Journal of Theoretical and Applied Information Technology
30th June 2017. Vol.95. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TAXONOMY BASED FEATURES IN QUESTION CLASSIFICATION USING SUPPORT VECTOR MACHINE

[1] ANBUSELVAN SANGODIAH, [2] ROHIZA AHMAD, [3] WAN FATIMAH WAN AHMAD

[1]Department of Information System, UTAR, Kampar, Malaysia

[2]Department of Computer and Information Sciences, UTP, Tronoh, Malaysia

[3]Department of Computer and Information Sciences, UTP, Tronoh, Malaysia

E-mail: [1]anbuselvan@utar.edu.my, [2]rohiza_ahmad@utp.edu.my, [3]fatimhd@utp.edu.my

ABSTRACT

One of the areas in text mining which is text classification has attracted much attention in various industries and fields lately. This is because the text classification has the ability in labelling text documents to one or more pre-defined categories based on content similarity. As text classification emphasizes on document level, question classification works at finer level such as sentence and phrase. Several studies on question classification in respect to Bloom taxonomy to measure cognitive level of learners in higher learning institutions have been carried out in the past. But, existing feature types in the past work may work reasonably well on data sets consisting of questions that are too specific to one particular field or area which will result in having multiple classifiers to be built for questions involving various fields or areas. Certainly, feature types play an important role in improving the accuracy of classifier. Past related work emphasizes on feature types such as bag of word (BOW) and syntactic analysis in question classification. In this study, a new feature type named taxonomy based is proposed to improve the accuracy of question classification for data sets having questions from various fields. The performance of question classification using the new feature type between data sets consisting of questions from specific and various areas will be compared. Support Vector Machine classifier will be used as it is known for high accuracy in text classification. The outcome of this study shows that the taxonomy based features has the ability in improve the accuracy of classifier involving data sets of questions from various fields.

Keywords: *Feature Type, Question Classification, Support Vector Machine, Bloom Taxonomy, Bag-Of-Words*

1. INTRODUCTION

Text mining which comes under the umbrella of data mining is a systematic process to extract and discover hidden patterns from a large scale of text data [1]. As the nature of text mining primarily deals with unstructured text data, therefore it has to work closely with other areas, such as information retrieval, information filtering, automatic summary, text clustering, text classification, natural language processing, artificial intelligence, machine learning, pattern recognition, statistics, visualization and so on [2], [3].

Text categorization or better known as document classification is one of the areas in text mining which is becoming increasingly popular. As its name implies, it is a task of labelling or assigning text documents to one or more pre-defined labels based on content similarity [4], [5]. This text classification is based on supervised learning technique where a model will be created using large collection of document with predefined labels or categories. This is to train the model so that later the same model can be used to label the unlabeled documents based on the likelihood inferred [6], [7]. The work in [5] has reported a better accuracy in classifying text documents. Also, reported in [8] where classifying documents is used in web environment. However, in question classification it will be a painstaking task.

Question classification works in the same manner as document classification. However, the former has a unique challenge which demands a separate study of its own. Due to the shorter length of questions as compared to documents' content, only little information or words can be worked upon when classifying the questions [9]. In other words, the minimal or limited words available have an impact on the discriminating power of classifiers [10]. Nevertheless, question classification has been popular in the areas of question answering system and information retrieval system and it is evident the work led by [11], [12].

While in the educational environment, much of the research work in question classification revolves around test or exam questions in the context of assessment. Using question classification to classify questions formulated using Bloom Taxonomy guidelines in accordance to Bloom Taxonomy level has received much attention. The accuracy of question classifier such as SVM certainly depends on feature extraction task in extracting relevant features in questions. Thus, the use of appropriate feature types has direct impact to the accuracy of classifier in classifying questions [13]. Despite there exist some researchers [14],[15],[16],[17],[18],[19] who have worked on classifying questions in accordance to BT through machine learning or rule based in the past, they may work well in classifying questions from one specific area or field only but not from different areas. This is because the question classifier using machine learning based on BOW (bag-of-words) feature type is area or domain specific which will work well in classifying question from a specific area but not from multiple areas. Therefore, if there are questions from multiple areas, then multiple classifiers will have to be built to classify questions For an instance, one question classifier will be built to classify questions from computing area while another one to classify questions from business area. At some point, the number of classifiers that need to be built will be proportional to the number of fields or areas.

In view of the above mentioned issue, it is essential to come up with a question classifier such that can classify questions from various areas or fields with a decent accuracy. In order to do that, the existing feature type that is BOW (bag-of-words) will have to be extended. Due to high variances between questions from different fields in terms of patterns and structure of questions, certainly, extending bag-of-words feature type by identifying and extracting other important and relevant features in questions is the way forward to address the aforementioned issue Therefore, in this study, taxonomy based feature type is proposed in order to build a classifier which is able to classify questions from various areas or fields with a decent accuracy. Also, with the development of the feature extraction model in extracting the proposed feature type, it is hoped that in the future the outcome of this study will witness web based learning management system being integrated with question classification component in the assessment module as reported in [20].

The rest of this paper will discuss the past related work, the proposed solution and the outcome of the study.

2. LITERATURE REVIEW

As stated in the introduction section, there already exists some research works which have been conducted in relation to question classification. For instance, in [20], Artificial Neural Network was used to classify questions in e-learning. In this work, questions were automatically classified into three difficulty levels. And the work has considered five aspects in feature selections which are query-text relevance, mean term frequency, length of Q&A, term frequency and distribution of Q & A in text. Using the selected features, it was reported that 78% accuracy was able to be obtained. In the next discussion, the research works which are very much relevant to the current study will be examined.

In another research work [21], the researcher uses Support Vector Machines with bag-of-word feature type to classify questions in accordance to Bloom taxonomy and though preliminary results show a satisfactory performance of SVM in accuracy and precision, however a poor recall and F-measure values have been reported. The same researcher in [15], uses SVM to classify classroom questions in accordance to BT. The work, generally, is good examination of term frequency and removal of stop words in classifying questions in accordance to Bloom taxonomy and the work compared with several classifiers and SVM is found to be superior compared to kNN and Naïve Bayes. This work focuses on bag-word-feature type. Also the same researcher also reported in [22] to investigate how linguistically motivate feature types such as POS Tagging, unigram, bigram, trigram, POS Bigram and Trigram perform with several classifiers. And the best classification result acquired by using SVM with unigram. Generally, the accuracies of classification of using each type of feature fall between 0.6 and less than 0.8. The results indicate that these linguistic features do improve classification accuracies. However, the performance of classifiers in classifying questions from various areas using

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

linguistically motivate feature types remains to be seen.

Similar work by [16] has used Artificial Neural Network classifier to classify questions in accordance to Bloom taxonomy. This research is very much focused on introducing three types of features sets to reduce dimensionality of the feature space and these feature sets are basically serve as feature reduction methods. The reduction methods (DF and CF-DF) have been found to have good performance in terms of speed of classifying questions in accordance to Bloom taxonomy. And the feature type used in the study is based bag-of-words.

Comparison between various classifiers such as SVM, Naïve Bayes and k-NN and feature selection methods such as Chi-Square, Mutual Information and Odd Ratio has been the main focus in the study reported [14]. The work uses bag-of-word (unigram) as feature type extracting features from exam questions based on Bloom taxonomy. The outcome of the study identifies best classifier and selection method in order to achieve a decent accuracy in classifying exam questions. The data set consists of exam questions confining to one particular area which is programming.

All the aforementioned work done by past researchers use exam question based on English language. However, question classifier such as SVM can also be used to classify questions in different language such as Indonesia as reported in [23]. Though the data set used in the work consists of exam questions in Indonesian language, Bloom Taxonomy is used to categorize the difficulty level of questions. The feature types that were used are bag-of-words (unigram), POS tagging, question length and keyword. The term weighting that was used is based on frequency that is counting the occurrence of terms in a question. The size of data set is not substantial that is 130 questions. The outcome of the work has shown some promising results in classifying questions in Indonesian language.

Rule based approach can also be used in classifying questions and the work by [18], [19] achieves it. Though the work shows some reasonable findings, however there is a need to have a lot of rules in order to improve in classifying questions. Similar work also reported in [24] where the work proposed rule based system to capture sequence patterns in order to categorize questions according to Bloom taxonomy.

Other work by [17],[25], categorizes Bloom taxonomy question based on matching technique. No text classifier was used in the research work.

Classification of non-exam based questions can also be done using unsupervised approach which reported in the work [26]. The study has proposed to use clustering technique using k-means algorithm to cluster questions from Stack Overflow based on topic modeling. Using this technique questions asked in the Stack Overflow can be categorized in accordance to Bloom taxonomy.

In a nutshell, most of the previous work measure the performance of question classifier with data sets consisting of exam questions from a specific area or field using bag-of-words feature type. And the accuracy of question classifier using bag-of-words feature type may not go down well with data set consisting of questions from various areas or fields. In view of this, this study proposes a new feature type called taxonomy based to improve the accuracy of question classification involving data sets from multiple areas or fields. The performance of the new feature type in question classifier will be compared between data sets from multiple areas and specific area respectively.

3. METHODOLOGY

A conventional text classification framework consists of preprocessing, feature extraction, and classification stages. The preprocessing stage usually contains the tasks such as tokenization, stop-word removal, lowercase conversion, and stemming or lemmatization [27]. The feature extraction stage generally utilizes the vector space model [28] that makes use of the bag-of-words approach [29] and followed by determining the appropriate term weighting to represent features. Finally, the classification stage uses well-known and successful pattern classification algorithms, e.g., support vector machines, decision trees, artificial neural networks, and naïve Bayesian classifier [30].

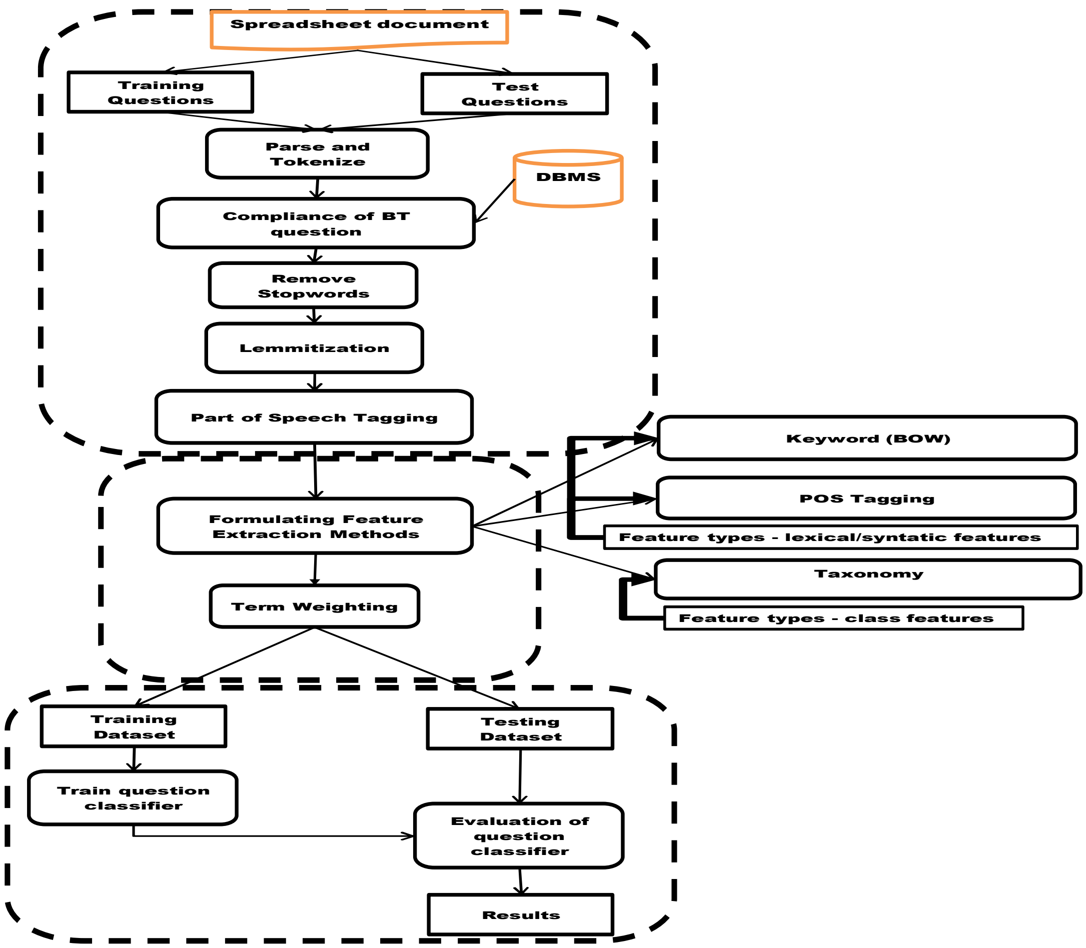

Figure 1: Question Classification Framework

A preliminary study was conducted to investigate the effectiveness of the proposed framework. A data set consisting of 415 questions have been collected from several universities in Malaysia which follows strictly the guidelines and principles of BT in preparing questions to assess learners according to BT cognitive level. Several domain experts have involved to validate the assigning of labels (BT levels) to questions. The collected questions have gone through a thorough preprocessing stage where questions were checked if they are in compliance with BT guidelines. Besides that, questions which are highly similar with one another were dropped in order to avoid of extracting similar or duplicated features from the data set. Obtaining unique features as many as possible from questions from various fields is the priority. In this study, two different datasets have been prepared. The first data set comprises questions from various areas or fields (Computing, Business, Multimedia, Programming, Social Science, Mathematics, Science) and the questions are segregated into six levels of BT. The second data set comprises questions from only Business area with the size of data set consisting 168 questions. The questions from each field obtained from various subjects or courses and each question may consist of more than one BT headword. The first and the second data set will be referred as multi data set and single data set respectively.

As shown in Figure 1, all the questions are stored in spreadsheet document and it will go through the preprocessing stage. This preprocessing starts with parsing and tokenizing of questions, breaking down each question into tokens. Each question in the form of tokens will be checked against the BT keyword list stored in database if the question consists of at least one BT keyword. Any question that does not contain BT keyword will not

Journal of Theoretical and Applied Information Technology
30th June 2017. Vol.95. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

be processed further. Stop words removal task will come into play by removing insignificant words such as a, the, an and etc. To remove the stop words, a list of stop words called a stop list is created. This process can also be considered as a feature reduction method, because it reduces the size of the initial feature set. Lemmatization is used instead of stemming as the meaning of words is important in extracting semantic features. WordNet lemmatize has been used in this study as this is free and publicly available [31]. Next part-of speech-tagging task is performed where Stanford Parser has been used as it has a better accuracy as reported in [32]. The result from these steps is an initial feature set equivalent to having bag-of-word features.

As shown in Figure 1, the feature types consist of three types which are keyword or better known as bag-of-words, POS Tagger and Taxonomy.

Keyword (Bag-Of-Words): A question is made up of several words and each word is a feature by itself and it belongs to one of the categories of word such as verb, noun, adverbs, adjectives, conjunction and others. It has been known that the word which belongs to noun is considered significant in classifying documents into an appropriate category or class. Therefore, words which are in noun category will be closely examined for significance in questions. Bag of words feature type is common and is widely used in both document and question classifications. In the context of natural language processing, it is known as lexical feature as this type of feature is captured directly from the sentence.

POS Tagging: This is the process of classifying or categorizing the words of a sentence into its part of speech (morphological classes) and to label them is known as part-of- speech tagging (POS tagging), or simply tagging. Determining for each word whether it is a verb, noun, adjective or something else is what POS tagging is all about. So, the feature of POS will be the category or classification of words such as VBZ, NN and so forth. The POS feature is resulted from syntactic extraction as it requires the understanding of structures of words and between words in sentence. And this type of feature is classified as syntactic feature. An example of a sentence with their tags: [[('Advise', 'VB'), ('the', 'DT'), ('five', 'CD'), ('steps', 'NNS'), ('as', 'IN'), ('recommended', 'VBN'), ('by', 'IN'), ('Porter', 'NNP'), ('and', 'CC'), ('Millar', 'NNP'), ('that', 'WDT'), ('can', 'MD'), ('be', 'VB'), ('applied', 'VBN'), ('by', 'IN'), ('business', 'NN'), ('managers', 'NNS'), ('in', 'IN'), ('taking', 'VBG'), ('advantage', 'NN'), ('of', 'IN'), ('information', 'NN'), ('technology', 'NN')]]

Taxonomy: In general, this type of feature is also known as class based features as it has strong link with BT cognitive levels. This taxonomy features are divided into two types which are general and specific. In practice, a question formulated in accordance to Bloom Taxonomy usually will have at least one verb keyword which has strong link to a cognitive level or category of BT. As such, the verb keyword will be referred as headword as it plays important role in determining the BT level of a question. Usually, the headword occurs as a first word in a particular question or in sub question and there may have more than one headword in one question. These headwords will be extracted and treated differently whereby it will become part of features besides bag-of-word features. For example, in the question: Interpret and discuss the concept that can be explained from the graph theory.

The keyword or bag-of-word features for the above-mentioned question will result in the following set: bag-of-words = {interpret, discuss, concept, explain, graph, theory}. And followed by general taxonomy features, the following features will be obtained; taxonomy = {interpret, discuss}. So, the final extracted features from the question will be; final set = {interpret, discuss, concept, explain, graph, theory, interpret, discuss}.

The headwords interpret and discuss occurring in the beginning of the set is treated as keyword or bag-of-word features while those at the end of the set (final set) that have been separated and treated differently are taxonomy features. By doing this, the accuracy of classifier in classifying questions in accordance to BT can be enhanced. As a matter of fact, the researcher [13] in the field of information retrieval has mentioned that separating main keywords from other keywords that have strong link to labels might enhance the performance of classifier. This type of feature called general taxonomy because the features that are extracted are important keywords which involve all the levels or categories of BT and not confining to a specific BT level.

The taxonomy specific feature is defined as features that are specific to a particular BT level or category unlike general taxonomy features. This

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

feature reflects the characteristics of a specific BT level and are pre-determined. Thereafter, the questions are checked for any occurrences of these pre-determined features. In this case, certainly, binary term weighting scheme is the most appropriate to be used compared to other popular term weighting schemes such as TF-DIF and Term Frequency (TF). Besides improving the accuracy of classifier with introducing taxonomy based features to address data sets from various fields, these taxonomy feature types can also be used to address ambiguous questions resulted from the existence of shared or overlapping headwords in Bloom taxonomy as reported in [19],[17],[26],[24]. From initial investigation of the data sets, it shows that most of the ambiguous questions involves BT levels which are Comprehension, Apply and Analysis. Therefore, in this study, the pre-determined features are derived from Apply and Analysis BT levels. Certainly, this taxonomy specific feature type can also be extended to other BT levels. Future work may consider deriving features from other BT levels. In this way, the most important features specific to each BT level are determined.

According to Bloom Taxonomy handbook [33], the category of Apply in BT is about the ability to use learned materials in a new situation or a given situation. The 'learned materials' can be referring to rules, method, laws or theories while 'new situation' can be domain or environment in which the 'learned materials' are applied. Therefore, any question which has the component of learned materials and new situation reflects the characteristic of Apply BT level and these two components are pre-determined features that are specific to only Apply BT level.

For example, in the question: Define and discuss how the bubble sorting technique can be applied in web environment. The question which contains applied keyword where its base form is apply belongs to Apply BT level. Based on this keyword, the question can be rewritten in the following way which truly reflects the characteristics of Apply level in BT.

Active sentence: *Apply bubble sorting technique in web environment*.

The above-mentioned question contains the two components which are {bubble sorting technique} and {web environment}, thus these predetermined features learned materials and new situation are said to have occur in the question. Determining the appropriate BT level for the question based on merely bag-of-word features, the first headword define belongs to Knowledge level while the second headword discuss belongs to the Comprehension BT level and eventually the question will be classified into Comprehension level as the headword which is in the higher level takes precedence over other headword of lower level of BT. However, by doing this, the question is misclassified into Comprehension level when it is supposed to be at Apply BT level. The work by [34] indicated that among the keywords in a question, the one which corresponds to the highest level of taxonomy must be assigned to the question.

The level or category of Analysis in BT is about the ability to break down materials into its component parts. Complex relationships including part-whole is one of the characteristics of this level. The predetermined feature derived from this level will be the occurrence of part-whole. And this feature generally is specific to Analysis BT level.

For example, in the question: Explain all the main characteristics of bubble sorting technique and choose one characteristic that closely matches with other characteristics of different types of sorting technique.

The above-mentioned question contains part-whole relationship where there is a need recognize one of the characteristics of bubble sorting technique out of several characteristics. The whole aspect covers all the main characteristics that make up bubble sorting while part aspect refers to the one of them in whole. This question requires the need to examine in detail each characteristic before it can be compared with other characteristics.

The aim of taxonomy features both general and specific is to improve the accuracy of question classification for questions from various areas or fields. Thus, by having these features, it is expected that question classifier will be able to classify questions from various areas or fields with a decent accuracy.

In order to extract all the aforementioned features, four features extraction methods have been developed. Each feature extraction method will extract different type of feature.

Once all the feature extraction methods have been applied to questions, the next step is to

Journal of Theoretical and Applied Information Technology
30th June 2017. Vol.95. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

derive some measures for the extracted features so that an appropriate term weighting can be formed and used before training a classifier to classify questions. The term weighing that is used in this study is binary. Though there are popular term weighting schemes such as TF-DIF and TF, these schemes work best in a situation in which there are frequent occurrences of a particular word in a document or in a question. But occurrences of a particular word more than once in a question is rare as it contains limited words unlike in documents. In this study, feature selection using statistical techniques such as chi-square to extract keywords which are highly correlated with class labels is not done as to avoid losing important keywords that rarely occur in question. It also reported in [22] that these infrequent occurrences of keywords or noise may have predictive value.

The library sckit-learn [35] which has been tested in [36] has been used to develop the classifiers and evaluate its accuracy in classifying questions. This library is known to be popular as reported in [37]. As for preprocessing involving natural language processing tasks, standard NLTK library has been used.

The final step in the Figure 1 is to split the data set of questions into training set and test set and to train classifiers using training data set and thereafter the trained classifier will be used to classify unlabeled questions in the test data set.

4. RESULTS AND DISCUSSION

In this study, a linear multi class Support Vector Machine (one-against-one) and binary term weighting were used. In order to train and validate the classifier, 10-fold cross validation method has been deployed. The next analysis and discussion will begin with the accuracy of SVM classifier based on several types of features in multi data set, followed by a comparison between accuracy of SVM classifier and number of features based on several types of features in multi data set. The same analysis aforementioned will be repeated for single data set.

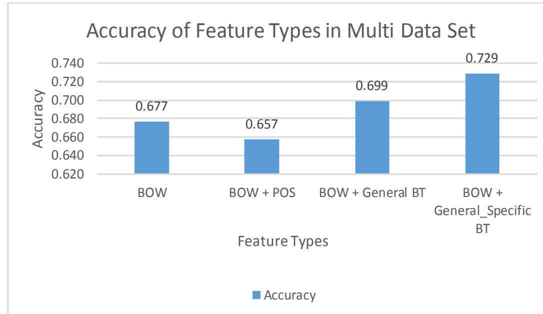

Figure 2: Accuracy of feature types in Multi Data Set

For the sake of simplicity, the following abbreviations will be used for feature types.

BOW → BW
BOW + POS → BWP
BOW + General Taxonomy → BWG
BOW + General_Specific Taxonomy → BWGS

Based on figure 2, all accuracies are greater than 0.6 but less than 0.8. And BWGS features perform better compared to all the other feature types. The second is BWG with the accuracy of 0.699, the third is BOW with the accuracy of 0.677 and the fourth is BWP (0.657). In this study, the accuracy of classifier using BWP feature recorded poor with 3% difference in accuracy compared with base line feature type which is BW. Certainly, the inclusion of class label features (BWG/BWGS) have performed better than the BW. Despite the improved accuracy hovers around 8% between BW and BWGS, a higher percentage in accuracy can be expected with a larger data set. The difference in accuracy between BWG and BWGS is not as substantial as between BW and BWGS as the former affects all the questions in the data set while the latter only affects selected questions which belong to Apply and Analysis BT levels. Nevertheless, BWG and BWGS features reflects the importance of taxonomy features to improve the accuracy in classifying BT questions. The base line feature is BW which has been used in the past research work.

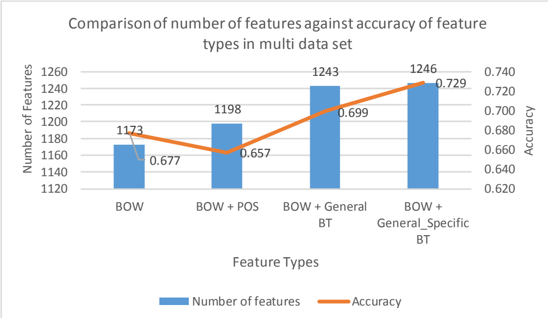

Figure 3: Comparison Of Number Of Features Against The Accuracy Of Feature Type In Multi Data Set

Based on figure 3, it shows the number of features obtained against each type of feature. For example, in BW, there are 1173 features while in BWP is 1198. As the total number of features increases, generally, the performance of the classifier improves gradually from BW to BWGS but not from BW to BWP. This shows that extracting relevant features that have discriminating power in classifying questions is more important. Meanwhile, a small amount of increase in features (around 6%) between BW and BWG can result in an increase in performance of classifier when the extracted features have some discriminating power. The outcome of figure 3, clearly indicates that the inclusion of taxonomy features can have positive impact on the performance of the classifier with a small amount of increase in taxonomy features.

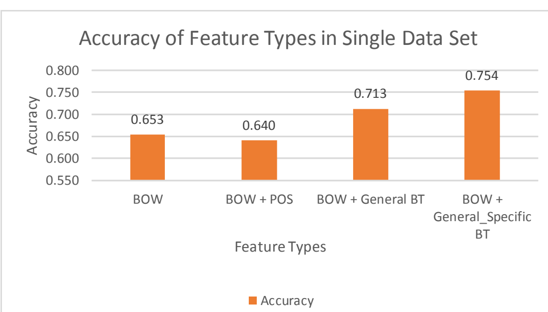

Figure 4: Accuracy Of Feature Types In Single Data Set

Based on figure 4, the accuracies of all feature types take the same shape as the accuracies of feature types in multi data set. However, the magnitude of increase in accuracy between BW and BWGS is 15% which is certainly higher than that of in multi data set. This shows that the taxonomy features also work reasonably well in the single data set. Also, the difference in accuracy between BWG and BWGS recorded (6%) higher than that of in multi data set. This could be due to the ratio of occurrence of specific taxonomy features to overall features in single data set is higher than that of in multi data set. As with BWP in multi data set, the accuracy declines from BW and BWP. A larger single data set and using term weighting TF-DIF may record a positive accuracy when including POS tagger feature type with unigram.

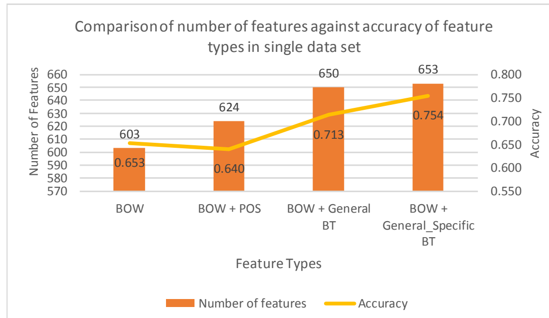

Figure 5: Comparison Of Number Of Features Against The Accuracy Of Feature Type In Single Data Set

Results from Figure 5 show that as the total number of features increases, the performance of the classifier improves gradually from BW to BWGS despite it takes dip from BW to BWP. The impact of an increase of features (BWG) around 8%, provides better discriminating power with the increase of accuracy of 9%. The outcome of figure 5, shows favorable result of extending BW with taxonomy feature type to increase the performance of classifier even in single data set. This is consistent with the results obtained using multi data set. Extending bag-of-words feature type with the proposed taxonomy based feature type shows an increase in accuracy in classifying questions for questions from both single and multi-data sets. However, the specific taxonomy feature type will play its part in accuracy of question classifier only when data sets contain ambiguous questions resulted from the existence of overlapping or shared BT keywords in questions.

5. CONCLUSION AND FUTURE WORK

Apparently, this study investigates the effectiveness of taxonomy based features in respect of accuracy of classifier. The outcome of preliminary study is promising where inclusion of class features to existing base line feature type (bag-of-word) has the potential to improve the accuracy of classifier in classifying questions of data sets of different fields or areas. Also, observed that the taxonomy feature type has the ability to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

improve the performance of classifier in classifying questions from the area/domain specific data set. The taxonomy based feature type also has the potential to address the issue of ambiguous questions resulted from the existence of overlapping or shared BT keyword where it can appear in more than level or category of BT. In short, the main contribution of this study is to extend existing feature type such as bag-of words with taxonomy based feature type to improve the accuracy of question classification for questions from specific area and various areas.

Future work would involve in comparing the performance of SVM using taxonomy feature type against other well-known classifiers. Besides that, a comparison between term weighting (binary) and TDFIF will be investigated with different feature types. A separate validate data set without class labels will be prepared to check the performance of classifier with unseen data.

REFERENCES:

[1] V. Gupta and G. S. Lehal, "A survey of text mining techniques and applications," *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.

[2] R. Agrawal and M. Batra, "A Detailed Study on Text Mining Techniques," *Int. J. Soft Comput. Eng.*, no. 26, pp. 2231–2307, 2013.

[3] J. a N. H. Kroeze, M. C. Matthee, and T. J. D. Bothma, "Differentiating Data- and Text-Mining Terminology," *Proc. 2003 Annu. Res. Conf. South African Inst. Comput. Sci. Inf. Technol. Enablement through Technol.*, vol. 6, no. December, pp. 93–101, 2003.

[4] A. Alahmadi, A. Joorabchi, and A. E. Mahdi, "A new text representation scheme combining Bag-of-Words and Bag-of-Concepts approaches for automatic text classification," in *2013 7th IEEE GCC Conference and Exhibition, GCC 2013*, 2013, pp. 108–113.

[5] L. Dan, L. Lihua, and Z. Zhaoxin, "Research of text categorization on Weka," in *Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications, ISDEA 2013*, 2013, pp. 1129–1131.

[6] H. Cao and H. Jia, "Tibetan text classification based on the feature of position weight," *Proc. - 2013 Int. Conf. Asian Lang. Process. IALP 2013*, pp. 220–223, 2013.

[7] N. Zechner, "The Past, Present and Future of Text Classification," in *2013 European Intelligence and Security Informatics Conference*, 2013, pp. 230–230.

[8] X. Guixian, G. Xu, Z. Xiaobing, and Y. Guosheng, "Automatic Classification of Uighur Web Pages," in *2013 Third International Conference on Intelligent System Design and Engineering Applications*, 2013, pp. 390–393.

[9] L. Xin and R. Dan, "Learning Question Classifiers," *COLING '02 Proc. 19th Int. Conf. Comput. Linguist.*, pp. 1–7, 2002.

[10] F. Colace, M. de Santo, L. Greco, and P. Napoletano, "Improving Text Retrieval Accuracy by Using a Minimal Relevance Feedback," *Commun. Comput. Inf. Sci.*, vol. 348, pp. 126–140, 2013.

[11] M. Mishra, V. K. Mishra, and S. H.R., "Question Classification using Semantic, Syntactic and Lexical features," *Int. J. Web Semant. Technol.*, vol. 4, no. 3, pp. 39–47, 2013.

[12] D. Zhang and W. S. Lee, "Question classification using support vector machines," *SIGIR '03 Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, pp. 26–32, 2003.

[13] D. Metzler, D. Metzler, W. B. Croft, and W. B. Croft, "Analysis of Statistical Question Classi cation for Fact-based Questions," *Inf. Retr. Boston.*, vol. 8, pp. 1–30, 2005.

[14] D. A. Abduljabbar and N. Omar, "EXAM QUESTIONS CLASSIFICATION BASED ON BLOOM ' S TAXONOMY COGNITIVE LEVEL USING CLASSIFIERS COMBINATION," *J. Theor. Appl. Inf. Technol.*, vol. 78, no. 3, pp. 447–455, 2015.

[15] A. A. Yahya, A. Osman, A. Taleb, and A. A. Alattab, "Analyzing the Cognitive Level of Classroom Questions Using Machine Learning Techniques," *Procedia - Soc. Behav. Sci.*, vol. 97, pp. 587–595, 2013.

[16] N. Yusof, "Determination of Bloom's Cognitive Level of Question Items using Artificial Neural Network," in *Information Systems*, 2010, pp. 866–870.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[17] W. Chang and M. Chung, "Automatic Applying Bloom's Taxonomy to Classify and Analysis the Cognition level of English Question Items," in *Pervasive Computing (JCPC), 2009 Joint Conferences*, 2009, pp. 727–734.

[18] S. S. Haris and N. Omar, "Determining Cognitive Category of Programming Question with Rule-based Approach," *Int. J. Inf. Process. Manag.*, vol. 4, no. 3, pp. 86–95, 2013.

[19] N. Omar and S. S. Haris, "A Rule- based Approach in Bloom ' s Taxonomy Question Classification through Natural Language Processing," in *7th International Conference on Computing and Convergence Technology (ICCCT), 2012*, 2012, pp. 410–414.

[20] A. Sangodiah, R. Ahmad, W. Fatimah, and W. Ahmad, "An Integration of Machine Learning Approach in Item Bank Test System," in *2016 3rd International Conference On Computer And Information Sciences (ICCOINS)*, 2016, pp. 180–184.

[21] A. A. Yahya and A. Osman, "Automatic Classification of Questions Into Bloom's Classification," in *Proceedings of the International Arab Conference on Information Technology*, 2011, pp. 1–6.

[22] A. A. Y. ADDIN OSMAN, "CLASSIFICATIONS of EXAM QUESTIONS USING LINGUISTICALLY- MOTIVATED FEATURES : A CASE STUDY BASED on BLOOM ' S TAXONOMY Research Questions Research Aim," in *The Sixth International Arab Conference on Quality Assurance in Higher Education*, 2016.

[23] S. F. Kusuma, D. Siahaan, and U. L. Yuhana, "Automatic Indonesia ' s Questions Classification Based On Bloom ' s Taxonomy Using Natural Language Processing," in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2015, p. 15.

[24] N. Omar, S. S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. F. A. Zainal, and R. Zulkifli, "Automated Analysis of Exam Questions According to Bloom's Taxonomy," *Procedia - Soc. Behav. Sci.*, vol. 59, no. 1956, pp. 297–303, 2012.

[25] S. Hasnah, S. Fattah, R. S. Sulong, and M. Mamat, "Mining Exam Question based on Bloom ' s Taxonomy," *https://www.semanticscholar.org/paper/Mining-Exam-Question-based-on-Bloom-s-Taxonomy-Tanalol-Fattah/97c252f2124d554c5fe5732cb8eb2047a23e8558*, 2008. .

[26] M. Dubey, "Classifying Stack Overflow Questions Based on Bloom's Taxonomy," 2016.

[27] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification," *Inf. Process. Manag.*, vol. 50, no. 1, pp. 104–112, 2014.

[28] G. Salton, a Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Mag. Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[29] T. Joachims, "A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization," *14th Int. Conf. Mach. Learn. (ICML '97)*, pp. 143–151, 1997.

[30] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic P. 2008.

[31] T. Miner, "Text Mining Online." [Online]. Available: http://textminingonline.com/tag/wordnet-lemmatizer.

[32] E. A. Toader, "USING STANFORD PARSER METHOD FOR ASSESSING THE COMPETENCIES OF IT PROFESSIONALS," no. April, pp. 1099–1109, 2015.

[33] B. S. Bloom, *Taxonomy of Educational Objectives Book 1: Cognitive Domain*. Addison Wesley Publishing Company, 1984.

[34] K. M. Boon and L. T. Lim, "An Examination Question Paper Preparation System with Content-Style Separation and Bloom's Taxonomy Categorisation," in *The Third International Conference on E-Learning and E-Technologies in Education (ICEEE2014)*, 2014, pp. 39–47.

[35] B. et Al, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[36] "PyCharm IDE," 2016. [Online]. Available: https://www.jetbrains.com/.

[37] B. Lorica, "Six reasons why I recommend scikit-learn," 2015. [Online]. Available: https://www.oreilly.com/ideas/six-reasons-why-i-recommend-scikit-learn.