C.-I., "Development of a patent document classification and search platform using a back-propagation network", *Expert Systems with Applications*, pp. 755–765, 2006 .

[27] Silvia Quarteroni, Alessandro Moschitti, Suresh Manandhar, "Advanced Structural Representations for Question Classification and Answer Re-ranking," t*he University of York, York YO10 5DD, United Kingdom, and Springer-Verlag Berlin Heidelberg,* 2007.

[28] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*, Cambridge University Press, Cambridge, UK, 2000.

[29] Y. Yang and X. Liu, "A Re-examination of Text Categorization Methods," In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99*), pp. 42-49, 1999.

[30] Rishika Yadav, Megha Mishra, SSCET Bhilai, "Question Classification Using Naïve Bayes Machine Learning Approach," *International Journal of Engineering and Innovative Technology (IJEIT),* Volume 2, Issue 8, February 2013.

[31] Fabrice Colas and Pavel Brazdil, "Comparison of SVM and Some Older Classification algorithms in Text Classification Tasks", *IFIP International Federation for Information Processing*, Springer Boston Volume 217, Artificial Intelligence in Theory and Practice, pp. 169-178, 2006.

[32] Min-Yuh Day, Chorng-Shyong Ong, and Wen-Lian Hsu, "Question Classification in English-Chinese Cross-Language Question Answering: An Integrated Genetic Algorithm and Machine Learning Approach," 1-4244-1500-4/07, 2007.

[33] Xu-Dong Lin, Hong Peng, Bo Liu, "Support Vector Machines for Text Categorization in Chinese Question Classification," *Proceedings Of The 2006 Ieee/Wic/Acm International Conference On Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06)0-7695-2747-7/06*, 2006.

[34] Bo Yu a, Zong-ben Xu b, "A comparative study for content-based dynamic spam classification using four machine learning algorithms", *Elsevier, Knowledge-Based Systems* 21, pp. 355–362, 2008.

[35] Ting Fei, Wei Jyh Heng, Kim Chuan Toh, Tian Qi, “Question Classification for E-learning by Artificial Neural Network”, *Institute for Infocomm Research National University of Singapore*, 2003.

[36] Arun D Panicker, Athira U, S. Venkitakrishnan, “Question Classification using Machine Learning Approaches,” *International Journal of Computer Applications* (0975 – 888) Volume 48, No. 13, June 2012.