The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 4 (2018) 239–242

www.elsevier.com/locate/icte

Classification of factoid questions intent using grammatical features

Alaa Mohasseb*, Mohamed Bader-El-Den, Mihaela Cocea

School of Computing, University of Portsmouth, United Kingdom

Received 23 September 2018; accepted 16 October 2018
Available online 3 November 2018

Abstract

In question-answering systems, question classification is a fundamental task. Identifying the accurate question type enhances the retrieval of more accurate answers. Factoid questions are the most challenging type of question to classify in which many approaches have been proposed with the objective of enhancing the classification of this type of question. In this paper, a grammar-based framework is used. The framework makes use of three main features which are, grammatical features, domain specific features and patterns. Using machine learning algorithms for the classification process, experimental results show that our approach has a good level of accuracy.

Keywords: Information retrieval; Question classification; Factoid questions; Grammatical features; Machine learning

1. Introduction

In question and answering systems (QASs), questions classification is a fundamental task. Identifying the accurate question type enhances the retrieval of more accurate answers. However, the continuing growth of the amount of web content makes the retrieval of relevant answers difficult. Factoid questions are the most challenging type of question to classify. Various approaches have been proposed with the objective of enhancing the identification and the classification of factoid questions; most of these are approaches based on semantic features and bag-of-words. Several question taxonomies have been proposed [1–5]. The most popular classification taxonomy of factoid ('wh-') questions is Li and Roth's categories [5]. Many researchers focused on Li and Roth classification of question [6–17]. Their two-layer taxonomy consists of a set of six coarse-grained categories which are Abbreviation, Entity, Description, Human, Location and Numeric value, and fine-grained classes such as Expression, Manner, Color, Event and City.

The classification of the questions performed in QASs directly affects the answers. Authors in [18] stated that most errors happen due to miss-classification of questions performed in QASs in which the task of generating answers to the users questions is directly related to the type of questions asked.

Classifying 'wh-' questions into proper semantic categories is found more challenging than classifying other types in question answering systems [17]. In addition, features are the key to obtain an accurate question classifier and linguistic features play an important role in developing an accurate question classifier [10]; recent studies classified users' questions using different features like bag-of-words [12,6,13] and uni-gram and word shape features [14]. Moreover, authors in [3] integrated pattern matching and machine learning techniques for the classification of questions, while [19] classified questions by their expected types of responses. According to [2] a question type is defined as a certain semantic category and is characterized by common properties.

Furthermore, machine learning algorithms have been used by many previous studies for the classification of questions. Support Vector Machine (SVM) is one of the most used algorithms [7,4,14,20,8,9]. Combining different features such as syntactic, lexical and semantic attributes with a SVM classifier improves the classification accuracy [13]. Other works like [12]

* Corresponding author.
E-mail addresses: alaa.mohasseb@port.ac.uk (A. Mohasseb), mohamed.bader@port.ac.uk (M. Bader-El-Den), mihaela.cocea@port.ac.uk (M. Cocea).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2018.10.004

and [13] used other machine learning algorithms besides SVM such as Naive Bayes, Nearest Neighbors and Decision Tree.

In a previous study [1], a Grammar-based framework for Questions Categorization and Classification (GQCC) was proposed. In this study, the framework is applied to question classification according to Li and Roth's [5] categories of intent, using three main features which are: (1) Grammatical features (2) Domain specific features and (3) Patterns. These features transfer the question into a new representation which has the advantage of preserving the grammatical structure of the question. The aim is to assess the influence of using the structure of the question and the domain-specific syntactic categories and features on the classification performance. The rest of the paper is organized as follows: Section 2 outlines the previous work in question classification. Section 3 describes the proposed approach and the grammatical features used. The experimental setup and results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Question classification methods

In this section, we review related work about question classification methods and machine learning algorithms using Li and Roth's question categories.

Authors in [6] used compositive statistic and rule classifiers combined with different classifiers and multiple classifier combination methods. In [8] a method was proposed using feature selection algorithm to determine appropriate features corresponding to different question types. While authors in [7] proposed a statistical classifier which is based on SVM. Furthermore, a SVM-based approach for question classification was proposed in [9].

In [13] question classification method was proposed using three different classifiers, k-Nearest Neighbor (kNN), Nave Bayes (NB), and SVM. In addition, features such as using bag-of-words and bag-of-ngrams were used and a set of lexical, syntactic, and semantic features were also used. Authors in [12] used five machine learning algorithms which are, KNN, NB, Decision Tree (DT), Sparse Network of Winnows (SNoW), and SVM. In addition, two features were used; bag-of-words and bag-of-ngrams. Moreover, in [14] a head word feature was proposed and two approaches were presented to augment semantic features of such head words using WordNet. Moreover, authors in [10] proposed a compact feature set that uses typed dependencies as semantic features.

In [15] authors used unlabeled questions in combination with labeled questions for semi-supervised learning. In addition, Tri-training were selected to improve the precision of question classification task. In addition, a two-level hierarchical classifier for question classification was proposed in [16]. The proposed classifier classifies the question sequentially two times by a coarse classifier and one of the six fine classifiers. Moreover, different machine learning algorithms were used for the coarse classifier and fine classifiers such as supervised and semi-supervised learning. Finally, in [17] authors classified what-type questions by head noun tagging. In addition, different features such as local syntactic feature, semantic feature and category dependency were integrated.

3. Proposed approach

3.1. Factoid questions grammatical features

This analysis was first introduced in [1]. Wh-questions (factoid) have its own characteristics, features, and structure that help in the identification and the classification process.

The main feature of a factoid question (Wh-Questions) is the presence of question words, this kind of question starts with a question word, such as *What, Where, Why, Who, Whose, When, Which*, e.g. *"What did the only repealed amendment to the U.S. Constitution deal with ?"*. In addition, this question could start with question words that do not start with "wh" such as *how, how many, how often, how far, how much, how long, how old*, e.g. *"How long does it take light to reach the Earth from the Sun?"*

In addition, the structure of this type of question could begin with a Preposition followed by a question, *"P + QW"* such as *"In what year did Thatcher become prime minister?" OR "At what age did Rossini stop writing opera?"*. Also in many cases the question word could be found in the middle of the question, e.g. *"The corpus callosum is in what part of the body?"*.

Most factoid questions are related to facts, current events, ideas and suggestions and could formulate an advice question, e.g. *"How do you make a paintball ?"*. In addition, some factoid questions could contain two types of question words, for example *"What does extended definition mean and how would one write a paper on it ?"*. Furthermore, factoid questions could have any kind of information given as an answer or response.

3.2. Question classification features

Three main features have been used for question analysis and classification which are, (1) Grammatical Features, (2) Domain specific Features and (3) Grammatical Pattern Features, these features transfer the question into domain specific grammatical pattern in which this new representation has the advantage of preserving the grammatical structure of the question.

3.2.1. Grammatical features

Grammatical Features have been used for the purpose of transforming the questions (by using the grammar) into a new representation as a series of grammatical terms, i.e. a grammatical pattern. The grammatical features consist of the seven major word classes in English, which are Verb, Noun, Determiner, Adjective , Adverb , Preposition and Conjunction. In addition, we added a category for question words that contains the six main question words: "how", "who", "when", "where", "what" and "which". Some word classes like Noun can have subclasses, such as Common Nouns, Proper Nouns, Pronouns and Numeral Nouns as well as Verbs, such as Action Verbs, Linking Verbs and Auxiliary Verbs. In addition, it consists of other features such as singular and plural terms.

3.2.2. Domain specific grammatical features

Domain-specific features (i.e. related to question-answering) were identified, which correspond to topics. Instead of further classifying the question to fine grained which is based on a large number of categories, we have used domain specific features to determine the type of question. For example, question type *ENTY* consists of fine grained categories such as religion, disease/medicine, event, product. These types could be identified using the following domain specific grammatical features: religion = religious terms PN_R, disease/medicine = health terms CN_{HLT} and PN_{HLT}, product = Products PN_P, event = events PN_E. Hence the domain specific grammatical features contain less categories than the fine grained categories proposed by Li and Roth but still could identify the different coarse categories.

3.2.3. Grammatical patterns

The question grammatical pattern help in the final identification of the question type, each factoid question type has a certain structure. For example, the following question which represent (HUM) type of question *"Who killed Gandhi?"* has the following grammatical pattern $QW_{Who} + AV + PN_C$. While, the question which represent (LOC) type of question *"What is the smallest country in Africa?"* has the following grammatical pattern $QW_{What} + LV + D + Adj + CN_{OS} + P + PN_G$. The different pattern representation helps in distinguishing between different factoid question type.

A full description of how these features are used is provided in the following sections

3.3. Question classification framework

A Grammar-based framework for Questions Categorization and Classification (GQCC) is used [1]. The question classification framework takes into account the grammatical structure of the questions and combines grammatical features with domain-related information and grammatical patterns. The framework consists of three main phases; (1) Question Parsing and Tagging, (2) Pattern Formulation and (3) Question Classification. The following question from Li and Roth datasets will be used *"What causes asthma?"* to illustrate how these phases work.

(1) **Question Parsing and Tagging:** this step is mainly responsible for extracting users question terms. The system simply takes the question and parses to tag each term in the question to its terms' category. In this phase parsing the keywords and phrases is done by; first parsed compound words then single words. In addition, the term tagging is done by tagging each term to its grammar terminals; each term will be tagged to its highest level of abstraction (domain specific).

For the given example the question will be parsed and tagged as follows:

Question: "What causes asthma?"

Terms extracted: What, causes, asthma

After parsing, each term in the question will be tagged to one of the terms category using tag-set that was proposed by [21] and [1]. The final tagging will be:

Question Terms Tagging: What=QW_{What} , causes= AV, asthma= CN_{HLT}

Table 1
Data distribution.

Question type	Total number of questions
ABBR	45
DESC	655
ENTY	710
HUM	655
LOC	457
NUM	478

(2) **Pattern Formulation:** in this phase after parsing and tagging each term in the question, the pattern is formulated. This is done by matching the question with the most appropriate question pattern to help facilitate the classification processing and the identification of the factoid question type in the next phase.

For the given example, the following pattern will be formulated:

Question Pattern: $QW_{What} + AV + CN_{HLT}$

(3) **Question Classification:** This phase is done by using the patterns generated in Phase (2), the aim of this phase is to build a model for automatic classification. The classification is done by following the standard process for machine learning, which involves the splitting of the dataset into a training and a testing dataset. The training dataset is used for building the model, and the test dataset is used to evaluate the performance of the model.

For the given example, the question will be classified to the following question type.

Question Type: DESC

4. Experimental study and results

In the experimental study we investigate the ability of machine learning classifiers to distinguish between different question types based on grammatical features and question patterns. Two machine learning algorithms, were used for question classification; Support Vector Machine (SVM) and J48. We used 3000 questions that were selected from Li and Roth.[1] Their distribution is given in Table 1. Questions in the dataset are classified into two categories; coarse and fine, in this experiment coarse categories have been used.

To assess the performance of proposed features and the machine learning classifiers experiments have been conducted using the Weka[2] software [22] The experiments were set up using the typical 10-fold cross validation.

4.1. Results

In this section we present and analyze the results of the machine learning algorithms. Table 2 shows the accuracy for GQCC based classifiers.

Table 3 presents the classification performance details (Precision, Recall and F-Measure) of the classifiers that have used SVM and J48 using the proposed grammatical features.

[1] http://cogcomp.org/Data/QA/QC/.
[2] http://www.cs.waikato.ac.nz/ml/weka/.

Table 2
Accuracy of GQCC based classifiers.

Classifiers	Accuracy
$GQCC_{SVM}$	95.5%
$GQCC_{J48}$	95.8%

Table 3
Performance of the classifiers — Best results are highlighted in bold.

Class:	$GQCC_{SVM}$			$GQCC_{J48}$		
	P	R	F	P	R	F
ABBR	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
DESC	**0.998**	0.998	0.998	**0.998**	**1.000**	**0.999**
ENTY	0.917	**0.920**	0.918	**0.937**	**0.920**	**0.928**
HUM	0.998	**0.998**	0.998	**1.000**	**0.998**	**0.999**
LOC	**0.861**	**0.908**	**0.884**	0.859	0.904	0.881
NUM	**0.987**	0.931	0.958	0.970	**0.948**	**0.959**

Results show that Decision Tree ($GQCC_{J48}$) identified correctly (i.e. Recall) 95.8% of the questions while $GQCC_{SVM}$ identified correctly 95.5% of the questions. Furthermore, when the classifiers were evaluated, both $GQCC_{J48}$ and $GQCC_{SVM}$ had nearly similar performance, both classifiers had (100%) recall for class type ABBR and similar recall for classes such as ENTY and HUM. However, $GQCC_{J48}$ has higher precision and f-measure for these classes. In addition, $GQCC_{SVM}$ has better performance for LOC class while for classes such as DESC and NUM $GQCC_{SVM}$ has higher precision and $GQCC_{J48}$ has higher recall and f-measure.

These results indicate that in terms of precision, recall and f-measure; $GQCC_{J48}$ had the better performance. In addition, the results validate that combining grammatical features and domain specific grammatical features improved the classification of these types and enable the machine learning algorithms to better differentiate between different class types.

5. Conclusion

In this paper, a Grammar-based framework for Questions Categorization and Classification (GQCC) was adapted. The framework make use of three main features which are, grammatical features, domain specific features and patterns. These features help in preserving the structure of the questions. In addition, the performance of different machine learning algorithms (J48 and SVM) was investigated for the classification of factoid questions. The results show that our solution led to a good performance in classifying questions.

Conflict of interest

The authors declare that there is no conflict of interest in this paper.

References

[1] A. Mohasseb, M. Bader-El-Den, M. Cocea, Question categorization and classification using grammar based approach, Information Processing and Management, 2018.

[2] O. Kolomiyets, M.-F. Moens, A survey on question answering technology from an information retrieval perspective, Inf. Sci. 181 (24) (2011) 5412–5434.

[3] F. Bu, X. Zhu, Y. Hao, X. Zhu, Function-based question classification for general qa, in: Proceedings of the 2010 conference on empirical methods in natural language processing, Association for Computational Linguistics, 2010, pp. 1119–1128.

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

[5] X. Li, D. Roth, Learning question classifiers: the role of semantic information, Nat. Lang. Eng. 12 (03) (2006) 229–249.

[6] X. Li, X.-J. Huang, L.-d. Wu, Question classification using multiple classifiers, in: Proceedings of the 5th Workshop on Asian Language Resources and First Symposium on Asian Language Resources Network, 2005.

[7] D. Metzler, W.B. Croft, Analysis of statistical question classification for fact-based questions, Inf. Retr. 8 (3) (2005) 481–504.

[8] N. Van-Tu, L. Anh-Cuong, Improving question classification by feature extraction and selection, Indian J. Sci. Technol. 9 (17) (2016).

[9] S. Xu, G. Cheng, F. Kong, Research on question classification for automatic question answering, in: Asian Language Processing (IALP), 2016 International Conference on, IEEE, 2016, pp. 218–221.

[10] P. Le-Hong, X.-H. Phan, T.-D. Nguyen, Using dependency analysis to improve question classification, in: Knowledge and Systems Engineering, Springer, 2015, pp. 653–665.

[11] R. May, A. Steinberg, Al, Building a Question Classifier for a Trec-Style Question Answering System, AL: The Stanford Natural Language Processing Group, Final Projects.

[12] D. Zhang, W.S. Lee, Question classification using support vector machines, in: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, ACM, 2003, pp. 26–32.

[13] M. Mishra, V.K. Mishra, H. Sharma, Question classification using semantic, syntactic and lexical features, Int. J. Web Semant. Technol. 4 (3) (2013) 39.

[14] Z. Huang, M. Thint, Z. Qin, Question classification using head words and their hypernyms, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, 2008, pp. 927–936.

[15] T.T. Nguyen, L.M. Nguyen, A. Shimazu, Using Semi-Supervised Learning for Question Classification, vol. 3, Information and Media Technologies Editorial Board, 2008, pp. 112–130.

[16] T.T. Nguyen, L.M. Nguyen, Improving the accuracy of question classification with machine learning, in: Research, Innovation and Vision for the Future, 2007 IEEE International Conference on, IEEE, 2007, pp. 234–241.

[17] F. Li, X. Zhang, J. Yuan, X. Zhu, Classifying what-type questions by head noun tagging, in: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1, Association for Computational Linguistics, 2008, pp. 481–488.

[18] D. Moldovan, M. Paşca, S. Harabagiu, M. Surdeanu, Performance issues and error analysis in an open-domain question answering system, ACM Trans. Inf. Syst. 21 (2) (2003) 133–154.

[19] F. Benamara, Cooperative question answering in restricted domains: the webcoop experiment, 2004.

[20] T. Hao, W. Xie, F. Xu, A wordnet expansion-based approach for question targets identification and classification, in: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, Springer, 2015, pp. 333–344.

[21] A. Mohasseb, M. El-Sayed, K. Mahar, Automated identification of web queries using search type patterns. in: WEBIST (2), 2014, pp. 295–304.

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, ACM SIGKDD Explorations Newslett. 11 (1) (2009) 10–18.