

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

Predicting Question Subjectivity in E-Commerce

Divya Panicker¹, Archana Chaugule²

P.G. Student, Department of Computer Engineering, DYPIET, Pimpri, Pune, Maharashtra, India¹

Associate Professor, Department of Computer Engineering, *DYPIET*, Pimpri, Pune, Maharashtra, India²

ABSTRACT: As human we are always in the pursuit to get information related to our topic. This information can be needed from student in school to professional working in different domain. Question answering system will help user to get accurate answer of their particular question in natural language. This system will provide explicit, short and most accurate answer to the user than giving set of relevant answer. Question answering system provides the user with functionality to get precise answer of the question articulated in their natural language. Question classification is a vital part of any question answering system. The accuracy of answer provided by question answering system heavily depends on the way question is classified. Accurate question classification leads to retrieval of exact answer in question answering system. Extracting whether a question is subjective or objective helps in analyzing the actual structure of answer expected by the users.

KEYWORDS: Machine learning, Natural Language Processing, Neural networks, Question answering system, Questions classification

I. INTRODUCTION

Question answering system can be used to get exact answer to different questions that are asked by different users. Many questions in question and answering system are presented in the natural language form, where users are free to give answer to the question in any of natural language form. Question processing, answer retrieval and answer generation are the three main component of any Question answering system. In Question Processing, natural language question which is given by users are parsed and form machine readable form questions. To form this question, various approaches are used which is then followed by answer retrieval. In answer retrieval candidate answer will be extracted using various intermediate to represent the question. Finally, user understandable answer is generated in this phase. This generated answer is provided to users which are precise as well as accurate.

Question classification is most important part of most of the question answering system. If question are not classified properly then it will generate answer which will not be accurate. For example, the question which is discovered by first Smartphone should be classified into the type of human. This information classified will help to know that answer to that particular question must be name of some human being. If the question is not classified correctly as located then those retrieved answer will not be relevant to as per the question. Through classification of question information will be extracted and it will help to reduce the criteria and size for search to identify correct answer from dataset which is available.

Most Artificial neural networks (ANN) explain various form of learning rules that will be assigned to it so that it can alter its weights of associated connection as per different input pattern which are given with different types of pattern. This paper contains introduction, related work where various methodologies used to classify the question in question answer system are explained in section II, Section III describes proposed system architecture. Section IV explains the Experimental setup which is followed by section V that will conclude the paper.

II. RELATED WORK

Various approaches have been used for classification of question as per respective question type. Classification of question can be done using various approaches depending upon the type of question.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

M. R. Morris et al. [1] has described different facts of why questions are asked by sending messages to the status of social network. They have also discussed in details about their study of data, different ways to ask question, various types of question that can be asked and instead of using traditional tool like Google, many of the users get motivated by asking questions.

S. A. Paul et al. [2] has introduced different method to tag human using Amazon Mechanical Turk. This tagging can be done by searching similar outcomes or Wikipedia articles values should be worth of. They have also used Amazon Mechanical Turk to distinguish and tag different tweet by the way the question are asked. They have shown the importance of micro task by means of organizing and by doing various studies and including task for brainpower of different persons. For completion of task different instruction are given but this instruction are not identified by system.

D. Moy et al. [3] has discussed the uses of different machine learning techniques and to differentiate question as informational and conversational, various technique can be used and all this differentiation can be done automatically. Due to this, it will help to find all technical groups that have good effect and various types of questions that will be regarded as categorical, linguistic, and social differences are trained to proceed. Even they have introduced a classifier; this classifier can be used to classify quantitative features which are taken from different data. This assumed feature will be different from the assumed, as assumed classifier will be displayed as informational and conversational questions. Some extra features are added by author to improve model of social networking, because of it performance will be better.

A. Pal et al. [4] has introduced a novel concept in question answer communities of question temporality where different measures are considered while measuring the length of answer of particular question which have expected value. The time duration between when the asked question was answered and Search engine will help the search engine and allow the question answer provider to observe appealing framework, it will help to select better answer from the user experience. Using feature such as non absolute time or text can be what, where and who can be used to guess temporal questions.

M. Jiang et al. [5] has introduced a relatively new and interesting text issues categorization way that will do group blogging such as political, liberal or conservative. This all classification depends on political learning. Author will categorized the sentence depending whether its subjective with a clue that sentence will not contain more than one subjective clues and this clues depend upon the inquirer dictionary and subjective sentence are recognized from this category. From the opinion expression of sentence and various different features are used to build a classifier for political learning.

P. Biyani et al. [6] has introduced a novel approach where supervised machine learning can be used. This approach can be used in online forum to differentiate question as subjective or objective. From this approach various feature can be generated such as N-gram and POS tagging for data which will give good outcomes for trends in forum.

B. Li et al. [7] has discussed about the different problem where orientation of subjective question orientation can be identified automatically In Question answering communities. Author has also found solution for supervised learning which contains different features and this features should be differentiated from each other as they are designed for different task. Characters that are not case sensitive 3 gram features will be very helpful as they will represent their task and this is proved by experiment conducted by them. They have even explored different answer given by different user..

L. Chen et al. [8] have introduced a novel model where this model is capable to categorize this question into three different categories. These categorizations are done on the basis of user consideration of question as objective, social, subjective. From machine learning author has build a predictive model where this model can be used to identify way question can be asked. So as the output system has grouped features into 2 groups where this feature will not be depending on each other and for prediction this feature can be used. In structure of supervised learning, author has considered two views where they have taken a larger amount of question which are not label and a small amount of question which are labeled manually.

J. Wiebe et al. [9] have showed that resultant generated by his subjectivity classifier along with annotated text can be used for training. Performance can be improved using supervised leaning approaches. Using extraction pattern the art state of objective sentence can be classified and it can be connected with objectivity and create objective classifier that get higher recall.

C. Lin et al. [10] have introduced model where LDA subjectivity is detected, where sentence level classification o subjectivity is done. Existing approach required extraction for linguistic pattern or label copra, due to this it is different from others. Evaluation on MPQA dataset can be done using model. Previous result proved that if objective call excluded then subjLDA can perform well on LDA over all metrics that are considered for evaluation. If both

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

approaches are considered the newly introduced approach produces better result than previous one. If more subjective words are added then it will produce better results, this observation made during experiment. By adding extra neural words performance can be increased.

Anderson et al [11] has proved a better way where user can focus on to group that will process for the creation of knowledge from one of answer. Author has considered all types of question where answer collection are analyzed for fundamental which can act on it. In question answer characterizations all previously work done are considered. Different activities that will shape the community to set the answer are considered. Answer and voter that are present during that time are considered. Even dynamic activity that has eventual outcome is too considered. They have proved that understanding this property will be useful during predicting different quantities which have values for larger term answer and question for that question.

Zhe Liu et al [12] have proposed a model which can classify the question. Due to rapid growing interest of people in social networking sites has given a new way to take various information through online. In these social networking sites, people post various question in their status and from this status various user can get various information. In this paper author has model a system which will classify the question as subjective or objective. To classify the question, they have considered many lexical, syntactical and contextual features. Using this feature they have build a classifier and their experiment proved better result.

Geerthik.S et al [13] has discussed about question where large no of answer Are given by different user it is very difficult to identify the best answer from it. So different methods and methodologies are used to identify best answer and those answer should be ranked according to its quality and appropriateness. But to rank this answer various issue arises where the problem is to identify best answers by CQA sites where author of that sites has many followers. The second is if a question is posted and if the best answer is written by a user late after many answer, the probability of the answer getting ranked as top answer with high rank is very less. So author has introduced a new algorithm ie reward rank algorithm . The RewardRank algorithm treats users with more followers and less followers equally as well as the answer written at various time intervals by different users is also treated equally.

III. SYSTEM ARCHITECTURE

Question answering system is very important application to most of its diverse users where we get precise and accurate information to user questions. Questions can be classified as subjective questions which deal with subjective form of question or objective question which deals on facts only. Most of the previously created question answering systems did not include classifying of question as subjective or objective before retrieving answers for the user questions. Subjective or objective question can be classified as it requires human level of perception. Neural networks can be used to improve the accuracy of subjective vs. objective classification of question. Most of the question answering system tends to give precise and accurate answer for user questions but many systems still fails to provide single and accurate answer to question. Reward ranking algorithm can aid in ranking the list of answers provided by question answering system and provide most related answers for user question. By integrating reward ranking algorithm with question answering system we get much accurate question answering system, designed for e-commerce domain where user get precise answer to their question which can aid the users in deciding which product to buy from e-commerce site. The proposed system will help the e-commerce site users to get relevant answers regarding the product of their interest based on the reviews provided by other customers. As information in the internet is growing drastically, so search engine need to be more powerful than before. Question answering system are designed such that accurate answer portion is given as answer to the question asked to a query. Different method of information retrieval, natural language processing and machine learning are used by question answering system. Question processing, passage retrieval and final answer processing are the three phases of our system.

A. Question Processing: It is to investigate the question and make a appropriate IR query as well as detecting the individual category of the answer, a category name which specifies the type of answer. The first task is called question reorganization and the second is called question classification.

B. Passage Retrieval (From Database): Passage retrieval is to question over the IR engine, route the comeback credentials and return candidate passages that are likely to surround the reply. Question classification comes

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

multipurpose at this time: it can establish the explore approach to recover candidate passages. Depending on the question class, the explore question can be changed into a form which is most suitable for judgment the answer.

C. Answer Processing: A QA classification is to method the nominee passages and extracts a section of words that is likely to be the answer of the question. Question classification over again comes helpful here. The candidate answers are ranked according to their likelihood of being in the same class as question class and the top ranked answer(s) will be considered as the final answer(s) of the question

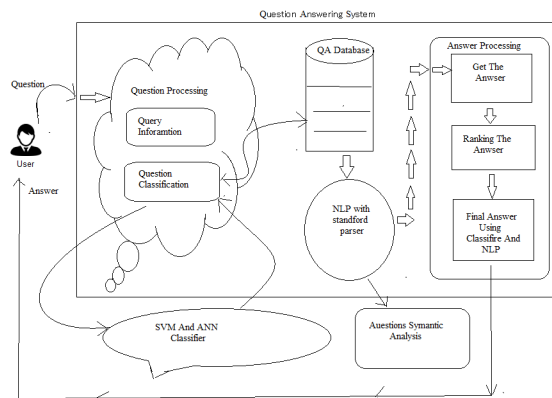


Fig. QA System Architecture

IV. EXPERIMENTAL RESULTS

EXPERIMENTAL SETUP

Evaluation Metrics

For accurately evaluating QAS systems, there is a need of matrices which measures the actual performance of our system. In any Question answering system it is very much important to retrieve explicit and precise answer to which will satisfy the user. There are different performance measures available to evaluate question answering systems.

Dataset

we have created the dataset of amazon where we have considered some course such as desc, ent, loc etc. Question will be differentiated using this course.

RESULT EVALUATION

Using different types of classifier, the question classification is studied. In this work we adopted SVM as well as neural networks.

1. Support Vector Machines:

Support vector machine is a supervised learning method for classifying data. It is especially successful for high dimensional data. SVM is a linear discriminant model. This model will try to learn maximum margin hyper lane which will help them to separate the classes.

2. Back-Propagation Neural Networks:

Neural network consist of multi layer feed forward neural network. This multi layer feed forward neural network are trained with back-propagation learning rule. This multi-layer feed forward neural network consists of multi layers where layers can be divided as input layer, output layer and multiple hidden layer. Each of this layer consist of multiple neurons and each neuron has forward connection with all neurons of its subsequent layer. This subsequent layer connection can be used to calculate the weight parameter. Input Weight of each neuron is calculated using sum of input signals and output weight is calculated using input signal function and optional threshold parameter.

On the basis of fine grained classes so as to reduce space, comparison between SVM and BPNN classifier is shown below

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

No Of Features	SVM Accuracy	BPNN Accuracy
50	75	83
100	80	84
150	82	87
200	83	88
250	78	91
300	79	88
350	81	91
400	80	92
450	79	92
500	73	91

Graphical representation of SVM and BPNN algo



1

V. CONCLUSION

Question classification system will provide more accurate answer to the given question if question answering system will classify question accurately. Nature of question can be identified as subjective or objective by classify the question and its will target the input question to retrieve related subjective objective answer form from the dataset which is available. Machine learning and neural networking will help to classify the question so that accuracy will be improved when multiple features are considered. Question answering system is very important as it will be applicable in many domain such as education, medical, tourism as well as organisation.

REFERENCES

- [1] M. R. Morris, J. Teevan, and K. Panovich, "What do people ask their social networks, and why?: A survey study of status message q&a behavior," in Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 1739–1748, 2010.
- [2] S. A. Paul, L. Hong, and E. H. Chi, "What is a question? Crowdsourcing tweet categorization," in Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 578–581, 2011.
- [3] F. M. Harper, D. Moy, and J. A. Konstan, "Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites," in Proc. SIGCHI Conf. Human Factors Comput. Syst., pp. 759–768, 2009.
- [4] A. Pal, J. Margatan, and J. Konstan, "Question temporality: Identification and uses," in Proc. ACM Conf. Comput. Supported Cooperat. Work, pp. 257–260, 2012.
- [5] M. Jiang and S. Argamon, "Exploiting subjectivity analysis in blogs to improve political leaning categorization," in Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., pp. 725–726, 2008.
- [6] P. Biyani, C. Caragea, and P. Mitra, "Predicting subjectivity orientation of online forum threads," in Proc. Comput. Linguistics Intell. Text Process., pp. 109–120, 2013.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 8, August 2017

- [7] B. Li, Y. Liu, A. Ram, E. V. Garcia, and E. Agichtein, "Exploring question subjectivity prediction in community QA," in Proc. 31st Annu.Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., pp. 735–736, 2008.
- [8] L. Chen, D. Zhang, and L. Mark, "Understanding user intent in community question answering," in Proc. 21st Int. Conf. Companion World Wide Web, pp. 823–828, 2012.
- [9] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in Proc. 6th Int. Conf. Comput. Linguistics Intell. Text Process, pp. 486–497, 2005.
- [10] C. Lin, Y. He, and R. Everson, "Sentence subjectivity detection with weakly-supervised learning," in Proc. 5th Int. Joint Conf. Natural Lang.Process., pp. 1153–1161, Nov. 2011.
- [11] Anderson, Ashton, et al. "Discovering value from community activity on focused CQA sites: a case study of stack overflow."Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012
- [12] Zhe Liu and Bernard J. Jansen, "Understanding and Predicting Question Subjectivity in Social Question and Answering." Ieee transactions on computational social systems, vol. 3, no. 1, march 2016, pp. 32-41
- [13] Geerthik.S, Venkatraman.S, K.Rajiv Gandhi," RewardRank:A Novel Approach for Positioning Uses Answers in Community Question Answering System", Conference On Information Communication And Embedded System,2016, pp.1-6.