ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A REVIEW ON QUESTION CLASSIFICATION USING MACHINE LEARNING BASED ON SEMANTIC FEATURES

S.Jayalakshmi*, Dr. Ananthi Sheshasaayee
***** Research Scholar, Periyar University, Salem
Associate Prof. & Head, PG & Research Dept. of Computer Science,
Quaid-e-Millath Government College for Women (Autonomous), Chennai

ABSTRACT

One of the most important aspects of the learning process is the assessment of the knowledge acquired by the learners. In typical assessment like Exam, Assignment or Quiz, a grader provides students with feedback on their answers to questions related to the subject matter. In this way Question classification plays an important role in Question answering. Its main role is to assign a suitable semantic category to the question posted in natural language that represents the type of the required answers. It is a major challenge for the automated Question Classification function. In order to solve this kind of complexity, learners use lexical, syntactic and semantic features to analyze the questions. This paper presents a review on various approaches for Question classification using Machine learning approach based on Semantic Feature.

KEYWORDS: Question Classification, Natural Language, Syntatic features, Semantic features, Machine Learning

INTRODUCTION

An important step in question answering system is to classify the question to the anticipated type of the answer. The World Wide Web is an attractive feature for searching information as it offers a massive amount of textual and pictorial information to the general public especially in the student community. The amount of information growth may lead to increase the difficulty level in finding the specific information. The traditional information retrival scheme works well in several aspects but users are not satisfied of searching an answer in the collection of thousands of documents. Question answering schemes solve this problem by providing Natural language interfaces in which users can put forward their information queries in terms of Natural language question and retrieve the exact answer instead of retrieving a set of relevant files. Question Answering is a research that attempt to build a system that can retrieve accurate answers to questions posted in Natural language from a vast collection of documents like www.

This paper mainly focuses on Question Classification, Passage Retrieval and Answer Grading. Question classification is one of the main role is to assign a suitable semantic category to the question posted in natural language that represents the type of the required answers. For example the question start with "where" it requires an answer type of "location" and similarly "when" questions require an answer type of "date" or "time". The consequent step of the question classification is to retrieve a relevant passage as an answer to a given question. The two main aspects of passage retrieval are query formulation and query submission. Query Formulation translates the question into a suitable representation that can be exploited by an information source to retrieve relevant passages. This representation can be made based on an unstructured set of keywords. In the query submission step, the query generation in the previous step is to feed input to any of the information sources to retrieve the required passage relevant to the given question.

The most important facet of the learning process is the assessment of the knowledge acquired by the students or learners. The exam grader or teachers validate the student's performance by providing better feedback to their answers. There are certain circumstances in which a large number of worldwide sites have only limited teachers or instructors to validate the knowledge of the subject matters. Therefore the automated computer assessment plays a major role to assign a large number of student's test performance.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

QUESTION CLASSIFICATION (QC) CHARACTERISTICS

QC is a vital component of the question answering system that aids the system to predict the expected answer type. The prediction can work on the three different statistical features, namely lexical, syntactic and semantic features.

A. Lexical features: Lexical features represent the relationship between words that mostly extracted from the question. Most of the scenario uses word level n-grams as lexical features. It also includes the techniques of stemming and stop-word removal that reduces the dimensionality of feature set.

B. Syntactic features: It represents syntax related features that parse the structure of the question based on the grammar. Syntactic features include question head-words and part-of-speech (POS) tags.

C. Semantic features: Semantic feature is a semantically related word that associated with an exact question class. Semantic features include the WordNet relation, Hypernyms, named entity, antonyms, synonyms and semantic headwords.

- Based on the question type, different approach can be selected to extract an appropriate answer.
- A mis-classified question not able to retrieve the correct answer as it leads to incorrect guess about the answer.
- QC helps in deciding the exact class in which the answer belongs to. Thus, it reduces the possible candidates for extract answers.
- The motivation of QC is to find the answer category that reduces the search space.
- The classification of questions into a number of semantic categories not only find the answer but also different processing strategies.
- A great deal of QC used carefully for standard expression and handwritten language rules to parse the question.

QUESTION CLASSIFICATION APPROACHES

The QC approach is classified into three different approaches as shown in Fig.1.

- Rule-based approach
- Machine learning approach
- Hybrid approach

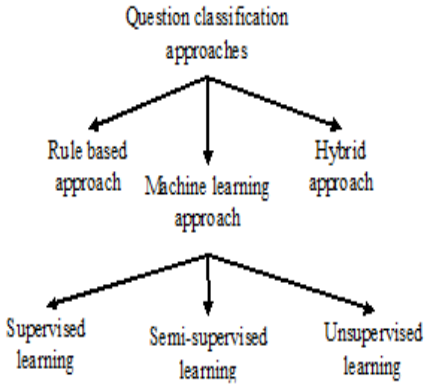

Figure. Question classification approach

The above figure represents the types of approaches for Question Classification.The remaining part of this paper focuses on review of Machine learning Question classification approaches based on Semantic feature.

VARIOUS MACHINE LEARNING QUESTION CLASSIFICATION APPROACHES

The Machine learning (ML) is one of the major part of artificial intelligence. The machine learning technique automatically classifies the question based on training and testing. The aim of ML technique is to classify questions in the natural language. The capability of machine learning improves the QC performance. The ML based QC overcomes the drawback in rule based QC. The ML techniques need human resource for manually label documents in order to use as a training set.

LITERATURE SURVEY ON MACHINE LEARNING QUESTION CLASSIFICATION BASED ON SEMANTIC FEATURE

In Machine learning question classification, the training knowledge is a set of questions along with their correct classes. Furthermore, in the training set the number of documents and its type is not straightforward. The machine learning algorithm is widely divided into three broad categories: supervised learning, semi-supervised learning and unsupervised learning. The ML techniques based QC uses lexical, syntactic and semantic features. An Extreme Learning Machine (ELM) uses semantic features to improve both training and testing compared to the benchmark of the SVM classifier [1]. The function of ELM is to classify the semantic features of statistical QC.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

The machine learning algorithm based QC analyze the semantic information on text classification [2]. Its prime goal is to classify questions into different semantic types that impose constraints on potential answers. The layered semantic hierarchy of the answer type guides hierarchical classifier and finally classifies into fine-grained classes. The benefits of semantic analysis features will be extracted for this class.

The QC system uses machine learning approaches. It presents an experimental result for question parsing and question answering types [3]. It classifies the questions to explain the different answer types termed as Qtargets.

SUPERVISED LEARNING APPROACH

A supervised learning method is defined as a useful structure with labeled classes. Supervised learning learns a function from a training set of pairs of inputs and relevant outputs. It is said to be supervised as it requires a supervisor to guide the learning process that offers the expected output to the corresponding input. The supervised machine learning approach trains the classifier using training set and tests the accuracy of the classifier using a test set. It achieves high accuracy of the QC. The supervised learning chooses different classifiers such as Support Vector Machine (SVM), Maximum Entropy model (ME), Advanced kernel method, Sparse Network of Winnow (SNoW), Language Modeling, k-Nearest Neighbor, Naïve Bayes algorithm and Decision Tree.

The automatic QC approach based on a machine learning model called supervised learning classification problem [4]. The main goal is to classify natural language questions to improve the entity type of answer. It extracts different features such as lexical, syntactical, and semantic from a question. In order to improve the accuracy of QC, it combines lexical, semantic and syntactic features. The technique in [5] improves the QC using term weighting methods in an efficient way. The task of question type classification is to find out the type of a question and it is completely different from the content based text classification. Therefore, the proposed scheme investigates well known supervised and unsupervised term-weighting methods for QC.

SEMI-SUPERVISED LEARNING APPROACH

The semi-supervised learning methods used both labeled and unlabeled data in the training set [6]. A kernel function in the semi-supervised machine learning that acquires unlabeled data using latent semantic information in which kernel allows supervised learning using bag-of-word representation [7]. This type of kernel method is flexible that supports different languages and domains.

UNSUPERVISED LEARNING APPROACH

An unsupervised learning method is defined as the useful structure without labeled classes. The unsupervised learning method learns a function from a set of input without prior knowledge of the output. The example of the unsupervised learning approach is clustering that group similar input object together into clusters. The clusters make use of informative answer in each cluster, frequently longest answer. The complex question answering system uses an unsupervised learning approach [9]. The empirical method of two unsupervised learning techniques, namely K-means and Expectation Maximization computes the relative importance of sentences.

CONCLUSION

The Question Classification is a crucial component of the question answering system. The QC is a hard problem. The QC is broadly divided into rule based, machine learning and hybrid approaches. This paper presents machine learning approaches which are not enough to execute the Question Answer(QA) system efficiently. It needs natural language processing techniques and semantic and similarity features to provide better results. The next generation of the QC system needs to understand more complex questions and different forms of questions. In order to categorize questions, effective NLP and machine learning are necessary to meet the future requirements of QC.

REFERENCES

1. Hardy & Yu-N Cheah, "Question Classification Using Extreme Learning Machine on Semantic Features" Journal of Information and Communication Technology Research & Applications, Vol. 7, No. 1, pp. 36-58, 2013
2. Xin Li and Dan Roth, "Learning Question Classifiers: The Role of Semantic Information" Natural Language Engineering, Vol.12, No.3, pp.229-249, 2004
3. Ulf Hermjakob, "Parsing and Question Classification for Question Answering" in Proceedings of the Workshop on Open-Domain Question Answering, Association for Computational Linguistics, Vol.12, pp. 1-6, 2001
4. Megha Mishra ,Vishnu Kumar Mishra and Dr. H.R. Sharma, "Question Classification Using Semantic, Syntactic and Lexical

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Features" International Journal of Web & Semantic Technology,Vol.4, No.3, 201
5. Xiaojun Quan, Wenyin Liu, and Bite Qiu, "Term Weighting Schemes for Question Categorization" IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 33, No. 5, 2011
6. Ratsaby, J. and Venkatesh, S, "Learning From a Mixture of Labeled and Unlabeled Examples with Parametric Side Information", In Proceedings of the 8th Annual Conference on Computational Learning Theory, pp. 412-417, 1995
7. Tomas and Claudio Giuliano, "A Semi-Supervised Learning Approach to Question Classification" In Proceedings of European Symposium on Artificial Neural Networks (ESANN) - Advances in Computational Intelligence and Learning, pp. 35- 40, 2009
8. Xiaojin Zhu and Andrew B. Goldberg, "Introduction to Semi-Supervised Learning" Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers Vol.3, No. 1, pp.1-13, 2009
9. Zhengtao Yu, Lei Su, Lina Li, Quan Zhao, Cunli Mao, and Jianyi Guo, " Question Classification Based on Co-Training Style Semi-Supervised Learning" Pattern Recognition Letters, Vol. 31, No. 13, pp. 1975-1980, 2010
10. Yllias Chali, Shafiq R. Joty and Sadid A. Hasan, "Complex Question Answering: Unsupervised Learning Approaches and Experiments" Journal of Artificial Intelligence Research Vol.35, No.1, pp.1-47, 2009
11. Michael Collins and Yoram Singer, "Unsupervised Models for Named Entity Classification" In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pp.100–110, 1999
12. Zhengtao Yu, Lei Su, Lina Li, Quan Zhao, Cunli Mao, and Jianyi Guo, " Question Classification Based on Co-Training Style Semi-Supervised Learning" Pattern Recognition Letters, Vol. 31, No. 13, pp. 1975-1980, 2010
13. Alexander Clark, "Inducing Syntactic Categories by Context Distribution Clustering" In Proceedings of the 2nd Workshop on Learning Language in Logic And The 4th Conference on Computational Natural Language Learning Vol. 7, pp. 91–94, 2000.
14. Joan Silva, Luısa Coheur, Ana Mendes, and Andreas Wichert, "From Symbolic to Sub Symbolic Information in Question Classification" Artificial Intelligence Review, Vol.35, No.2
15. pp: 137–154, 2011
16. Bo Qu, Gao Cong, Cuiping Li, Aixin Sun, Hong Chen, "An Evaluation of Classification Models for Question Topic Categorization" Journal of the American Society for Information Science and Technology, Vol. 63, No. 5, pp. 889-903, 2012
17. Abdullah M. Moussa and Rehab F. Abdel-Kader, "QASYO: A Question Answering System for YAGO Ontology" International Journal of Database Theory and Application, Vol. 4, No. 2, 2011
18. Minh Le Nguyen, Thanh Tri Nguyen and Akira Shimazu, "Subtree Mining for Question Classification Problem" The International Joint Conferences on Artificial Intelligence (IJCAI), pp. 1695-1700, 2007
19. Santosh Kumar Ray, Shailendra Singh and B. P. Joshi, "A Semantic Approach For Question Classification Using Word Net and Wikipedia" Pattern Recognition Letters, Vol.31, No. 13, pp. 1935-1943, 2010
20. Andreas Merkel, Dietrich Klakow, "Improved Methods for Language Model Based Question Classification" In 8th Annual Conference of the International Speech Communication Association, pp. 322-325, 2007
21. Ratsaby, J. and Venkatesh, S, "Learning From a Mixture of Labeled and Unlabeled Examples with Parametric Side Information", In Proceedings of the 8th Annual Conference on Computational Learning Theory, pp. 412-417, 1995