AUTOMATIC CLASSIFICATION OF QUESTIONS INTO BLOOM'S COGNITIVE LEVELS USING SUPPORT VECTOR MACHINES

Anwar Ali Yahya*, Addin Osman*

*Faculty of Computer Science and Information Systems, Najran University, Najran, Kingdom of Saudi Arabia
aaesmail@nu.edu.sa, aomaddin@nu.edu.sa

ABSTRACT

In recent years, E-learning has increasingly become a promising technology in educational institutions. Among numerous components of E-learning systems, question bank is a primordial component. Question bank is a repository of questions that assists students and instructors in the educational process. In question bank, questions are annotated, stored and retrieved based on predefined criteria such as bloom's cognitive levels. Definitely, for question bank management, the automatic classification of questions according to Bloom's cognitive levels is of particular benefit. This paper explores the effectiveness of support vector machines (SVMs), in tackling the problem of question classification into Bloom's cognitive levels. To do so, a dataset of pre-classified questions has been collected. Each question is processed through removal of punctuations and stop words, tokenization, stemming, term weighting and length normalization. SVMs classifiers, namely linear kernel, have been built and evaluated on approximately 70% and 30% of the dataset respectively, using SVM-Light software package. The obtained preliminary results show a satisfactory effectiveness of SVMs with respect to classification accuracy and precision. However, due to the small size of the current dataset, the results of the classifiers' recall and F-measure suggest a need for further experiments with larger dataset to obtain conclusive results.

Keywords: *E-learning, Question bank, Text classification, Bloom's taxonomy, Machine learning.*

1. INTRODUCTION

It is undeniable fact that the advent of computer and internet technology has dramatically influenced the educational systems in many ways and the continuous growth of E-learning systems in educational institutions is unequivocal evidence. A primordial component of E-learning system is question bank, in which questions are stored in a database so as to be retrieved for a test or practice by users. Usually, questions are periodically collected at each test or exam time from year to year and stored based on predefined criteria such as difficulty levels, area of the curriculum or type of skill being tested (bloom's cognitive levels [21]), etc. Question bank can be then used for designing a more effective assessment by allowing a unique subset of questions to be chosen for each test or student where specific or personalized skills and levels of competence need to be examined. Therefore, question bank requires the best management e.g., organization, classification and retrieval for full utilization by users. Usually, the classification of questions is done manually, which is not only time consuming but also tedious and prone to mistake. To avoid these difficulties, systematic and automatic methods to manage question bank are needed [15].

Among the criteria that are used to categorize questions in question bank, Bloom's Cognitive Levels (BCLs) is one of the most important [19]. Generally speaking, in the field of education, Bloom's taxonomy is an essential concept that guides educators in writing learning objectives, preparing the curriculum, and creating assessment. In his effort to classify the thinking behaviors, Benjamin Bloom [1], identified three domains: cognitive (mental skills), affective (growth in feelings or emotional areas) and psychomotor (manual or physical skills). For the sake of this paper, only the cognitive domain is presented. The cognitive domain [2] involves knowledge and the development of intellectual skills. There are six major classes, which are listed in order below, starting from the simplest behavior to the most complex.

- **Knowledge**: Recall data or information or specific items, remember definition of some terms.
- **Comprehension**: Recall but do a little more (e.g. paraphrase, define and discuss to some extend), understand the meaning, translation, interpolation, and interpretation of instructions and problems.
- **Application**: Do all of the above, but can take information of an abstract nature and use in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.
- **Analysis**: Break down a communication into its constituent parts, revealing the relationships among them. Separates material or concepts into component parts so that its organizational structure may be understood.
- **Synthesis**: Pull together many disorganized elements or parts so as to form a whole. Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.
- **Evaluation**: Makes judgments about the value of material or methods. Make judgments about the value of ideas or materials.

Obviously, the task of automatic classification of questions into BCLs can be casted as text classification problem. In the information systems field, text classification is the automated assignment of natural language texts to predefined categories based on their content [17]. It is also viewed as instance of text mining, a subfield of data mining that try to extract probably useful information by analyzing large quantities of text and detecting usage patterns. Since its first appearance, dated back to the early 60's, it has been used in a good number of applications either explicitly as the main technology or implicitly, as a supportive technology, in the context of other applications [17]. In the field of E-learning, text classification has been used in a number of applications such as [10, 11, 13, 19].

This paper introduces a new application of text classification techniques in the field of E-learning. More specifically, the paper proposes a use of SVMs to tackle the problem of automatic question classification into different BCLs. Despite a wide variety of text classification techniques which have been developed, SVMs has been selected due to its proven superiority over others [14].

The remainder of this paper is organized as follows. Section 2 introduces the related works. Section 3 presents SVMs and describes how it can be used for BCLs question classification. Section 4 presents the obtained results and discusses them. Section 5 is devoted for conclusion.

2. RELATED WORKS

As mentioned above, the aim of this paper is to harness the remarkable performance of SVMs in text classification to tackle the problem of the problem of question classification into different BCLs. Therefore, this section sheds light on text classification techniques in general and their particular applications in question classification. Text classification enjoys quite a rich literature. As reported in [17], the works on text classification can be characterized by two stages. In the ’80s, the most popular approach of creation of automatic text classifiers is knowledge engineering techniques. Typically a set of manually defined logical rules, one per class, of type, *if DNF formula i then hcategory i*, where a DNF (disjunctive normal form) formula is a disjunction of conjunctive clauses. The text is classified under class *i* iff it satisfies the formula, that is, iff it satisfies at least one of the clauses. The drawback of this approach is the knowledge acquisition bottleneck well known from the expert systems literature.

The second stage of text classification works starts in the early ’90s, where machine learning approach has gained popularity and has eventually become the dominant one, at least in the research community. In this approach, a general inductive process (also called the learner) automatically builds a classifier for a given class c_i by observing the characteristics of a set of text manually classified under c_i or $\overline{c_i}$ by a domain expert; from these characteristics, the inductive process gleans the characteristics that a new unseen text should have in order to be classified under c_i.

As it has been pointed out above, since its first appearance, dated back to the early 60's, it has been used in a good number of applications, of which the most worth mentioning are controlled vocabulary indexing, routing and packaging of news and other text streams, content filtering, word sense disambiguation, hierarchical categorization of web pages. Besides that it has been applied implicitly, as a supportive technology, in the context of other applications such speech categorization by means of a combination of speech recognition and text classification, multimedia document categorization through the analysis of textual captions, author identification for literary texts of unknown or disputed authorship, language identification for texts of unknown language, automated identification of text genre, and automated essay grading [17].

In the context of the automatic classification of questions, the use of text classification techniques has been reported in several works. In one of these works [19], artificial neural network is proposed to question classification, in which back-propagation neural network is used as text classifier to classify question into three difficult levels that is easy, medium, and hard. In this work a five dimension feature vector is used as input to the back-propagation neural network. This five-dimension feature vector consists of query text relevance, mean term frequency, length of question and answer, term frequency distribution, and distribution of question and answer in text.

Another work on question classification using text classification techniques is reported in [3], which focuses on a specific type of questions, called open-ended questions. Questions of this type can be broken down into classes that identify the format and content of the expected response. In this work, SVMs are used successfully in the classification of open-ended questions.

An interesting work on question classification is presented in [15]. In this work, an adaptable learning assistant tool for managing question bank is presented. The tool is not only able to automatically assist educational users to classify the question items into predefined classes by their contents but also to correctly retrieve the items by specifying the class and/or the difficulty levels. The system adapts the categorization learning model to improve the system’s classification performance using the incoming questions. The system is tested and evaluated in terms of both system accuracy and user satisfaction. The evaluation result shows that the system accuracy is acceptable and satisfies the need of the users.

3. SVMs for BCLs QUESTION CLASSIFICATION

SVM is an emerging machine learning approach that has attracted much attention as a more viable alternative to other more mature approaches such as neural networks and nearest neighbor algorithms [8]. In the late seventies, the approach was introduced as statistical learning techniques as a result of research conducted by Vladimir Vapnik and was furthered in the 1990s by Vapnik and others at AT&T Bell Laboratories [20]. The SVM is a supervised machine learning algorithm that is trained to separate between two sets of data using training examples of both sets. In its simplest form, when used as a binary classifier, the training of SVM will construct a hyperplane, which acts as the decision surface between the two sets of data. This is achieved by maximizing the margin of separation between the hyperplane and those points nearest to it [4], as shown in Figure 1. Once the training is complete, new data can be classified by determining where it lies in relation to the hyperplane.

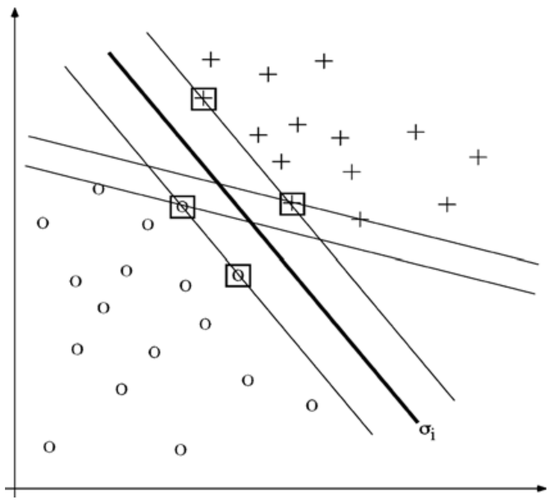

Fig. 1: Binary SVMs classifier

SVM approach was introduced to text classification by [9] and subsequently used by many researchers [5, 6, 12, 18, 22]. In these researches, several authors have shown that SVMs provide a fast and effective means for learning text classifiers from examples. Joachims [9] argued that SVMs are very suited for text classification due to the inherent characteristics of text such as high dimensional input space, few irrelevant features, the sparseness of documents vectors, and most text classification problem are linearly separable. With their ability to generalize well in high dimensional feature spaces, SVMs eliminate the need for feature selection making the application of text classification considerably easier. Another advantage of SVMs over the conventional methods is their robustness. Furthermore, SVMs do not require any parameter tuning, since they can find good parameter setting automatically. All this makes SVMs a very promising and easy to use method for learning text classifier from example.

The use of SVMs to design a text classification system requires three main steps: text representation, classifiers building, and the classifiers evaluation. In the text representation step, a text representation method is used to map a text into a compact representation of its content that is suitable for the subsequent steps. In the classifier building step, the SVM classifiers are automatically built for each class c_i by observing the characteristics of a set of texts manually classified under c_i or $\overline{c_i}$ by a domain expert.

In the classifiers evaluation step, the SVM classifiers are evaluated by gleaning the characteristics that a new unseen text should have in order to be classified under c_i. In the following subsection a detailed description of how SVMs are used to tackle the problem of question classification into BCLs is given.

3.1 QUESTION REPRESENTATION

As it has been pointed out above, texts cannot be directly interpreted by SVMs, therefore, a conversion procedure to map a text of a question q_j into a compact representation of its content needs to be uniformly applied. In this work, the adopted method to represent a question q_j is a vector of term weights $<w_{1j}, \ldots, w_{Tj}>$, where T is the set of terms (sometimes called features) that occur at least once in at least one question , and $0 \le w_{kj} \le 1$ represents, how much term t_k contributes to the semantics of question q_j. The term weight can be a binary weights (1 denoting presence and 0 absence of the term in the question); or non-binary depending on the classifier building algorithm used. For SVMs, non-binary weights are used. More precisely, the standard *tfidf* function is used, which defined as

$$tfidf(t_k, q_j) = \#(t_k, q_j).\log\frac{|Tr|}{\#Tr(t_k)} \tag{1}$$

where $\#(t_k, q_j)$ denotes the number of times t_k occurs in q_j , and $\#Tr(t_k)$ denotes the question frequency of term t_k, that is, the number of questions in which t_k occurs.

In order to apply the above representation a preprocessing of question should be applied which includes:

- Reducing of the question text to lower case characters.
- Punctuation removal: all types of punctuations are removed from the question.
- Stop word removal: any occurrence of words from the stop word list of the SMART system, found at ftp://ftp.cs.cornell.edu/pub/smart/english.stop, is removed.
- Tokenization: a token is a maximal sequence of nonblank characters. In this process, tokens consisting purely of digits are discarded.
- Stemming: The tokens were stemmed with Porter stemmer [16].

After the preprocessing of a question text, term weighting is computed as in Eq. 1, and length normalization is applied as follows

$$w''_{q_j}(t_k) = \frac{w_q(t_k)}{\sqrt{\sum_l w_{q_j}(t_l) \times w_{q_j}(t_l)}} \quad (2)$$

The vectors with the new term weights are used as input to the subsequent steps.

3.2 SVM CLASSIFIER BUILDING

As it has been pointed out above, building SVM algorithms find a linear decision surface (hyperplane) with maximum margin between it and the positive and the negative training examples for a class [9]. In this step, a single SVM classifier has been trained for each class using part of the dataset called training set. This step can be accomplished using one of the currently available SVMs tools. In this work SVM-Light [9] package, version 6.02 has been used. It is freely available from http://svmlight.joachims.org/.

It should be mentioned that SVMs using nonlinear kernel functions can be used, but have not shown a significant advantage in past text categorization studies [14], therefore it has not been investigated here.

3.3 SVM CLASSIFIER EVALUATION

The effectiveness of SVMs on a single class of BCLs can be evaluated using several measures [14, 17]. The computation of these measures depends essentially on a contingency table obtained from the classification of the testing set of each class. The contingency table consists mainly of the following values

- A : the number of documents a system correctly assigns to the category (true positives),
- B : the number of documents a system incorrectly assigns to the category (false positives),
- C : the number of documents that belong to the category but which the system does not assign to the category (false negatives)
- D : the number of documents a system correctly does not assign to the category (true negative),

The following are the common measures used to evaluate the effectiveness of SVMs classifiers.

- **Precision**: it is the probability that if a random document dx is classified under c_i , this decision is correct. It can be viewed as the "degree of soundness" of the classifier with respect to the class. That is

$$P = \frac{A}{A+B} \quad (3)$$

- **Recall**: it is the probability that if a random question ought to be classified under c_i , this decision is taken. It can be viewed as the degree of completeness of the classifier with respect to the class. That is

$$R = \frac{A}{A+C} \quad (4)$$

- **F_β measure**: it is the harmonic mean of recall and precision which is defined, for β=1.0, as follows

$$F_{1.0} = \frac{2RP}{R+P} \quad (5)$$

- Accuracy: the accuracy of a classifier is defined as follows

$$Acc = \frac{A+C}{A+B+C+D} \quad (6)$$

In addition to the above measures for a classifier of a single class, the effectiveness across a set of classes can be measured by the macroaverage (unweighted mean of effectiveness across all classes) and the microaverage (effectiveness computed from the sum of per-class contingency tables) [14].

4. RESULTS AND DISCUSSIONS

The currently used dataset has been collected from a number of Web sites on Bloom's Taxonomy literature. The collected dataset have been processed as described in section 3.1 and divided into training set (≈70% of the dataset) and testing set (≈30% of the dataset). Table 1 presents the statistics of the dataset. As it is shown in the table the size of dataset is 272 and the size of the training set and testing set is 190 and 82 respectively.

Table 1: Dataset statistics

Bloom's Cognitive Level	training	Testing
Knowledge	17	11
Comprehension	32	12
Application	28	13
Analysis	32	16
Synthesis	44	14
Evaluation	37	16
Total	190	82

Table 2 shows samples of questions and their corresponding BCLs class.

In the building step of SVM with SVM-light, all parameters were left at default values. This meant, in particular, that a linear kernel has been used (by leaving -t unspecified), equal weighting of all examples whether positive or negative (by leaving -j unspecified), and set the tradeoff C between training error and margin to the reciprocal of the average Euclidean norm of training examples (by leaving -c unspecified). Since cosinenormalized training examples were used, leaving -c unspecified meant C was set approximately to 1.0.

Table 2: Dataset question examples

Bloom's Cognitive Level	Question example
Knowledge	Identify the standard components of a computer
Comprehension	Explain what a poem means.
Application	Compute the area of actual circles
Analysis	Compare this book to the last book you read
Synthesis	Construct a device that would assist an athlete in their training.
Evaluation	Evaluate a work of art, giving the reasons for your evaluation.

The obtained results are summarized in Table 3, Table 4 and Table 5. Table 3 contains the results of the contingency table for each BCLs classes. It also presents the contingency table results over all classes of the SVM classification for each BCLs class and the total over all BCLs. These results obtained from the SVM classification of the testing set.

Table 3: Contingency Table for BCLs classes

Bloom's Cognitive Level	A	B	C	D
Knowledge	1	0	10	71
Comprehension	3	3	9	67
Application	1	0	12	69
Analysis	3	1	13	65
Synthesis	9	1	5	67
Evaluation	8	0	8	66
Total	25	5	57	405

Table 4 contains the reported results of the classification effectiveness measures (Accuracy, Precision, Recall and F-measure) for each BCLs class. These results are computed using the results of contingency table as described in section 3.3.

Table 4: Classification effectiveness for each BCLs class

Bloom's Cognitive Level	Accuracy	Precision	Recall	$F_{1.0}$
Knowledge	87.80	100.00	9.09	16.67
Comprehension	85.37	50.00	25.00	33.33
Application	85.37	100.00	7.69	14.28
Analysis	82.93	75.00	18.75	30
Synthesis	92.68	90.00	64.29	75
Evaluation	90.24	100.00	50.00	66.67

Additionally Table 5 presents the results of the Macro-average and Micro-average values computed for measure.

Table 5: Macro(Micro)-average over all BCLs classes

	Accuracy	Precision	Recall	$F_{1.0}$
Macro-average	87.4	85.83	29.14	39.33
Micro-average	87.4	83.33	30.49	44.64

From the above results and by comparing them with the reported results in the literature on the effectiveness of SVMs for text classification in other domain[14], it can be concluded that the accuracy and precision measures are satisfactory, whereas the result of recall measure is not satisfactory which consequently affect the values of F-measure. Although it can be noticed that for certain BCL the recall values is somewhat acceptable, the overall effectiveness is poor. The above results can be interpreted in light of of the small size of the dataset which has been used and also the short size of the questions.

5. CONCLUSION

This paper introduces text classification techniques, to a new application in the field of E-learning. It explores the effectiveness of SVMs in the classification of questions into Bloom's cognitive levels which of particular importance in question bank management systems. A dataset of pre-classified questions have been processed and divided into training set and testing set and through the use of SVM-light tool, a linear kernel SVM has been built and evaluated using several measures of effectiveness. The preliminary results obtained from the conducted experiment show a satisfactory performance of SVM, with respect to accuracy and precision measures, however a poor recall and F-measure values have been reported. The poor results of recall can be attributed to the small size of dataset and the lake of cues term for each class due to the relatively short size of questions. The next stage of this works will focus on experimenting with a large size of dataset in order to obtain more conclusive results with respect the recall and F-measure. In future the syntactic structures and the semantic knowledge of questions will be exploited to improve the effectiveness of recall and F-measure.

ACKNOWLEDGEMENT

This work is supported by the Deanship of Scientific Research in Najran University, Kingdom of Saudi Arabia under research project number NU 21/11.

REFERENCES

[1] Anderson L.W., Sosniak L.A., Bloom's taxonomy: a forty-year retrospective. *Ninety-third yearbook of the National Society for the Study of Education, Pt.2.* , Chicago, IL., University of Chicago Press, 1994.

[2] Bloom B. S., Taxonomy of educational objectives, *Handbook I: The Cognitive Domain.* New York: David McKay Co Inc. 1956.

[3] Bullington J., Endres I., Rahman M., Open ended question classification using support vector machines. *MAICS 2007*. 2007.

[4] Burges, C. J. C., A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[5] Drucker H., Vapnik V., Wu, D. Automatic text categorization and its applications to text retrieval. *IEEE Trans. Neural Network. 10*(5) 1048–1054, 1999.

[6] Dumais S. T., Chen H. Hierarchical classification of Web content. In *Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval* (Athens, Greece, 2000), 256–263, 2000.

[7] Dumais S., Platt J., Heckerman D., Sahami, M. Inductive learning algorithms and representations for text categorization. In *Proceedings of the Seventh International Conference on Information Retrieval andKnowledge Management (ACM-CIKM-98),* 148–155, 1998.

[8] Hearst, M. A., Schölkopf B., Dumais S., Osuna E., and Platt J., Trends and controversies-support vector machines. *IEEE Intelligent Systems*, 13(4): 18-28, 1998.

[9] Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of the Tenth European Conference on Machine Learning (ECML '98)*, Lecture Notes in Computer Science, Number 1398, 137–142, 1998.

[10] Jung-Lung H., A text mining approach for formative assessment in e-learning environment, PhD thesis, National Sun Yat-sen University, Taiwan, 2008.

[11] Karahoca A., Karahoca D., Ince, F. I., Gökçeli, R., Aydin N., Güngör, A. Intelligent question classification for E-learning by ANFIS. *E-learning conference '07*, 156 – 159, 2007.

[12] Klinkenberg, R. Joachims, T. Detecting concept drift with support vector machines. *Proceedings of ICML-00, 17th International Conference on Machine Learning*, 487–494, Stanford, CA, 2000.

[13] Larkey, L. S. Automatic essay grading using text categorization techniques. *Proceedings of SIGIR-98, 21st ACM International Conference on Research and Development in Information Retrieval,* Melbourne, Australia, 90–95, 2000.

[14] Lewis D. L., Yang Y., Rose T. G., Li F., RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5: 361-397, 2004.

[15] Nuntiyagul A., Naruedomkul K., Cercone N., Wongsawang D., Adaptable learning assistant for item bank management. *Computers & Education* , 50(1): 357-370, 2008.

[16] Porter M., F., An algorithm for suffix stripping. *Program*, 14(3):130–137, 1980.

[17] Sebastiani F., Machine learning in automated text categorization. *ACM. Computing Surveys* 34 (1) 1–47, 2002.

[18] Taira H., Haruno M., Feature selection in SVM text categorization. In *Proceedings of AAAI-99, 16th Conference of the American Association for Artificial Intelligence,* 480–486, Orlando, FL, 1999.

[19] Ting, F., Wei, J. H., Kim, C. T., Tian, Q. Question classification for e-learning by artificial neural network. *Proceedings of the 2003 Joint Conference of the Fourth International Conference*. 3, 1757- 1761, 2003.

[20] Vapnik, V. *Statistical Learning Theory*. John Wiley and Sons, Inc. New York. 1998.

[21] Wiggins, G. Educative assessment designing assessments to inform and improve student performance. California, CA: Jossey-Bass Inc., Publishers, 2003.

[22] Yang Y. Liu X., A re-examination of text categorization methods. In *Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval*, 42–49. Berkeley, CA, 2003.