CLASSIFICATIONS of EXAM QUESTIONS USING LINGUISTICALLY- MOTIVATED FEATURES: A CASE STUDY BASED on BLOOM'S TAXONOMY

ADDIN OSMAN[1] and ANWAR ALI YAHYA[1, 2]
[1] College of Computer Science and Information Systems
Najran University, Saudi Arabia
[2] Faculty of Computer Science and Information Systems
Thamar University, Thamar, Yemen
aomaddin@nu.edu.sa, aaesmail@nu.edu.sa

Abstract

In quality assurance, Bloom's taxonomy can be used to automatically classify educational goals, objectives, learning outcomes and questions. Some educational organizations such as accreditation bodies are in need to check the correctness of the classification of exam questions according to Bloom's Cognitive Levels and they might be faced with an enormous number of questions which would be difficult to check manually. Therefore, the usage of automatic classification of questions based on Bloom's taxonomy is highly needed. This paper aims to test and compare different machine learning methods (Naïve bayes, support vector machine, logistic regression, and decision trees) to automatically classify exam questions based on the cognitive levels of Bloom's taxonomy. The features used in the classification were based on linguistically-motivated features which are the bag of words, part of speech (POS), and n-grams. A database contains 600 exam questions for English language course was used. The results of the study show that the machine learning methods together with linguistically-motivated features perform satisfactorily in the automatic classification based on cognitive levels of Bloom's taxonomy.

Keywords: *Bloom's taxonomy, question classification, machine learning algorithms, natural language processing.*

Introduction

"All our knowledge results from questions, which is another way of saying that questioning is our most important intellectual tool." (Neil Postman)

Generally, nobody can learn without thinking which, can be considered as the centre of all learning processes. Furthermore, every educational institution stays alive only to the extent that valuable questions are developed as the driving force in a process of thinking. In addition, questions play effective role in teaching, understanding, quality assessment and evaluation. Universally, it is conventional that educational institutions convene exams to their students to evaluate their comprehension level in a subject area. The level of questions in all exams is significant; students who are given questions based on higher levels of thinking will tend to think more creatively and divergently. Effective style of questioning is always an issue to help students reach the intended learning outcomes [16].

Accreditation bodies and quality organizations worldwide were in need for a hierarchy of levels to assist teaching staff members in categorizing questions levels of abstraction - thus providing a useful structure which can be used to classify questions in terms of cognitive skills and abilities. Bloom's Taxonomy [1] which was developed by Benjamin Bloom and a group of specialists is one of the hierarchies in educational institutions, which is widely accepted and used as an important framework for teachers to assist them in crafting questions to test different cognitive levels and to ensure student's cognitive mastery [7, 10, 12, 17].

Bloom's Taxonomy contains six cognitive levels as depicted in Figure 1. The levels were arranged in a hierarchical form, starts with knowledge (recalling information), which is the lowest and the simplest level of cognition, and moves to comprehension, application, analysis, synthesis and ends up with evaluation (making judgment about something), which is the highest and most complex level of cognition. The levels were cumulative; to master any level a learner needs to master the earlier levels [1, 6].

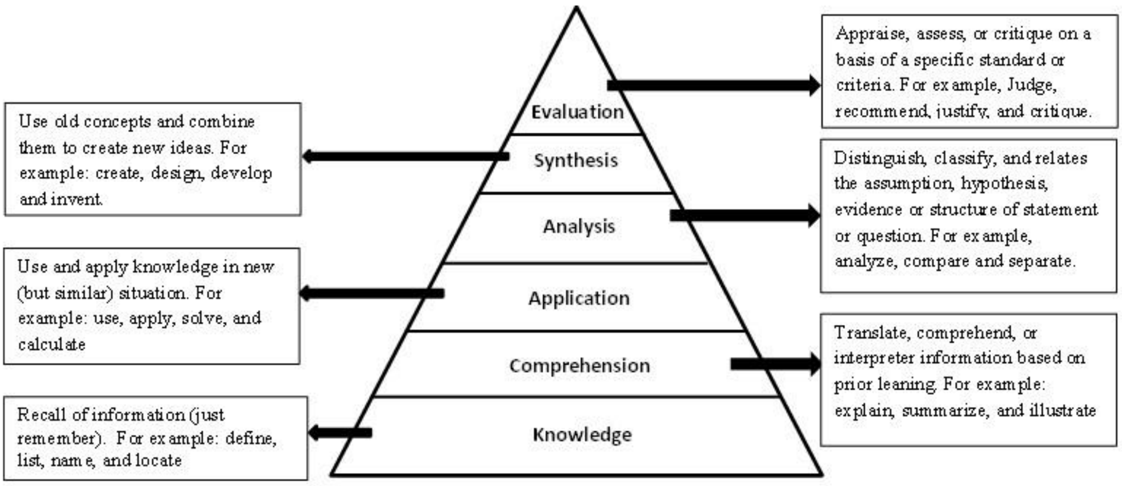

Figure 1: The Six Cognitive Levels of Bloom's Taxonomy.

Crafting questions based on Bloom's cognitive levels (BCLs) would be a very effective method of changing the weak trend of only transferring facts and recalling information to a trend of focusing on mastery of subjects and promotion of higher forms of thinking. Based on this, *BCLs* can be considered as a class label for questions. The classification of questions based on Bloom's taxonomy remains a central challenge for educators in modern times due to the lack of educators' knowledge about the taxonomy or the difficulty of the concept of the taxonomy. In addition to this, some educational organizations such as accreditation bodies are in need to check the correctness of the classification of questions according to *BCLs* and they might be faced with an enormous number of questions which, would be difficult to check manually. Therefore, the usage of automatic classification for the questions based on Bloom's taxonomy is highly needed.

Research Aim

The aim of this research is to test and compare different machine learning classifiers (Naïve Bayes, Support Vector Machine, Logistic Regression, and Decision Trees) and linguistically-motivated features to automatically classify exam questions based on the cognitive levels of Bloom's Taxonomy. As known, most content-based classification approaches frame the problem as a text categorization task; given that questions are almost always contain some form of textual content and considered as unstructured data. This research also compare and test different linguistically-motivated features such as bag of word**s** (BOWs), part of speech (*POS*), and *n*-grams and their combination in questions classification using the classifiers mentioned above. Since building text classifiers by hand is difficult and time-consuming, it is advantageous to learn classifiers from examples.

Research Questions

This paper seeks to address a number of key research questions in the course of this study:

- Can a high level of accuracy be achieved on the question classification based on Bloom's taxonomy using some machine learning models?
- Is the combination of linguistically-motivated features, especially those derived from a 'deep' syntactic analysis of text are useful for questions classification?
- In applied *NLP* classification, is it more important to focus on linguistically-motivated features or the machine learning classification models?

Related Work

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 [24], the 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 (easy, medium, and hard).

Another work on question classification using text classification techniques is reported in [5], 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.

A different work was 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.

An interesting work on question classification is presented in [2]. This work explored 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 showed 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. In addition, the work did not test the natural language syntactic and semantic knowledge in questions classifications.

Sangodiah et al. [21] performed a comprehensive review on question classification. This study showed that most the research work on question classification use a deep syntactic and semantic analysis instead of just using surface features such as bag-of-words and n-gram methods. In most of the reviewed research works they found Support Vector Machine classifier predominantly perform well when working with unstructured text data. Nevertheless, some research work has used CRF classifier to achieve better.

Mathew and Das [14] proposed a question classification algorithm based on Naïve Bayes Classifier and question semantic similarity, which mainly focused on *Numeric* and *Location* type questions.

Different questions classification using different machine learning methods and different features were proposed in [1, 8, 13, 22].

Text Classification

Supervised text classification is the process of assigning a predefined class label or labe17ls for a given instance (natural language text) based on its unstructured content. Supervised text classifiers are built based on training corpora containing the correct label or labels for each instance which is considered as a bag of words. In our case, we would like to assign a class label, which is one of the six *BCLs* for a given question. The framework used by a question supervised classifier used in this paper is depicted in *Figure 2*.

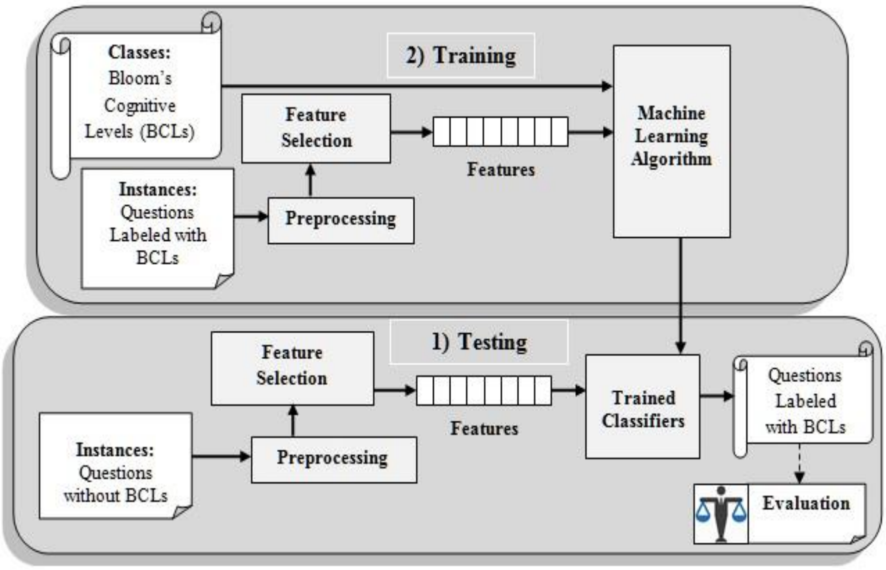

Figure 2: Framework for Supervised Classifier of Questions.

The framework is quite similar to the conventional text classification framework which consists of input (instances of questions with the class labels; *BCLs*), preprocessing, feature selection, and classification stages which contains the evaluation process of the classifier; assessing the performance of the classifier experimentally.

Preprocessing

The preprocessing stage used in the framework contains the mostly used preprocessing tasks in *NLP* [11], which are:

- Tokenization: Breaking up a stream of text into words, phrases or symbols; tokens.

- Part of Speech Tagging (*POS)*: *POS* tagging; assigning parts of speech to each word or token, such as noun, verb, adjective, etc.
- Stemming: Is the process for reducing or transforming inflected words to their base or root words; normal form. For instance converting "flew" and "flying" both into "fly".
- Lemmatizing:
- Lowercase Conversion
- Singular Value Decomposition.
- *N*-gram: sequence of *N* adjacent tokens. Therefore, 3-gram would be three consecutive words. This creates the bag of words model containing some information about word ordering.

The bag of words model is a conventional method to represent documents in matrix form.
Consequently, $n \times t$ document-term matrix would be formed, where n is the number of documents, and t is the number of unique terms. Every unique term would be represented by a column, and each cell (*i*, *j)* keeps the number of term *j* which are in document *i*. A simple term frequency bag of words can be used. Whereas, one of the widely used weighting factors is the term frequency- inverse document frequency (*tf-idf*). It uses other formula (as shown in Equation 1) to calculate the term frequency than just the numbers.
TF-IDF is the most widely used term weight algorithm nowadays. tf-idf is the best known weighting scheme in information retrieval.

Feature Extraction and Selection

To analyze unstructured text data, one has to extract info from text and turn that into numerically structured data matrix, which is a representation of documents. The feature extraction stage generally utilizes the vector space model [20] that makes use of the bag-of-words approach. Some Natural Language Processing (NLP) techniques are used. Provide features for feature selection.
The feature selection stage can be viewed as a problem of dimensionality reduction. Occasionally, it is described as the exclusion of noisy words. However, Joachims showed that even the words considered as noise sometimes can have predictive value [9]. In this stage a decision must be made about what to be considered as a feature term.
It forms and important subset within the much large area of supervised text classification, most of the time, employs the filter methods such as document frequency, mutual information, information gain, chi-square, Gini index, and distinguishing feature selector.

Classification Models

Finally, the classification stage uses well-known and successful pattern classification algorithms, e.g., naïve Bayes classifier, support vector machines, decision trees, artificial neural networks, and naïve Bayesian classifier.
If a huge amount of data is available, then the choice of classifier probably has little effect on the results and the best classifier choice may be unclear [3]. One big challenge in educational data mining is the availability of data. Actually there are formidable amount of scattered data (e.g. exams and tests) but not collected, organized and formalized. Whenever, the data to be used for training a supervised classifier is fairly little, the machine learning theory advices to use a classifier with high bias/low variance. Based on that we decided to use Naïve Bayes, SVM, logistic regression, and decision trees in this research.

Naïve Bayes Classifiers

A naive Bayes (*NB*) classifier [19] is a simple probabilistic classifier based on applying Bayes' theorem with naïve assumption independence between every pair of features. Assume a variable *C* denotes the class of an observation *O*. The class of the observation O can be predicted using the Bayes rule, we need to calculate the highest posterior probability of [23]:

$$P(C|O) = \frac{P(C)P(O|C)}{P(O)} \tag{1}$$

In the *NB* classifier, using the assumption that features $O_1, O_2, \ldots, O_n$ are conditionally independent of each other given the class, we get [23]:

$$P(C|O) = \frac{P(C)\prod_{i=1}^{n} P(O_i|C)}{P(O)} \tag{2}$$

Corpus Processing

We tested two text preprocessing procedures on our system: stemming and lemmatization. Stemming is a transformation in which a given word is stripped of its suffixes [18], thereby reducing it to its stem. In this manner, multiple different words can be normalized into a single morphological form, thus reducing the overall number of different words. Hence, the application of stemming to a text corpus ultimately leads to the creation of a smaller semantic space. This can, in effect, reduce the overall computational costs. In our system we utilized the standard Porter stemmer (Porter, 1980).

Methodology

In this research, different algorithms have been used to automatically assign a *BCL* to a given teacher's classroom questions. The *NLP* features such as bag of word**s**, part of speech (*POS*), *n*-grams, and others (e.g. punctuation inclusion, stem *n*-grams, stop words and stretchy patterns) are used in this research. In doing so, six main steps have been carried out: questions collection, questions preprocessing, terms selection, questions representation, questions classifiers learning, and questions classifiers evaluation.

Questions Collection

The dataset used in this work is a set of questions collected over a period of three semesters of the academic years 2011and 2012, from a set of courses lectures in computer science program at Najran University. The procedure of data collection was basically based on courses lecturers, who were asked to keep records of questions they ask their students in their classrooms. For this purpose, a questions collection form was designed and used by the lecturers and submitted by the end of semester. All program's lecturers are informed about Bloom's taxonomy and the objectives of this work.

The total number of collected questions is 600 questions (100 for each BCL) in English language which is the adopted medium of teaching in the program. After collection process, annotation of questions according to the six levels of BCLs was carried out. It worth mentioning that although an initial annotation of the collected questions was carried out by the lecturers during questions collection process, a second cycle of annotation was carried out with a help of pedagogical expert. In this process a kappa statistic [7] is used to measure the agreement between the two cycles of annotation and the obtained kappa, 0.82, indicates a very good agreement.

Questions Preprocessing

In this step, each question in the data set is manipulated by applying the following preprocessing steps:

- Reducing of the question text to lower case characters.
- Punctuation removal: all types of punctuations are removed from the question.
- Removing low frequency terms: all terms with frequency < 3 are removed.
- Tokenization: a token is a maximal sequence of nonblank characters. In this process, tokens consisting purely of digits are discarded.
- Stemming: tokens were stemmed with Porter stemmer.

Terms Selection

In this step, a feature selection approach is applied to select from the original term set (a set containing all the terms from questions) a subset such that only the most representative terms are used. A computationally easier term selection approach is the filter approach, which selects a subset of terms that receive the highest score according to a function of term importance for the classification task.

Questions Representation

Basically, question's text cannot be directly interpreted by a classification algorithm, therefore, a conversion procedure to map a text of a question q_j into a compact representation of its content need to be uniformly applied. A common method for representing a question q_j is as a vector of term weights $<w_{1j}, \ldots, w_{Tj}>$, where T is number of terms selected in the term selection step, and $0 \leq w_{kj} \leq 1$ 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 learning algorithm used. The most common form of non-binary weight uses the standard *tfidf* function, which is defined as

$$tfidf\,(t_k, q_j) = \#(t_k, q_j).\log \frac{|Tr|}{\#Tr(t_k)} \qquad (3)$$

Where $|Tr|$ denotes the number of questions in the training set, $\#(t_k, q_j)$ 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. Term weighting is computed as in Eq. 3, and length normalization is applied as follows:

$$w''_{q_j}(t_k) = \frac{w_q(t_k)}{\sqrt{\sum_l w_{q_j}(t_l) \times w_{q_j}(t_l)}} \qquad (4)$$

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

Evaluation and Results

We conducted a series of experiments to compare and evaluate different machine learning algorithms and different natural language processing features to classify the exam questions. To perform the experiments some settings were needed at the beginning.

Experimental Design

To test the performance of the classification models, the randomized 10-fold cross-validation was used. The question data were randomly divided into the folds. One fold was selected in rotations until all folds were selected. In each rotation, the selected fold was kept for testing and the rest nine folds were used for training the model.

Several metrics for evaluating the extracted features were used:

- **Total Hits:** The number of documents in the training set that contains this feature.
- **Target Hits:** The number of documents with the target annotation containing this feature.
- **Precision/Recall/ F-Score**: Measures of sensitivity and specificity those are common in language technologies research.
- **Kappa:** The discriminative ability over chance of this feature for the target annotation.
- **Correlation:** For numeric prediction, Pearson's between class value and feature alone (uninformative for nominal classification).

Extracting Features

The LightSide, which uses the Stanford POS tagger is the tool used in this research. The tool requires a simple representation of format. The data contained in a spreadsheet, with every row representing a training example, except the first, which lists the names of the fields of the data. In our case we have two columns; the first column contains the six Bloom's Taxonomy Cognitive Levels. The second column contains the question text. The first process in the tool after uploading the questions file to the tool is to convert (extract features) from the file. The basic features were extracted using the tool. The basic features are: Unigrams, Bigrams, Trigrams, POS Bigrams, POS Trigrams, Word/POS Pairs, and Stem N-grams.

The features were extracted by considering all combinations of the 32 feature selection methods mentioned above; that means 32 groups of features were created. For example, when we have used the Unigrams features we got 217 features, the punctuations were used, trach feature hit location was true and the type of features was nominal. And when we used all features selection methods we got 3535 features, including the punctuations, track feature hit was true and the type of feature was nominal.

Classification:

Multiple machine learning classifiers (Naïve Bayes, Logistic Regression, Support Vector Machines, Decision Trees, and J48) were trained and tested using the features created previously. The reliability of the models was reported as accuracy and kappa because their classifications were nominal.

A slightly more detailed description of the accuracy is given via a confusion matrix (as shown in Figure 3), describing the types of errors that are being made by a model. This confusion matrix is also often called a contingency table; accurate decisions are made along the diagonal, with prediction labels represented in each column and actual labels in each row.

Act \ Pred	Analysis	Application	Comprehension	Evaluation	Knowledge	Synthesis
Analysis	70	6	6	6	6	6
Application	11	55	9	1	10	14
Comprehension	7	5	68	3	14	3
Evaluation	9	4	5	60	9	13
Knowledge	4	3	4	0	88	1
Synthesis	10	11	6	9	5	59

Figure 3: Model Confusion Matrix for Naive Bayes Classifier.

Results

To answer our first research question of **c**an a high level of accuracy be achieved on the question classification based on Bloom's taxonomy task using some machine learning models (NB, Logistic Regression, SVM and Decision Trees)?, we tested the four machine learning models on different linguistically based features. The statistics from these four models are presented in Figure 4.

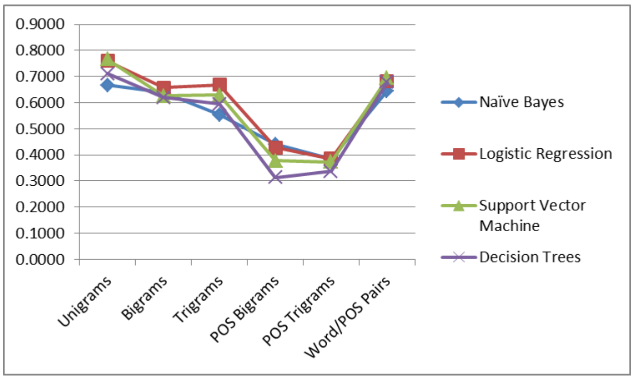

Figure 4: The Effect of Separate Feature Types in the Classifiers Accuracies Based on Single Features.

Results from Figure 4 show that the four models capture good classification accuracy (all of them above 0.6) when using the unigram features. In this case, the best classification result acquired by support vector machine (0.7667), the second is the logistic regression (0.76), the third is the decision trees (0.7417) and lastly the Naïve Bayes (0.6667). This shows that the four models are effective in classifying questions based on Bloom's taxonomy. The Word/POS pairs and Bigrams features also show good classification results; all accuracies are greater than 0.6 and less than 0.7. The accuracies of Trigrams, POS Bigrams and POS Trigrams were not good (mostly less than 50%).

To answer our second research question of is the combination of linguistically-motivated features, especially those derived from a 'deep' syntactic analysis of text, useful for questions classification?, we tested the four machine learning models on all 18 combinations of the six features shown in Figure 4. The statistics from these four models are presented in Figure 5. The results show a bit improvement to accuracies in all combinations. When we used single feature the highest accuracy was obtained by unigram and support vector machine model (0.7667) and in the combinations, the highest accuracy shows a bit increment (0.7683) that is by combining Unigrams and Bigrams using the logistic regression classifier.

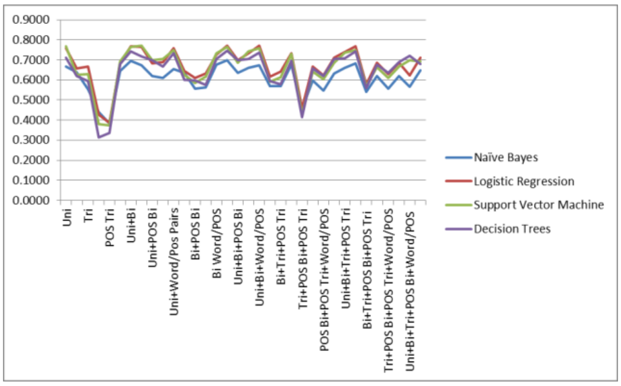

Figure 5: The Effect of Composed Feature Types in the Classifiers Accuracies.

To answer our last research question of in question classification, is it more important to focus on linguistically-motivated features, or the machine learning models?. Figures 4 and 5 show positive correlation between the four models used in this research; they are fluctuating together. Nevertheless, the support vector machine and logistic regression show better accuracies in most cases than Naïve Bayes and decision trees. As a result, it is important to focus on both.

Conclusion and Future Work

From our investigation we found that machine learning methods implemented with linguistically-motivated features, especially those derived from a "deep" syntactic analysis of text are useful for questions classification based on Bloom's taxonomy cognitive levels. The combinations of linguistically-motived features improved a bit the classification accuracies.

Future research will focus on enhancing the database by testing a huge question database from different academic fields. Further, we believe that the term frequency – inverse document frequency (TF-IDF) which is widely used in information retrieval and text mining should be implemented to identify its effect on questions classification based on Bloom's taxonomy. In addition, the semantic meaning of the text can be tested whether they improve the classification accuracy and efficiency.

References

1. Abduljabbar D. A., & Omar N., (2015), *Exam Questions Classification Based on Bloom's Taxonomy Cognitive Level Using Classifiers Combination*, Journal of Theoretical and Applied Information Technology, vol. 78, no. 3, pp. 447-455.
2. Anwar A. Y. & Addin O., (2011), *Automatic Classification of Questions into Bloom's Cognitive Levels Using Support Vector Machines, The International Arab Conference on Information Technology*, Naif Arab University for Security Science (NAUSS), Riyadh, Saudi Arabia, pp. 1-6.

It is worth mentioning that the number of the features increases exponentially when the number of the combined features increases (as shown in Figure 6).

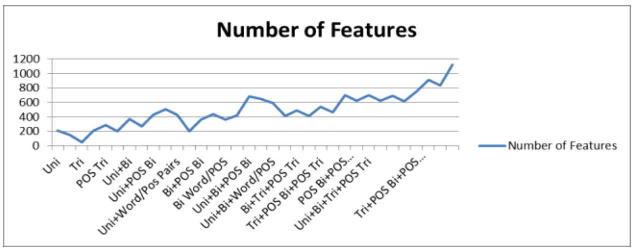

Figure 6: The Relationship between Feature Numbers and Combined Feature Types

3. Banko M. & Eric B., (2001), *Scaling to Very Very Large Corpora for Natural Language Disambiguation*, ACL '01 Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics Stroudsburg, PA, USA, pp. 26-33.
4. Bloom B. S., Englehart M. D., Furst E. J., Hill W. H., & Krathwohl D. R., (1956), *Taxonomy of Educational Objectives*, *Handbook 1: Cognitive Domain*, New York, Longmans, Green.
5. Bullington J., Endres I., & Rahman M., (2007), *Open-ended Question Classification Using Support Vector Machines*, MAICS 2007.
6. Bush H. M., Daddysman J., & Charnigo R., (2014), *Improving Outcomes with Bloom's Taxonomy: From Statistics Education to Research Partnerships*, Journal of Biometrics & Biostatistics, vol. 5, no. 4.
7. Chang W. C., & Chung M. S., (2009), *Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition Level of English Question Items*, IEEE, pp. 727-733.
8. Jayalakshmi S., & Sheshasaayee A., (2015), *A Review on Question Classification Using Machine Learning Based on Semantic Features*, International Journal of Engineering Sciences and Research Technology, vol. 4, no. 2, pp. 561-564.
9. Joachims T. , (1997), *A Probabilistic Analysis of the Rocchio Algorithm with tfidf for Text Categorization*, 14th International Conference on Machine Learning, Morgan Kaufmann Publishers Inc, pp. 143–151.
10. Jones K. O., Harland J., Reid J. M. V., & Bartlett R., (2009), *Relationship between Examination Questions and Bloom's Taxonomy, FIE'09 Proceedings of the 39th IEEE International Conference on Frontiers in Education Conference*, pp. 1314-1319.
11. Jurafsky D., & Martin J. H,, (2000), *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Prentice Hall Englewood Cliffs.
12. Kunen S., Cohen R., & Soman R., (1981), *A Levels-of-Processing Analysis of Bloom's*

Taxonomy, Journal of Educational Psychology, vol. 73, no. 2, pp. 202-211.

13. Marquardt, M., (2005) Leading With Questions- How Leaders Find the Right Solutions by Knowing What to Ask. San Francisco, CA: Jossey-Bass, pp. 32, 69–74.
14. Mathew J., & Das S. N., (2015), Question Classification using Naive Bayes Classifier and Creating Missing Classes using Semantic Similarity in Question Answering System, International Journal of Engineering Trends and Technology (IJETT), 23 (4), 155-160.
15. Nuntiyagul, A.; Naruedomkul, K.; Cercone, N. and Wongsawang D., (2008) Adaptable Learning Assistant for Item Bank Management, Computers & Education , 50(1): 357-370.
16. Omar N., Sufi Haris S., Hassan R., Arshad H., Rahmat M., Zainal N. F. A., & Zulkifli R., (2012), Automated Analysis of Exam Questions According to Bloom's Taxonomy, Elsevir, SciVerse ScienceDirect, Procedia - Social and Behavioral Sciences, vol. 59, pp. 297 – 303.
17. Oliver D., Dobele T., Greber, M., & Roberts T., (2004), This Course has a Bloom Rating of 3.9, Proceedings of the Sixth Conference on Australasian Computing Education, Australasian Computing Society Inc, Dunedin, New Zealand, vol. 30, pp. 227-231.
18. Porter M., (1980), An Algorithm for Suffix Stripping, Program, vol. 14, no. 3, pp. 130–137.
19. Rish I., (2001), an Empirical Study of the Naive Bayes Classifier, IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence.
20. Salton G., Wong A., & Yang C. S., (1975), Vector Space Model for Automatic Indexing, Communications of the ACM, vol. 18, pp. 613–620.
21. Sangodiah A., Muniandy M., & Heng L. E., (2015), Question Classificaiton Using Statistical Approach: A Complete Review, Journal of Theoretical and Applied Information Technology, vol. 71, no. 3, pp. 386-395.
22. Singh V., & Dwivedi S. K., (2015), Two Way Question Classification in Higher Education Domain, International Journal of Modern Education and Computer Science, vol. 9, no. 59-65.
23. Sonat T., & Musa M., (2013), Learning the Naïve Bayes Classifier with Optimization Models, International Journal of Applied Mathematics and Computer Science, vol. 23, no. 4, pp. 787–795.
24. Ting F., Wei J. H., Kim C. T., & Tian, Q., (2003), Question Classification for e-Learning by Artificial Neural Network, Proceedings of the 2003 Joint Conference of the Fourth International Conference, vol. 3, pp. 1757- 1761.