Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism

Vallikannu Ramanathan
Research Scholar
Alagappa University
Karaikudi, Tamilnadu
India
vallikannug@gmail.com

T.Meyyappan
Professor
Alagappa University
Karaikudi, Tamilnadu
India
meyyappant@alagappauniversity.ac.in

Abstract - Sentiment analysis plays vital role in the internet era due to extensive range of business applications and social media. Inspiration behind sentiment analysis is that it provides people's opinion about the product, which helps to improve the product quality. It also supports to take purchase/manufacturing decisions. In this paper we apply sentiment analysis to catch people feedback about Oman tourism using social media messages. For this we use tweeter data set to analyze tourist opinion about this country. In this paper, we recommend innovative sentiment analysis method based on common sense knowledge (Domain Specific Ontology). We created our own Oman tourism ontology based on ConceptNet. Entities are identified from the tweets using POS tagger and entities are compared with concepts in the domain specific ontology. Further the sentiment of the extracted entities are determined by the combined sentiment lexicon approach. Finally semantic orientations of domain specific features are combined with respect to the domain. We deliberate conceptual semantic as feature which can be combined with machine learning algorithm to enhance the performance of sentiment analysis of Oman tourism.

Keyword- Domain specific ontology, Combine sentiment lexicon approach, Semantic features, Conceptual semantic sentiment analysis, machine learning method.

I. INTRODUCTION

Sentiment analysis [SA] is the area which discuss about people's opinion towards product, service, tourism, movies, political issues, education system etc. Text mining refers to extracting the useful information from the natural language text[1]. Generally most of the people express their opinion through social media like Twitter, Facebook etc. Sentiment analysis or opinion mining is the study of extracting user's opinion from the given text. With rapid development of internet, people are using social media sites, blogs, channels, forums and review sites to express their opinion about any topic in the current trend.

Of late, trillions of users share their opinions using microblogging. In social microblogging website users are sharing their own experience, discussing current issues in all the topics.

As many people are writing their view about movies, products, organizations, political issues and social issues every one's opinion has become important in social microblogging websites. For example, companies are eager to know the following:

- Product review and feedback (quality, service, price etc.)
- Number of likes and dislikes to their product
- Future expectation of people on their existing product

Sentiment analysis is the role of identifying opinions, emotions and evaluations in a given text. The important concept in sentiment analysis is to identify the polarity of overall text. It helps to indicate whether the text is positive, negative or neutral.

In this paper we concern individual's opinion about Oman tourism since Oman is the beautiful country, all are interested to visit Oman. Since twitter is the large carpus of data, twitter message have taken for this research to know the people sentiment.

We propose novel method based on domain specific ontology. We constructed our own Oman tourism ontology. Generally nouns are consider as entities, so nouns are extracted from the tweets using POS tagger which could be compared with concepts in ontology. Further entities/opinion extraction to be done by inquiring object-attribute pair in ontology. In addition to that we construct our own sentiment lexicon based on three existing lexicons such as Sentistrength, SentiWordNet and Opinion lexicon.

Finally we introduce semantic sentiment analysis concept which concern about the semantic meaning of the word. It would be classified in to two types contextual semantic and conceptual semantic. Contextual semantic deals with considering about neighbouring word. Conceptual semantic depends on outside knowledge such as ontologies and semantic network. We consider conceptual semantic as feature which can be incorporate with Naïve Bayes machine learning method to increase the performance of sentiment analysis.

II. RELATED WORK

Sentiment analysis is the mission of categorizing positive or negative opinions, feelings in the text. SA

is the multidisciplinary study which deeds methods from text mining, machine learning and natural language processing. The two most popular approaches for SA is machine learning approach and lexicon based approach. Lexicon based approach also known as dictionary based approach and knowledge based approach.

Lexicon based approach adopts that the sentiment orientation of a given text document is the average of the sentiment orientation of its words and phrases. Sentiment analysis is a special text mining work used to define and extract subjective attitude or sentiment of people from a given text [12].

III. PROPOSED METHODLOGY

A. Domain specific Ontology

ConceptNet is a semantic network deals with common sense knowledge. ConceptNet also known as knowledge graph which is used for natural language processing. Common sense in the ConceptNet is donated by the web crawlers in the world wide web (WWW). It is the largest freely available common sense knowledge. It is conveyed as a graph which has nodes (concepts) linked by edges (relationship between concepts). Few relationship (assertion) between concepts in the ConceptNet are IsA, CapableOf,, HasA and so forth [6].

In our proposed method we apply ConceptNet to build domain specific ontology for Oman Tourism. Further Oman tourism spots are added to expand the ontology to cover all parts of Oman. We expand the ontology by integrating the ontologies of related domains for the best coverage of specific features.
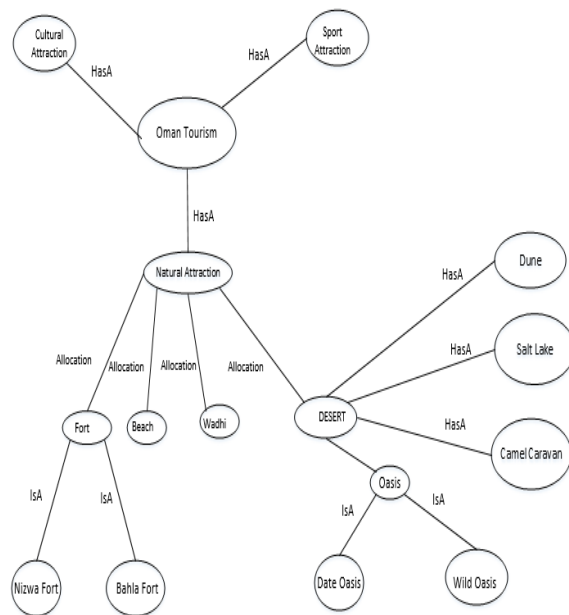

Fig. 1 . Ontology for Oman Tourism Domain

In fig. 1 we show the creation of ontology for Oman tourism, we can also construct ontology for

'Hotel'. Then we combine both ontologies 'Oman Tourism' and 'Hotel' by connecting with new relation 'CapableOf'. Similarly we can extend the tourism Ontology by adding 'Transport' ontology etc., to make the complete ontology domain for Oman tour.

B. Entity Specific Opinion Extraction

Entities are recognized using the tourism ontology designed in fig. 1. Consider the tweet message 'beautiful deserts are the main attraction of Oman', desert is an entity about which sentiment is expressed. Generally noun is treated as an entity. Here desert is noun and beautiful is the adjective which express the positive opinion. To trace out the entity in each tweet part-of-speech tags are used. We used Stanford English Twitter POS tagger for tagging the tweets.

Consider the example, 'Muscat is a peaceful city'. According to Penn part of speech tags, sentence is tagged as follows: 'Muscat NNP is VB a DT peaceful JJ city NN'. Here Muscat and city are treated as entities since we follow the rule of all nouns to be consider as entities. After the extraction of nouns, these nouns are matched with the domain specific ontology (Oman) concepts. All the inappropriate features are removed with the help of ontology. Actually we extract from our ontology data set such as objects (e.g., desert, beach, park, fort) and their attributes (sandy, beautiful, clean, largest) and compare with the features mined from the tweet. The assessment was completed by demanding twitter with object-attribute pairs in the ontology.

C. Lexicon based Approach

Sentiment lexicon contains words with their sentiment orientation values. We construct the new sentiment lexicon for our approach based on three existing lexicons such as Sentistrength, SentiWordNet and Opinion lexicon.

1) Sentistrength

Sentistrength is the publicly available resource developed by Thelwall [2]. It calculates the positive and negative sentiment score in short text. Positive sentiment value ranges from 1 (not positive) to 5 (extremely positive) and negative sentiment value ranges from -1 (not negative) to -5 (extremely negative). It has mainly developed for social media web text. For example 'I enjoy playing in the park but the fear of insects stops me from doing so', the term enjoy gives positive score 3, the term fear gives negative score -4. So the overall sentiment of the above sentence is max[3,-4] .It yields -1 which means negative sentiment.

2) SentiWordNet

SentiWordNet is a sentiment lexicon holding polarity score of the opinion words[3]. It has approximately 3 million words including nouns, verbs, adverbs, adjectives. Generally terms in

SentiWordNet classified into four classes namely adjective, adverb, verb and noun. SentiWordNet was built by using WordNet, this is based on 'bag of Synet' model. This lexicon holds booster word list and emoticon lookup table. SentiWordNet has three scores.

- Positive sentivalue
- Negative sentivalue
- Objective sentivalue

Each word in the SentiWordNet has three scores (positive, negative and neutral) value ranges from 0 to 1. This is because some words could be positive, negative or neutral depending on the context in which they are used. The sum of all three values is always 1.0. If the positive score is more than the negative and objective score, it is marked as +1. Suppose negative score is maximum then it is marked as -1.The objective score 0 signifies that the word is neutral. It does not express any opinion.

3) Opinion Lexicon

Opinion lexicon was the one of the oldest sentiment dictionary. It has list of 10000 positive words and negative words. Some misspelled words are given in the lexicon, since these misspelled words are used in the social media[4,5]. Our idea is to combine all the above three lexicons and create a new lexicon which comprises the words and their sentiment score.

This approach does not involve training data since opinion word in the text is directly paralleled with wordlist in the lexicons. This method has few limitations. First sentiment dictionaries are made up of set of existing static words, they would not able to cover the informal words constantly appear in the tweets. Secondly the words in the lexicons have fixed previous sentiment value regardless of the context in which the term is used.

D. Semantic Sentiment Analysis

The sentiment of a tweet is implicitly articulated via the semantics of its words or the semantic relation between them [7]. Traditional sentiment analysis in twitter message is semantically weak because they are not bothering about semantics of words when calculating sentiment polarity in the tweet. Semantic sentiment analysis is categorized into two types: contextual semantic sentiment analysis and conceptual semantic sentiment analysis.

1) Contextual Semantic Sentiment Analysis

The key emphasis behind the contextual semantic sentiment analysis comes from 'You should know a word by the company it keeps'[8]. Contextual semantic inferred from the co-occurrence patterns of words. To change the context may lead to change the word's sentiment. Consider the sentences 'The movie is super hit' and 'The animation movie is super flop'. 'Super' is the significant in both sentences based on lexicon based approach. The

word 'super' provides positive sentiment in the context of 'hit' and negative sentiment in the context of 'flop'. The sentiment is changed based on neighbouring word.

2) Conceptual Semantic Sentiment Analysis

Conceptual semantics is often extracted from external knowledge sources such as ontologies and semantic networks [8]. In our proposed method first semantic concepts about Oman tourism has been developed. Then we use Alchemy API tool to retrieve entity extraction from the tweet and semantic concept mapping.

We consider conceptual semantic as feature which can be combined with supervised machine learning approach to enhance the performance of twitter sentiment analysis. We use Naïve Base supervised machine learning classifier, since it is simple, easy to implement and use.
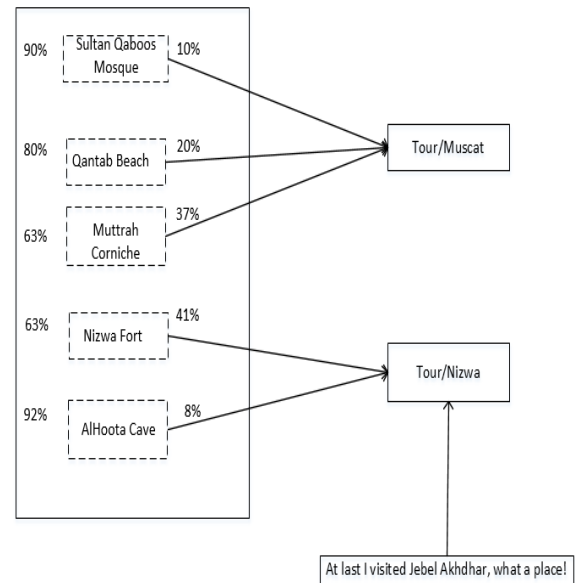

Fig. 2. Incorporating concpetual semantic concept with sentiment anlayis for Oman Tourism with positive /Negative sentiments

Fig. 2 shows the effectiveness of semantic concept as feature for sentiment training classifier. In this fig. 2 the left most box entities are appeared in the training data set. Positive and negative percentage of occurrence probability of these entities in tweet data set also shown in the box. For example, the entities 'Muttrah Corniche', 'Sultan Qaboos Mosque', 'Qantab beach' are more frequently appeared as positive sentiment in tweet data set and they all mapped to the semantic concept 'Muscat'. Similarly 'Nizwa Fort' and 'AlHoota Cave' are most probably appeared as positive sentiment and mapped to the semantic concept 'Nizwa'.

For example consider the tweet 'At last I visited Jebel Akhdhar, what a Place' from the test data set. Actually no sentiment word in this tweet. Using conceptual semantic concept we can find the

Method	Precision	Recall	F1
Entity Specific Opinion Extraction Method	75.45	74.23	75.44
Combined Lexicon Base Method	79.23	76.56	79.43
Conceptual Semantic Sentiment Analysis method	83.34	84.23	85.54

sentiment of this tweet. This tweet is most probably to have positive polarity since it contains the entity 'Jebel Akhdar' (which is not occurring in the training data) is mapped with 'Nizwa'. We know semantic concept 'Nizwa' shows majority positive sentiment, so the above tweet also offers positive sentiment.

Algorithm 1 Incorporate conceptual semantics concept to identify the sentiment from the tweet.

Input : The train data set O^{train} and test data set O^{test}

Output : Sentiment polarity from the tweet data set.

Create a new tweet train data set O^{train}

For each tweet t_n in $(t_1, t_2, t_3, ...t_m) \in O^{train}$ do

    Detect the named entities n_i from tweet data set O^{train}

        $N = (n_1, n_2, n_3, ...n_s)$

    Map each entity with concept in the ontology and calculate the polarity.

End for

Create a new tweet test data set O^{test}

For each tweet t_n in $(t_1, t_2, t_3, ...t_p) \in O^{test}$ do

    Detect the named entities n_i from tweet data set O^{test}

        If $n_i \in (n_1, n_2, n_3, ...n_s)$ then

            Retrieve the sentiment polarity

        else

            Use conceptual semantic concept to identify the sentiment.

End for

    Naïve Bayes is the probabilistic classifier whose task is to categorize the given tweet into any one of the sentiment class (positive or negative).

IV. EXPERIMENT AND RESULTS

We use TAGS to get tweeter data set for this experiment. Around 4,432 tweets are retrieved about Oman. We randomly divide 80% of data into training set and 20% of data into testing data. All the entity words extracted from the tweet to reduce their tweet size. We filter username and letter occurring more than one time is replaced with single occurrence of the letter. Easy negation handling method also used to deal the negation problem.

TABLE 1 ACCURACY PERCENTAGE OF VARIOUS METHODS

Method	Precision	Recall	F1
Baseline Method	67.83	65.34	66.62
Domain Specific Ontology Method	71.38	73.42	72.77

Baseline method we consider only any existing lexicon to calculate the accuracy. It yields only 66.62% accuracy. When we syndicate with domain specific ontology is improving the performance the sentiment analysis to 72.77%. In this experiment we extracted entity as feature and additionally those features matched into concepts in the ontology. When we combined lexicon base method the accuracy of the result into 79.43%. It is detected from the experiment that using conceptual semantic sentiment analysis which improved the performance of the domain to 85.54%. .

V. CONCLUSION

In this paper we examine the effect of four factors that is domain specific ontology, entity specific opinion extraction, combined lexicon based approach and conceptual semantic sentiment analysis to determine the sentiment analysis of tweets about Oman tourism. The new approach including conceptual semantic sentiment analysis expressively improve the performance of the sentiment analysis. Future work involves how to enrich the sentiment analysis using contextual and conceptual semantic sentiment analysis together. From our research we conclude that most of the people express positive opinion about Oman Tourism.

REFERENCES

[1] Vallikannu Ramanathan, T.Meyyappan 'An Exhaustive Exploration on Twitter Sentiment Analysis', Journal of Computer Science and Applications, International Research Publication house, Volume 6, Number 1, 2014. ISSN: 2231 -1270

[2] Bing Liu, "Sentiment Analysis and Opinion Mining" Handbook of natural language Processing, 2010.

[3] A.Esuli and F.Sebastiani. 'SentiWordNet: a publicly available lexical resource for opinion mining' in Proceeding of the 5th International conference on Language Resources and Evaluation, pp.417-422, 2006.

[4] Minging Hu and Bing Liu. 'Mining and Summarizing Customer Reviews', Proceedings of the ACM SIGKDD Interenational Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA.

[5] Bing Liu, Minging Hu and Junsheng Cheng. 'Opinion Observer: Analyzing and Comparing Opinions on the Web' Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan.

[6] C.Havasi, R.Speer, and J.Alonso, 'Conceptnet 3: a flexible, multilingual semantic network for common sense knowledge' in recent Advances in Natural Language Processing, pp.27-29,2007.

[7] Erik Cambria 'An Introduction to concept-level sentiment analysis', In Advances in Soft Computing and Its Applications, pages 478-483. Springer, 2013.

[8] John R. Firth. A Synopsis of linguistic theory. Studies in Linguistic Analysis, 1930-1955.

[9] P.Turney 'Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews'. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, 2002.

[10] Vallikannu Ramanathan, T.Meyyappan 'Expeditious approach to identify the product sentiment based on MORQA system and polarity shift mechanism', International Journal of Applied Engineering Research ISSN 0973-4562 volume 12, No.14, 2017 PP 4528-4533.

[11] Erik Cambria, Marco Grassi, Amir Hussain, and Catherine Havasi. 'Sentic computing for social media marketing. Multimedia tools and applications', 59(2):557–577,2012.

[12] G.Vinodhini and RM.Chandrasekaran June 2012 ' Sentiment Analysis and Opinion Mining; A Survey' ISSN:2277 128X, Voluume 2, Issue 6. [4].