Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

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

QUESTION CLASSIFICATION USING SUPPORT VECTOR MACHINE AND PATTERN MATCHING

[1]ALI MUTTALEB HASAN, [2]LAILATUL QADRI ZAKARIA
[1,2] Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
E-mail: [1]alimatlab65@yahoo.com, [2]lailatul.qadri@ukm.edu.my

ABSTRACT

Question classification plays a crucial role in the question answering system, and it aim to accurately assign one or more labels to question based on expected answer type. Nonetheless, classifying user's question is a very challenging task due to the flexibility of Natural Language where a question can be written in many different forms and information within the sentence may not be enough to effectively to classify the question. Limited researches have focused on question classification for Arabic question answering. In this research we used support vector machine (SVM) and pattern matching to classify question into three main classes which are "Who", "Where" and "What". The SVM leverage features such as n-gram and WordNet. The WordNet is used to map words in questions to their synonyms that have the same meaning. Five pattern were introduced to analyze "What" question and label the questions with "definition", "person", "location" or "object". The dataset set used in this research consist of 200 question about Hadith from Sahih Al Bukhari. The experimental result scored F-measure at 95.2%, 84.6%, and 83.6% respectively for "Who", "Where" and "What". The result show that the SVM classifier is useful to classify question in Arabic language.

Keywords: *Question Classification System; Machine Learning.*

1. INTRODUCTION AND RESEARCH BACKGROUND

Question-Answering (QA) System is an information retrieval system that takes the search queries in the form of natural language questions and returns results in the form of a short answer that is appropriate to the question asked. Automatic Question Answering (QA) has received researchers' attention over the last contract. It aims to find a short answer string which exactly addresses the users' questions instead of than overwhelm them with a large number of retrieved documents, which they must sort through to find the required answer. This goal is usually achieved using a combination of different techniques, including information retrieval, information extraction, and natural language processing opinion mining.

Question classification has a key role in automated question answering systems Although different types of question answering systems have different architectures, most of them follow a framework in which question classification plays an important role [1]. Furthermore, it has been shown that the performance of question classification has significant influence on the overall performance of a question answering system [2] [3] [4].

In fact, classifying users' questions in a question answering is very challenging problem [5]. The language flexibility where different users can write or ask about the same information in many different questions. For instance, the two questions: (Where did the ditch battle happen?) and (What is the name of mountain that happened the battle of Badr?), both have the same answer type. A central challenge of computational linguistics is to decide if these pieces of texts have similar meanings and requesting the same type of answer [6]. The second issue is the ambiguity of questions where the same keywords can have different answers [7]. For instance, what eat fish" and what fish eat", although these two questions uses the same keywords, they are different and require different information. The third issue is that question classification is challenging as compared

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

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

to other text classification tasks such as document classification as in question classification there exist only little information which may not be enough to effectively classify questions [8]. Work on other languages is still growing. Here are basically two different approaches for question classification: rule-based and learning based.

In this research, a question classification models designed for Arabic Hadith text. The overall objective of this paper is to discuss question classification approach by using support vector machine learning and pattern matching. The remainder of this paper is organized as follows. Section 2 present the methodology as well as the classification modules and feature extraction methods. In Section 3, we present the experiment setup and discuss the experimental results. Finally, we conclude our work and discuss future directions of research in Section 4.

2. RELATED WORK

Abdelnasser [9] design a question classification system for Quran. They used support vector machine as base classifier. they utilized several features number of words and number of named entities the data set consist of 230 classified question from Quran dataset.

Musa [10] used to improve the popularization performance of single the learning algorithm. Therefore, comparing classification performance between SVM and LR using bagging and ensemble is an interesting issue.

Al Chalabi [11] uses the linguistic patterns as the main features in the question classification task in order to identify the candidate answers types. These patterns are helpful in matching, parsing and identifying the candidate answer sentences. For instance, consider the question "ماهو الاسلام؟" (What is Islam?). The question classification process predicts this question as "Definition" question, and creates the searching patterns for this question.

3. METHODOLOGY

This section presents the methodology which has been used in designing our question classification model for Arabic Hadith text. This model uses pre-processing tasks to eliminate the incomplete, noisy and inconsistent data. Next, this model performs text representation and feature weighting. Each question from the raw data is represented as a row of values for each feature from predefined set of features. After that the model used this output as an input for the Support vector machine to classify the question. Figure 3.1 shows the overall architecture of this model models. As shown in the figure, the model involves the following phases:

- Data pre-processing phase
- Text representation and features weighting
- Question Classification phase.
- Evaluation phase.

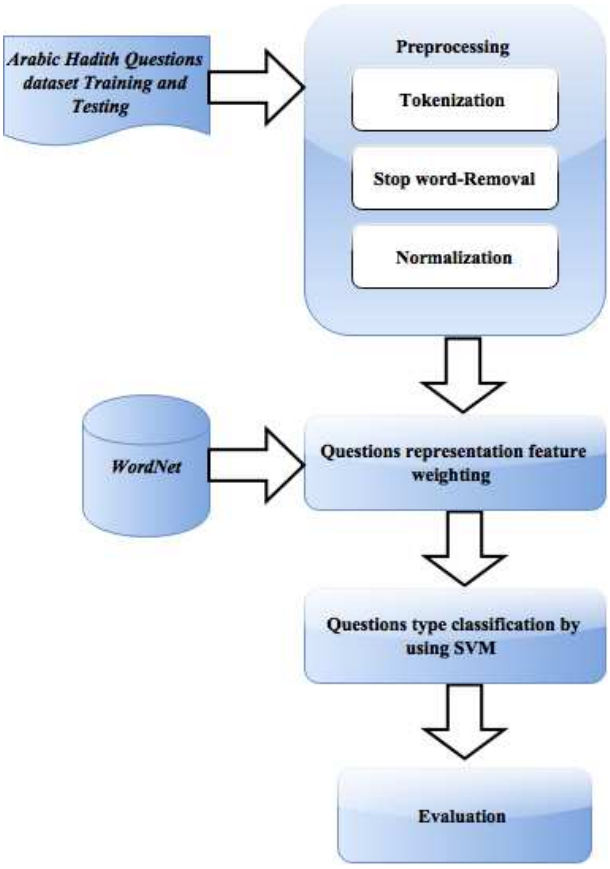

Figure 1. The Question Classification In Hadith - Domain Question Answering Architecture

3.1 Pre-processing

The pre-processing phases is an essential phase in designing a question answering system and it is also crucial to pre-process the data with machine learning approaches. In question answering both users' queries and hadith documents are passed through the pre-processing phase. The pre-processing phases plays an important role in the question answering system for many reasons. It

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

leads to the significant reduction of the space of words and processing time. It removes noisy words, which frequently occur in both hadiths and questions texts, such as prepositions, conjunctions and articles. Data pre-processing comprises three steps: normalization, tokenization, and stop word removal. All of the reviews involve a pre-processing stage. In the normalization process, diacritics are removed.

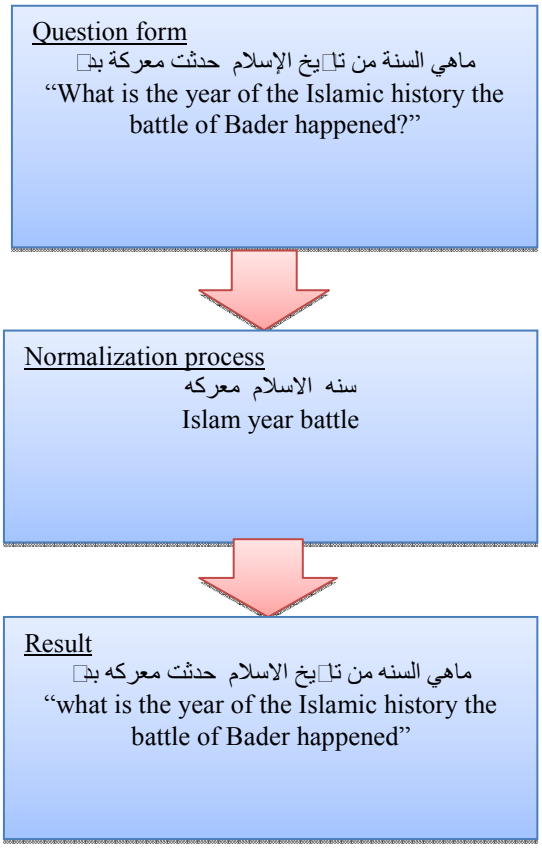

Figure 2. Normalization Process

3.2 Questions Representation and Feature Weighting

Before any classification task, one of the most essential tasks that need to be accomplished is that of text representation and feature weighting. In Question Classification in Hadith -domain question answering, a question is described by a vector of features (terms) and feature values, also called attributes and attribute values. In this his step, the Arabic Word Net are used to expand the question which has the same meaning. A simple mapping is used where each word and their synonyms are converted to single entry. For example, "معركة" in all training and testing questions are expanded with the Arabic words" موقعة" (battle), "معركة" combat, and" غزوة" (foray).

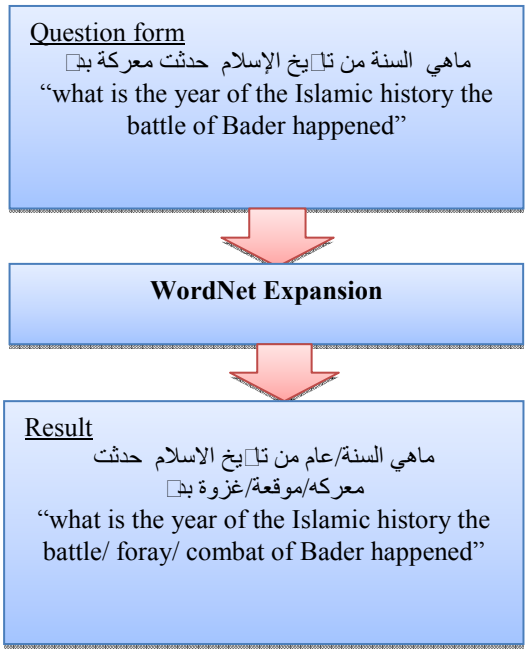

Figure 3. Wordnet Expansion

In addition, after pre-processing, we then assign the weight of each word, by calculated using the TF.IDF Term Frequency (TF) weighting is also recognized as a simple method for term weighting. TF.IDF weighting approach gives weight to the frequency of a term in a document with a factor discounting its importance in case when the appearance of it is found in most of the documents.

$$w_{i=TF_i}.\log(\frac{N}{n})$$

For example, the question "<What is the the year of the Islamic history the battle of Bader happened?)" Is represented in the following form

$$x = \{(What, 1)(year, 1)(Islamic, 1)(history, 1)(battle, 1)(bader, 1)(happened, 1)\}$$

Unigrams is a special case of the so-called n-gram features. To extract n-gram features, any n consecutive words in a question is considered as a feature. The bigram feature "what-name" can be viewed as a new term and the bigram pair {(what-name, 1)} will be added to the feature vector when bigram features are extracted.

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

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

$$bigram\ features = x =$$

$$\{(What_{name}, 1)(name_{Islamic}, 1)$$
$$(Islamic_{history}, 1), \ldots\ldots\}$$

$$trigram\ features = x$$

$$= \{(What_{name_{Islamic}}, 1)$$
$$(name - Islamic - history, 1)..\}$$

We found that considering only the first two or three words of a question as trigram features, performs as good. For example, consider the question "What is the name of the mountain at which Allah talked to Moses?" The trigram "what name_ mountain" is very useful in question classification.

3.4 Questions Representation and Feature Weighting

The question classification process is used to classify unseen questions. The SVM classifier is applied to the question vector is used for classification. The SVM classifier build a training model based on the feature vector of the training data from the three type of questions. Then the SVM classifier take as input the feature vectors of the new unseen questions (Ts1, Ts2…… Ts20).

SVM Classifier. A SVM is a relatively new class of machine learning techniques that was first introduced by. SVMs are a very popular technique for text categorization used in the machine learning community. They are considered to be one of the most effective classification methods according to their performance on text classification, as confirmed by many researchers.

Based on the structural risk minimization principle from computational learning theory, SVMs seek a decision surface to separating the training data points into two classes and to make decisions based on the support vectors that are selected as the only effective elements in the training set.

Multiple variants of SVMs have been developed. In this paper, our discussion is limited to linear SVMs due to their popularity and high performance in text categorization. The optimization procedure of SVMs (dual form) is to minimize the following:

$$\vec{\alpha} = \arg\min\left\{-\sum_{i=1}^{n}\alpha_i + \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \left\langle \vec{x_i}, \vec{x_j} \right\rangle\right\}$$

$$subject\ to: \sum_{i=1}^{n}\alpha_i y_i = 0;\ 0 \le \alpha_i \le C$$

3.3 Pattern matching

There are some questions inherently do not have any head word. For example, for the question "What is Islam?" there is no suitable head word as the entity type of the only noun in this question (Islam) does not contribute to classify this question as "type1: concept definition". We have also implemented 5 patterns-rules as shown Table 1 to investigate the contribution of this feature space in our classifier.

Table 1. The Developed Patterns-Rules

Name	Name Pattern
Type: definition pattern 1	The question begins with a meaning in English is "what" is/are { هوor {هم and follows by an optional a, an { لor {ع} or the and then follows by one or two words "What is Islam" {ماهو الاسلام}
Type: definition pattern 2	The question begins with what do/does {فعل or {هل and ends with mean "what do we means by Iman" {ماذا يعني□حن كتبها ايمان}
Type: ENTY: location pattern 3	The question begins with what and follows by one or two words contains location keyword "<What is the name of mountain that happened the battle of Badr?)" ما هو اسم الجبل الذي حدث في غزوة بد□ } }
Type: ENTY: person pattern 4	The question begins with what and follows by many words contains person keyword "<What is the name of the first prophet's wife's name?)" { ماهو اسم اول زوجة الرسول}
Type: ENTY: object pattern 5	The question begins with what and follows and follows by three to four words contains object keyword "What was the first mosque built in the name of Islam" { ماهو أول مسجد بني في الاسلام}

The pattern name can be viewed as a term and the feature value would be 1 if the question match with that pattern. For example, for the question

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

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

"What is Islam?" its head rule feature can be represented as follows:

$$\{(type:def-pattern,1)\}.$$

3.5 Experimental Setup

We conduct several experiments to evaluate our model. We evaluate the performance of the classification algorithms. We manually annotated 200 relevant Hadith-domain questions, classifying them into different categories. The data used in this research was collected based on Arabic books of Hadith. In this study using the cross-validation process, the corpus was randomly partitioned into 10 equal subsamples. A single subsample (20 questions) was retained as the validation data for testing the model, and the remaining 9 subsamples were used as the training data (180 questions). The cross-validation process was then repeated 10 times (10 folds). To measure the performance of the classification method, experimental results are sorted into the following: True Positive (TP) is the set of reviews that is correctly assigned to the given category, False Positive (FP) is the set of reviews that is incorrectly assigned to the category, False Negative (FN) is the set of reviews that is incorrectly not assigned to the category, and True Negative (TN) is the set of the set of reviews that is correctly not assigned to the category. However, we use the F1 and Macro-F1 measures. The following describes these metrics:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

$$\text{F1- measures} = \frac{2*\text{Re}call*\text{Pr}ecision}{\text{Re}call+\text{Pr}ecision}$$

4. RESULTS AND DISCUSSION FOR CLASSIFIER

To examine the SVM classifier's overall performance with the 1-Gram features. The first set of experiments is to evaluate the usefulness of 1-Gram feature in question classification. The performance of this experiment provides a base line for other experiments. Table 2 and Table 3 shows detailed results in term of evaluation matrices precision, recall and F-measure which are obtained by the Support vector machine with 1-gram and for each question type when (TF) and (TF×IDF) weighting methods are used respectively. These results indicate that the support vector machine classifier when using 1-gram feature and TFIDF weighting method is a good machine learning technique for Question Classification in Hadith - domain question answering.

Table 2 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (1-GRAM, TF Weighting) For Question Classification.

	Precision	**Recall**	**F-measure**
TYPE1	96.25	84.62	90.06
TYPE2	80	66.67	72.73
TYPE3	80	94.12	86.49
AVG	**85.42**	**81.8**	**83.09**

Table 3 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (1-GRAM, TF×IDF Weighting) For Question Classification.

	Precision	Recall	F-measure
TYPE1	92.5	83.15	87.57
TYPE2	85	65.38	73.91
TYPE3	79	92.94	85.41
AVG	85.5	80.49	82.3

The second set of experiments is to evaluate the usefulness of 2-Gram feature in question classification. Table 3 and Table 5 shows detailed results in term of evaluation matrices precision, recall and F-measure which are obtained by the Support vector machine with 2-gram and for each question type when (TF) and (TF×IDF) weighting methods are used respectively. As shown Table 3 and Table 5, the use of 2-gram feature give a significant improvement over the baseline models.

Table 4 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (2-GRAM, TF Weighting) For Question Classification

	Precision	**Recall**	**F-measure**
TYPE1	93.75	87.21	90.36
TYPE2	80	69.57	74.42
TYPE3	84	92.31	87.96
AVG	**85.92**	**83.03**	**84.25**

Table 5 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (2-GRAM, TF×IDF Weighting) For Question Classification.

	Precision	**Recall**	**F-measure**
TYPE1	93.75	87.21	90.36
TYPE2	75	88.24	81.08
TYPE3	89	91.75	90.36
AVG	**85.92**	**89.07**	**87.27**

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

The third set of experiments is to evaluate the usefulness of 3-Gram feature in question classification. In addition, Table 6 and Table 6 shows detailed results in term of evaluation matrices precision, recall and F-measure which are obtained by the support vector machine with 3-gram for each question type when (TF) and (TF×IDF) weighting methods are used respectively. As shown Table 6 and Table 7, the use of 3-gram feature give a significant improvement over the baseline models for each question type. These results indicate that the classifier Support vector machine classifier with 2-Gram Feature are most suitable technique for Question Classification in Hadith -domain question answering.

Table 6 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (3-GRAM, TF Weighting) For Question Classification.

	Precision	Recall	F-measure
TYPE1	97.78	88	92.63
TYPE2	59.09	92.86	72.22
TYPE3	89	90.82	89.9
AVG	**81.96**	**90.56**	**84.92**

Table 7 The Performance (Precision, Recall And F-Measure) For Each Question Type Obtained By Support Vector Machine (3-GRAM, TF×IDF Weighting) For Question Classification

	Precision	Recall	F-measure
TYPE1	95	90.48	92.68
TYPE2	65	86.67	74.29
TYPE3	92	91.09	91.54
AVG	**84**	**89.41**	**86.17**

5. CONCLUSION

SVM and Pattern matching help to improve question classification for Question Classification in Hadith-domain question answering. This paper presented several experiments to empirically compare the effectiveness different sets of features (bag-of-word features (i.e., 1-Gram), 2-grams and 3-grams) on the performance of support vector machine classification approach for Question Classification in Hadith-domain question answering. In addition, several experiments are conducted to study the effect of the different feature weighting methods (term frequency (TF) or Term Frequency × Inverse Document Frequency (TF×IDF)) on the performance of the classification approach. Finally, the experimental results demonstrate that the use of large n-gram feature set give a significant improvement over the baseline model. In addition, results also indicate that the results of support vector machine obtained using 2-gram feature method outperformed that obtained using other n-gram features for Question Classification in Hadith -domain. Finally, the experiments indicate that the SVM classifier with the 2-gram feature yields the best classification method, with a performance of 87.25%.

ACKNOWLEDGMENT

I would like to express my deepest and sincere gratitude to my supervisor, Dr. Lailatul-Qadri Zakaria not only for her immeasurable wisdom, patience, valuable suggestions but for also ensuring that my dream of becoming a Master degree holder is not shattered. You are cherished, I would like to thank my own National University of Malaysia, University Kebangsaan Malaysia for supporting this research.

REFRENCES:

[1] E. M. Voorhees, "The TREC question answering track," Natural Language Engineering, vol. 7, pp. 361-378, 2001.

[2] A. Ittycheriah, M. Franz, and S. Roukos, "IBM's Statistical Question Answering System-TREC-10," in TREC, 2001.

[3] E. Hovy, L. Gerber, U. Hermjakob, C.-Y. Lin, and D. Ravichandran, "Toward semantics-based answer pinpointing," in Proceedings of the first international conference on Human language technology research, 2001, pp. 1-7.

[4] D. Moldovan, M. Paşca, S. Harabagiu, and M. Surdeanu, "Performance issues and error analysis in an open-domain question answering system," ACM Transactions on Information Systems (TOIS), vol. 21, pp. 133-154, 2003.

[5] V. Punyakanok, D. Roth, and W.-t. Yih, "Natural language inference via dependency tree mapping: An application to question answering," 2004.

[6] C. Unger, L. Bühmann, J. Lehmann, A.-C. Ngonga Ngomo, D. Gerber, and P. Cimiano, "Template-based question answering over RDF data," in Proceedings of the 21st international conference on World Wide Web, 2012, pp. 639-648.

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

[7] L. Merhbene, A. Zouaghi, and M. Zrigu, "Ambiguous Arabic words disambiguation," in Software Engineering Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD), 2010 11th ACIS International Conference on, 2010, pp. 157-164.

[8] V. Lopez, V. Uren, E. Motta, and M. Pasin, "AquaLog: An ontology-driven question answering system for organizational semantic intranets," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 5, pp. 72-105, 2007.

[9] H. Abdelnasser, R. Mohamed, M. Ragab, A. Mohamed, B. Farouk, N. El-Makky, et al., "Al-Bayan: An Arabic Question Answering System for the Holy Quran," ANLP 2014, p. 57, 2014.

[10] A. B. Musa, "Comparative study on classification performance between support vector machine and logistic regression," International Journal of Machine Learning and Cybernetics, vol. 4, pp. 13-24, 2013.

[11] H. M. Al Chalabi, "Question Processing for Arabic Question Answering System," 2015.