



Automatic Identification of Potential Respondents to Subjective and Objective Questions

Sonali U. Tapase

PG Student, Ashokrao Mane Group of Institution,
Vathar, Kolhapur, Maharashtra, India

Prof. K. B. Manwade

Associate Professor, Ashokrao Mane Group of
Institution, Vathar, Kolhapur, Maharashtra, India

ABSTRACT

In recent years social networking sites are very popular for user needs. Mostly people use an important social networking information source such as Twitter to find answers to their questions. Twitter contains microblogging services for informal information interactions. In this paper, we classify the subjective and objective questions using Naïve Bayes algorithm and also find the respondent users. For classification we build the feature extraction techniques such as lexical, syntactical and contextual in terms of the way they are asked and answered. For applying the classification on a larger dataset using other social media and analysis of performance.

Keywords: *Information seeking, social search, Twitter, other social networks*

1. INTRODUCTION

Social networking sites (SNSs) or social media is a platform to build social networks or social relations among people who share similar personal and career interests, activities, backgrounds or real life connections. Because of the social networking sites the communication between people has made more diverse and convenient.

Social question and answering (social Q&A) provides people more easy and direct way to express their information need so that the individuals can broadcast their requests to all friends and receive more personalized responses using social Q&A as compared to the typical search engine services such as google and bing. Considering the increasing popularity of social Q&A there are variety of

questions such as subjective, objective knowledge or factual truth. Subjective questions requires more diverse replies based on their personal opinion and perspective and also it contains more contextual information so that it takes more working hours. Compared with the subjective questions the objective questions more focus on the accuracy of the responses and it takes the shorter time between posting and receiving responses. Subjective questions are more attracted from strangers than the objective ones. Subjectivity analysis using the comprehensive set of features from Lexical, Syntactical and Contextual perspectives. Using the Naïve Bayes algorithm it classifies the testing dataset in subjective and objective questions and subjective question list goes to the identification of respondent user and also the training set features and respondent list of the training dataset goes to the identification of respondent user.

2. RELATED WORK

Zhe Liu, Bernard J. Jansen to automatically decide which answering strategy to use, based on ASK question types using question features from the perspectives of lexical, topical, contextual, and syntactic as well as answer features[1]. The ASK taxonomy differentiate questions posted on social networking into three parts considering the nature of the questioner's inquiry of accuracy, social, or knowledge.

They develop and implement a predictive model based on features extraction using machine learning techniques. The automated method proves to be very effective in classifying ASK types of questions.

Bernard J. Jansen and Mimi Zhang, Kate Sobel, Abdur Chowdury. Uses [2] the Word of mouth (WOM) process of conveying information from person to person and plays a major role in customer buying decisions. The relationship between company and customers are affected from the effects of services in the commercial sectors.

Dejin Zhao, Mary Beth Rosson provide a new communication channel for people to broadcast information that they likely would not share otherwise using existing channels (e.g., email, phone, IM, or weblogs) and also provide a variety of impacts on collaborative work (e.g., enhancing information sharing, building common ground, and sustaining a feeling of connectedness among colleagues)[3].

F. Maxwell Harper, Daniel Moy, Joseph A. Konstan using [4] machine learning techniques to automatically classification questions in informational or conversational, learning in the process about categorical, linguistic, and social differences between different question types. To distinguishing Informational and Conversational Questions in Social Q&A Sites.

Baoli Li, Yandong Liu, Ashwin Ram, Ernest V. Garcia, Eugene Agichtein. Using [5] case-insensitive features and n-gram techniques to remove spelling errors and poor formatting, and POS features to attempt to capture simple grammatical patterns. To automatically identifying subjectivity orientation of questions in QA communities, and explore a supervised machine learning solution with different features.

3. PROPOSED WORK

In proposed system, the different Social Networking Sites (SNS) are used for the collection of different questions. Using the percentage ratio the questions are classified in training dataset and testing dataset. It uses the Lexical, Syntactical and contextual features for the extraction of data. In Lexical feature N-gram is used to count the frequencies of all unigram, bigram and trigram tokens that appeared in training data. POS (Part Of Speech) tagging used to distinguish the two types of questions, as it can add more context to the words used in the interrogative tweets. The MPQA subjectivity lexicon used to count the number of subjective clues in each question. The Syntactic features describe the format of subjective or objective information-seeking tweet. It also includes the length of the tweet, number of clauses or sentences in the

tweet, whether or not there is a question mark in the middle of tweet and also consecutive capital letters. The contextual features used to find the presence of hashtags, emoticons and mentions in the tweet. For the classification of testing dataset it uses Naïve Bayes algorithm. It is used for text retrieval and text categorization. The proposed system identifies the respondent users of subjective question list, training set features and the respondent user list of training set features.

A. System Architecture

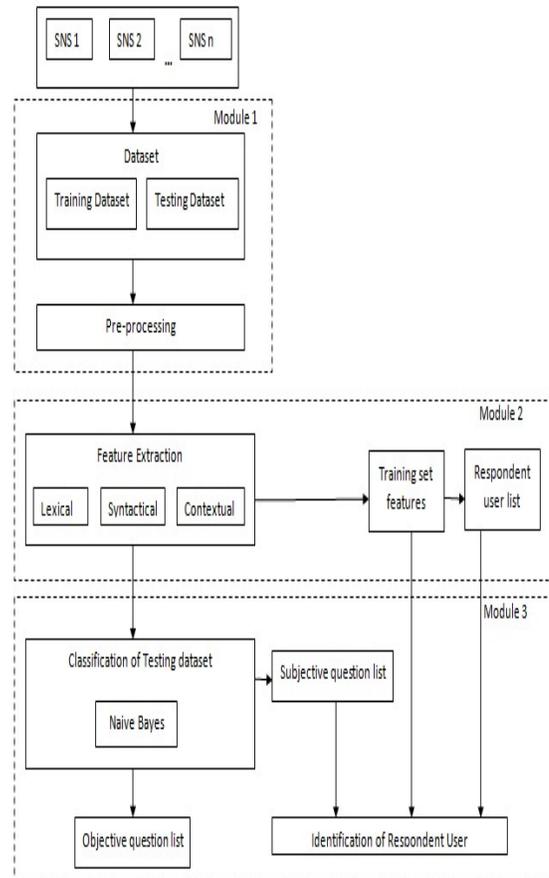


Figure 1: Architecture of Automatically Identify Potential Respondents to Subjective and Objective Question

The proposed system provides following modules:

Data Collection:

In this module, the proposed system refers different Social Networking Sites (SNS) for the collection of data. That data is divided in Training and Testing data for the processing.

Pre-processing:

In this module includes the cleaning, transformation, feature extraction and selection of data which is collected from the different Social Networking Sites (SNS).

Feature Extraction: It contains main three features are as follows:

1. Lexical:

The lexical features are N-gram, POS tagging and MPQA subjectivity lexicon. The N-gram feature is used to count the frequencies of all unigram, bigram and trigram tokens. Part Of Speech (POS) tagging used to distinguish the two types of questions as it can add more context to the words used in the interrogative tweets. The MPQA subjectivity lexicon is used to count the number of subjective clues in each question.

2. Syntactical:

It describes the format of subjective or objective information seeking tweet. It also includes the length of the tweet, number of sentences/clauses in the tweet, whether or not there is a question mark in the middle of the tweet and also consecutive capital letters in the tweet.

3. Contextual:

It includes the presence of hash tags, emoticons and mentions in the tweet.

Classification of Testing dataset using Naive Bayes: The Naive Bayes is supervised learning algorithm which is used to classify the data. It is used for text retrieval and text categorization. It only requires a small amount of training data to estimate the parameters necessary for classification. It is a model which is easy to build and particularly useful for very large data sets.

Identification of Respondent User: The subjective question list of training and testing dataset is automatically handover to the respondent user.

4. SCOPE OF THE WORK

Social Networking Sites (SNS) provide people to easy and convenient way for the communication and their individual needs. It automatically identifies the

potential respondents for subjective and objective questions using different Social Networking Sites.

The Pre-processing is used to remove the rare words, lowercase letters and stemming of data in the tweet. The features are extracted from the Lexical, Syntactical and Contextual features in tweet. In proposed system the Naïve Bayes supervised learning algorithm is used to classify the testing dataset.

The purpose of identifying potential respondent users for removing the confusion between subjective and objective questions and also to give the appropriate answer to the user.

5. CONCLUSION

In this paper, different features of extraction techniques were studied and a new system is proposed, that finds potential respondents to subjective and objective questions. The features are extracted from the Lexical, Syntactical and Contextual features in tweet and also remove the rare words, lowercase letters and stemming of data in the tweet. Naïve Bayes supervised learning algorithm is used to classify the testing dataset. The new proposed system will automatically finds the potential respondents for removing the confusion between subjective and objective questions and also gives the appropriate answer to the user.

REFERENCES

1. Zhe Liu, Bernard J. Jansen, "ASK: A Taxonomy of Accuracy, Social, and Knowledge Information Seeking Posts in Social Question and Answering", march 2016.
2. B. J. Jansen, M. Zhang, K. Sobe, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," J. Amer. Soc. Inf. Sci. Technol., vol. 60, no. 11, pp. 2169–2188, Nov. 2009.
3. D. Zhao and M. B. Rosson, "How and why people Twitter: The role that micro-blogging plays in informal communication at work," in Proc. ACM Int. Conf. Supporting Group Work, 2009, pp. 243–252.
4. 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., 2009, pp. 759–768.

5. 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., 2008, pp. 735–736.
6. P. Biyani, S. Bhatia, C. Caragea, and P. Mitra, "Using non-lexical features for identifying factual and opinionative threads in online forums," *Knowl.-Based Syst.*, vol. 69, pp. 170–178, Oct. 2014.
7. Z. Liu and B. J. Jansen, "Almighty Twitter, what are people asking for?" *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 49, no. 1, pp. 1–10, 2012.
8. N. Aikawa, T. Sakai, and H. Yamana, "Community QA question classification: Is the asker looking for subjective answers or not?" *IPSIJ Online Trans.*, vol. 4, pp. 160–168, Jul. 2011.

