Twitter Sentiment Analysis for Predicting Election Results

Chetana Thombare[1] Dr. B. K. Sarkar[2]
[1,2]PVPIT, College of Engineering, Pune, Maharashtra, India

Abstract— Twitter becomes one of the important platforms for interaction. Twitter allows people to have their own accounts to comment, express feelings and convey emotions via texts as well as emoticons. When a certain issue is discussed, monitoring such information becomes difficult since there are too many suggestions and the problems usually tend to be overlooked. Thus, this paper aims to identify the opinion mining and sentiment analysis components for extracting both English and Malay words in Twitter. Information, in terms of texts, are extracted and clustered into emotions.

Key words: Sentiment Analysis, Natural Language Processing, Training Data

I. INTRODUCTION

Sentiment analysis aims to detect opinions expressed regarding a given subject or topic from text. With the rapid growth of social media platforms such as micro blogging services, social networking sites and short messaging services, people increasingly share their views and opinions online. As such, sentiment analysis has attracted much attention since opinions or sentiments detected from text are potentially useful for downstream applications including recommender systems, social network analysis, market forecasting and the prediction of political topics.

In product reviews, it is observed that the distribution of polarity ratings over reviews written by distinct users on different products are often skewed in the real world. Opinion mining refers to the application of text mining, computational linguistics, and NLP, to identify or classify the opinion expressed in text message is either positive or negative. As such, incorporating user and product information would be helpful for the task of sentiment classification of reviews.

The sentiment analysis can be performed either at document level, at sentence level, or at aspect level. At the document level, opinion mining consists of identifying the overall sentiment polarity as expressed in a review.

A. Sentiment Analysis using Different techniques

Sentiment analysis perform using following techniques,

1) Word Wise Sentiment Analysis

In word based approach the criteria of selecting Twitter sentiments with the presence of words that express sentiments such as good, bad, excellent, trouble and etc. From these words it is possible to infer the sentiment present in the text. These words are used to determine the Twitter sentiments and must be created for each sentiment that is positive or negative according to the applications.

B. Using SentiWord Dictionaries

SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.

1) SentiWordmeans

- A general thought, feeling, or sense.
- Feelings, especially tender feelings, as apart from reason or judgment.
- Gentle or tender feelings, sometimes of a weak or foolish kind.

C. Using WordNet Dictionaries

WordNet is a combination of dictionary and thesaurus. It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members.

D. Using NLP & Sentiment Dictionaries NLP

Natural language processing is ontology-assisted way of programming in terms of natural language sentences.

Natural language processing (NLP), is a branch of artificial intelligence that concerned with automated interpretation and generation of human language.

1) POS Tagging Wise

Part-Of-Speech (POS) detects if the word token is noun, verb, and adjective. The word is assigned in accordance with its syntactic functions. In English the main parts of speech are noun, pronoun, adjective, determiner, verb, preposition, adverb, conjunction, and interjection.

2) NER Detection Wise

Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names. Classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.

II. EMOTICON WISE SENTIMENT ANALYSIS

In this approach the criterion to select Twitter sentiments for classification is the presence of at least one emoticon. Based on emoticons used in the message it is possible to infer the sentiment of the text

The Emoticon data set is created by collecting status and comments with positive ':)' and negative ':(' emoticons. In this approach messages are classified based on positive and negative emotions. Using Unicode's we can do sentiment analysis.

Emoticons	Feeling	Sentiments
:) :-)	Happy	Positive
:(:-(	Sad	Negative
:D :-D	Very Happy!	Positive
D: D=	Very Sad	Negative
_ *.* *-*	Fascinated	Positive
D:< D: D8	Horror, disgust, sadness	Negative
xD XD	Laughing, big grin	Positive
:| =| :-|	Straight face no expression	Neutral

Table 1: Sample of Emoticons Used.

III. LITERATURE REVIEW

"Mauro Dragoni, Fondazione Bruno Kessler" et al Both the authors give the brief information about the tool that are for enabling the building of multi-domain sentiment model that gives the linguistic overlaps between domains for inferring document polarity of each domain. It makes the use of a deep learning architecture using distributed vectors to represent words. It can be used for calculating the polarity of a given text by using domain-specific information. Dranziera protocol is used for evaluating the performance of the system.

"LakshmishKaushik, AbhijeetSangwan and John H. L. Hansen" et al LakshmishKaushik et al. proposed a new method for recognizing sentiment in audio using keyword spotting (KWS) for sentiment detection with the use of audio obtained from videos in youtube.com and UT-Opinion corpus. The system uses the iterative methodology to automatically extract sentiment behavior keywords from text. The Maximum Entropy based approach is used to enlarge the sentiment classifier.

"Yoonjung Choi, JanyceWiebe, and Rada Mihalcea" et al The knowledge-based +/-effect coarse-grained sense disambiguation method to convey the opinion towards +/-effect of the word sense depending on the surrounding context that positively or negatively affect entities. The system uses the selection preferences which are modeled using Latent Dirichlet Allocation (LDA).The system uses WordNet information is used to determine whether an instance of a word in the corpus is being used with a +effect, -effect, or Null sense.

"RuiXia, Jie Jiang, and Huihui He" et al A distantly supervised lifelong learning approach, for large-scale social media sentiment analysis is proposed in for addressing distant supervision work in terms of continuously increasing and constantly changing topics. The continuous sentiment learning in social media can learned the knowledge from past tasks, and continuously update the knowledge as new tasks appear. The system can adapt any single-task sentiment learning algorithms to the scene of lifelong sentiment learning. The lifelong sentiment classifier is evaluated on nine benchmark datasets.

Fuhai Chen, Rongrong Ji, Jinsong Su, Donglin Cao, and Yue Gao, et al gives a Weakly Supervised Multi-modal Deep Learning (WS-MDL) scheme towards robust and scalable sentiment prediction. The system learns convolutional neural networks iteratively and selectively from "weak" emoticon labels, which mostly contains noise. A probabilistic graphical model is explained to filter out the emoticon labels serve as noisy labels to be used and evaluated during training that simultaneously learn discriminative multi-modal descriptors and infer the confidence of label noise.

IV. ADVANTAGES

Researchers can find out sentiment or market trend for a celebrity.

V. APPLICATIONS

1) Political exit poll prediction
2) Best visiting places recommendation
3) Movies comparisons and calculate ratings.

VI. CONCLUSION

This work begins with transforming unstructured information into meaningful lexicons after extracting the Twitter's contents. All of the meaningful lexicons are stored in a database after manual identifications are carried out. With sentiment analysis, emotions are classified into happy (positive), unhappy (negative) and emotionless. The results are displayed by giving the percentage of sentiment categories so that it can be concluded that a selected Twitter post get positive or negative responses based on all comments received from users. As a case study, an issue on an examination results is posted and results of students' responses are determined. Th is study is significant of enabling the stakeholders such as administrators and businessmen to monitor any discussion issue for enhancing their services.

REFERENCES

[1] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub,"Arabic sentiment analysis: Lexicon-based and corpus-based," in AppliedElectrical Engineering and Computing Technologies (AEECT), 2013IEEE Jordan Conference on. IEEE, 2013, pp. 1–6.
[2] R. Duwairi, R. Marji, N. Sha'ban, and S. Rushaidat, "Sentiment analysis inArabic tweets," in Information and communication systems (icics),2014 5th international conference on. IEEE, 2014, pp. 1–6.
[3] A. Assiri, A. Emam, and H. Aldossari, "Arabic sentiment analysis: Asurvey," vol. 1, no. 6, 2015, pp. 75–85.
[4] M. Abdul-Mageed, M. T. Diab, and M. Korayem, "Subjectivity andsentiment analysis of modern standard Arabic," in Proceedings of the49th Annual Meeting of the Association for Computational Linguistics:Human Language Technologies: short papers-Volume 2. Associationfor Computational Linguistics, 2011, pp. 587–591.
[5] M. Nabil, M. A. Aly, and A. F. Atiya, "Astd: Arabic sentiment tweetsdataset." in EMNLP, 2015, pp. 2515–2519.
[6] M. Abdul-Mageed and M. T. Diab, "Sana: A large scale multi-genre,multi-dialect lexicon for Arabic subjectivity and sentiment analysis." InLREC, 2014, pp. 1162–1169.
[7] B. Pang and L. Lee, "A sentimental education: Sentiment analysis usingsubjectivity summarization based on minimum cuts," in Proceedings ofthe 42nd annual meeting on Association for Computational Linguistics.Association for Computational Linguistics, 2004, p. 271.
[8] M. Lan, S.-Y. Sung, H.-B.Low, and C.-L. Tan, "A comparative studyon term weighting schemes for text categorization," in Neural Networks, 2005. IJCNN'05.Proceedings.2005 IEEE International Joint Conference on, vol. 1. IEEE, 2005, pp. 546–551.

