

Priority Based Sentiment Analysis for Quick Response to Citizen Complaints

Kamalakshi V. Deshmukh

Department of Computer science
Marathwada Mitra Mandal's college of Engineering
Pune, India
kamalakshideshmukh.comp@mmcoe.edu.in

Prof. Sankirti S. Shiravale

Department of Computer Engineering
Marathwada Mitra Mandal's College of Engineering
Pune, India
sankirtishiravale@mmcoe.edu.in

Abstract—Today metropolitan citizens need a common platform to register their complaints. In the traditional telephonic system task of complaint registration is very time-consuming process so, citizen have to wait until call is received by the service executive. The proposed framework is developed for Pune Municipal corporation (PMC) will be very helpful for registering queries in natural language and get immediate response.

Understanding the short text is the main challenge of the system as short text do not follow syntax of the written language, short text does not have sufficient statistics to support approach of text mining, short text is noisy and ambiguous. So, traditional Part-of-Speech (POS) tagging tools cannot be easily applied. In proposed framework for understanding natural language semantic knowledge provided by well-known knowledgebase WordNet is used. In prequery citizens inserts complaint to system and get immediate response to query with the help of knowledgebase and machine learning algorithm. In postquery system analyses the citizen sentiment to handle grievance level and accordingly prioritize the citizens by sentiment analysis. The proposed framework will help many organizations to ensure quality service provision and customer satisfaction with less human efforts.

Index Terms— Short text understanding, WordNet, Sentence Similarity, text segmentation, NLP, type detection, concept labelling, Sentiment Analysis.

I. INTRODUCTION

In Computer Science Natural language processing (NLP) is emerging field where interactions between human and computers takes place. There are many challenges in natural language processing like speech recognition, natural language understanding, to understand human expressions or emotions and question answering systems or some combinations of above.

Information burst pinpoints the requirement for machines to well understand natural language texts. In this project, we emphasis on short texts which denote texts with limited context [1]. Short text is like tweet, microblogging services, user entered queries or web search etc. are generated in huge amount so, there is necessity to tackle short texts. As a result, improved understanding of short texts will bring great value. The sentiment analysis is typical problem in the NLP. The process of analyzing and summarizing the opinions expressed

in these huge user generated data is usually called Sentiment Analysis or Opinion Mining which is a very interesting and popular domain for researchers nowadays [20].

In the today's fast world no one has time to rely on the telephonic system where one has to wait long time to register their complaint. The proposed system is based on the concepts of NLP, sentiment analysis and machine learning. This system is developed for the resolving queries of the citizens in natural language. Also giving the priority to the citizens request based on the intensity calculated by sentiment analysis. Citizens enter the complaints or queries are considered to be short text. Correct interpretation of short text semantically and sentiment analysis of the user entered text are two important tasks of this framework.

II. RELATED WORK

To correct interpret the short text has many challenges like short texts does not follow syntax of written language, short text does not have sufficient statistics to support for approaches for text mining, short text is ambiguous and noisy. Therefore, traditional natural language tools such as POS tagging cannot be easily applied. In this system for understanding natural language processing semantic knowledge present in the recognized knowledgebase and sentence similarity measure can be used [1]. Wen Hua et.al proposed framework for considering semantic knowledge in order to understand the short text. Short text is noisy, ambiguous and are produced in massive amount thus it adds another trouble to tackle them. In this paper they have used knowledge intensive approach for tasks like text segmentation, NER, concept labelling and type detection for understanding short text [1].

Study of views and sentiments expressed in texts by means of text classifying algorithms is called as sentiment analysis [20]. we will consider the basic definition of sentiment as giving positive or negative view. "I liked the music concert" will be treated as positive mood however "Concert was having very poor performances" conveys a negative comment or mood. "I am going to my college" can be considered as neutral comment as it does not convey any moods. In the sentiment analysis many times we need to classify the text according to the sentiment polarity positive, negative or

neutral [20]. Hence, it is treated as text classifying problem and its main task is to label the texts among different classes by using various algorithms. Recently a lot of work has been done on the on the sentiment analysis with different approaches depends on the relation to the particular domain. Lexicon based and machine learning are two main approaches in sentiment analysis. Survey conducted in paper [8] covers popular techniques in sentiment analysis.

- Lexicon based approach: It is used in the [17], [19], [20] to determine the polarity of the tweets, news, facebook data respectively. In this approach the text is analyzed on the basis of the predefined lexicon library which produces the sentiment score of the given text. The key benefit of the lexicon based approach is that it does not need any labeled data set for classifying the text.
- Machine Learning approach: To determine the polarity of sentiment on the basis of trained and test data set machine learning approach is used. In machine learning approach we have learning phase and actual classification phase. In the learning phase it trains the model using the test data followed by classification phase. Resulting model does the actual work of classification. ML algorithms such as SVM, Maximum entropy, naïve bayes etc. can be useful for this task. The benefit of this method is the capability to adjust and create new trained models for specific application. Drawback of this approach is that it requires the availability of labeled data that could be very expensive. It also requires more time to train the model in the learning phase [3], [5], [6], [10], [12].

Support Vector Machine Classifier (SVM): This method is used to categorize the search space linearly for the separation of two dissimilar classes. SVM is supervised method which need labelled data set to train the model [21]. Vector representing optimum margin hyper plane is prime motive behind SVM. SVM is used in [3], [5], [6], [10], [12] for sentiment analysis for the data such as e-commerce data, internet movie database, comments on clothing product, stock market data and restaurant review data respectively.

Naïve bayes Classifier(NB): This method is based on the probability that some character in text is member of the specific label to determine the polarity of the text [21]. It uses bag of words (BOW) model for feature extraction. This model assumes that all the features are independent [12]. Generally, it is used when the dimension is very high. Naïve bias is used for classification of restaurant reviews [12].

Maximum Entropy(ME): This is probabilistic classifier which is built on the concept of Maximum Entropy. ME can be used when we don't have any assumption regarding the data and is not suitable for any assumption. It does take into consideration independence of the characteristics in data. As our data is words so, we can't consider words are independent. ME converts the labelled characters into vectors which is then used to determine the polarity of the text [5], [21].

TABLE I. COMPARATIVE STUDY OF DIFFERENT TECHNIQUES FOR SENTIMENT ANALYSIS

Author	Year	Algorithm	Data set
Hanhoon Kang et.al [12]	2012	Naïve bayes, SVM	Restaurant reviews
A. Moreo[19]	2012	Lexicon based	News
Alvaro Ortigosa et.al.[20]	2013	Lexicon based, machine Learning	Facebook data
Zhang et.al.[6]	2015	SVM	Chinese comments given on clothing products
Luo et.al.[10]	2015	SVM, NB, Decision Tree (DT)	Stock market data
Niu et.al.[11]	2016	BOW characteristic with TF and TF-IDF approach	Manually annotated Twitter data
Abinash Tripathy et.al.[5]	2016	SVM, Maximum Entropy, SGD, N-gram, NB	Internet Movie Database (IMDb)
Zimbira, David et.al [17]	2016	Lexicon based approach	Tweets
Paramita Ray and Amlan Chakrabarti [15]	2017	Lexicon based approach	Product Review
Hari Krishna M, Rahamathulla K et. Al[3]	2017	SVM	e-commerce data

Table 1 shows the comparative study of the different sentiment analysis techniques.

III. PROPOSED ARCHITECTURE

In the system architecture there are two main building blocks given below.

1. NLP (Natural language Processing)
2. Sentiment analysis

NLP part is used for the understanding of the short text and to get the semantic meaning of the short text whereas Sentiment analysis is component used to check the emotions or feeling of the citizens. Figure 1 shows components of the proposed architecture.

A. Natural Language Processing

There are given main steps to understand the short text.

- Text Segmentation - It is the process by which whole short text will be separated into words and phrases. Here we need to find the most semantically intelligible segmentation.
- Type Detection - In this step we need to find the best type of the word grammatically like subject, verb, object etc.

- Concept Labelling - In this step detect the concept underlying to each instance (e.g. India refers to country).
- NER (Named entity recognition) – It categorize the terms in predefined classes by locating the named entities. (e.g., Organization, States, people, period etc.)

Ambiguities such as ambiguous segmentation, Noisy short text, ambiguous type, ambiguous instance will be handled in the system with help of above steps and knowledgebase WordNet.

B. Sentiment analysis

In the prequery the queries matching the stored database will be answered to the user. In the postquery it will be transferred to the admin where sentiment analysis is done. Sentiment analysis will determine polarity of the citizens inserted short text after that intensity will be calculated on the basis of the sentiment analysis text score. In this case we are using SVM for sentiment analysis.

C. Flow of the System

- 1) User types a phrase in the system.
- 2) System sends the phrase to machine learning NLP engine.
- 3) NLP engine extracts user's intent and entities from given phrase and send back to system.
- 4) Intent is used to call upon proper service, using entity information to find proper data.
- 5) Data will be returned to the user.
- 6) System packages data in proper response for display to client.

Figure 2. shows flow diagram of the proposed system.

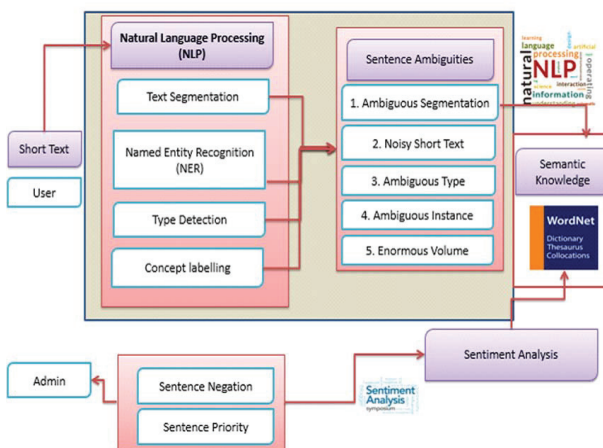


Fig. 1. System Architecture

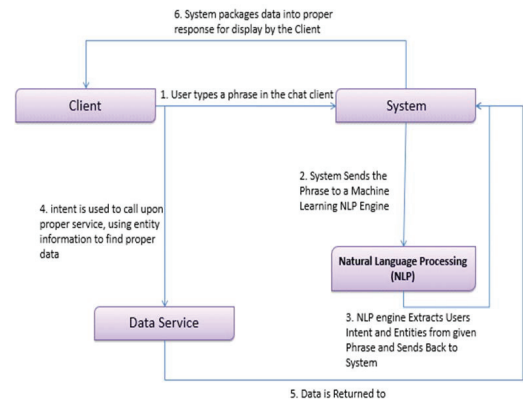


Fig. 2. Flow of the System

IV. ALGORITHM STRATEGY

A. Step1: NLP (Natural Language Processing)

To understand the short text inserted by the user is having 3 important steps.

- Text segmentation is done in the first step where the short text inserted by the user is divided into words.
- Type detection is the next step where the type of every word is determined grammatically like subject, object and verb.
- Concept labeling is the step where the words are mapped into the particular concept in from the real world.

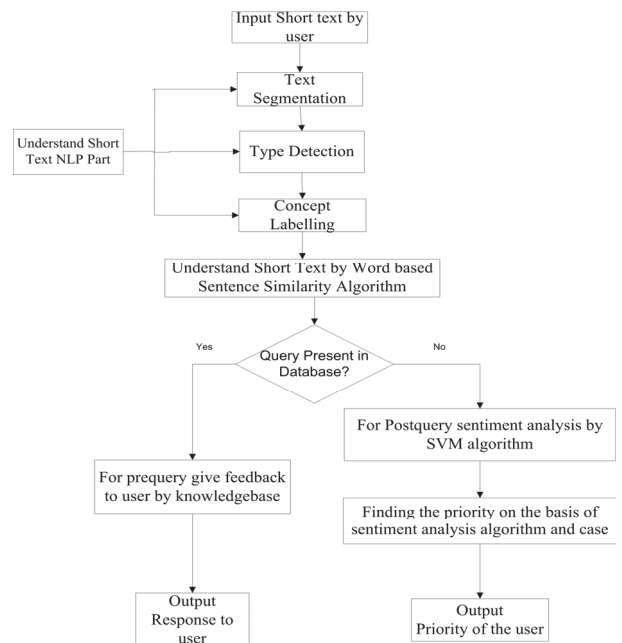


Fig. 3. Algorithm Strategy

- Sentence similarity is calculated with the help of the word-based sentence similarity algorithm and WordNet.
- Sentence similarity is the measure to detect whether two sentences have the same meaning or looking for same subject.

At this stage recognize the meaning of the short text is completed.

B. Stage 2: Sentiment Analysis and Priority decisions

- Later stage output 1 is the Prequery gives feedback to the user.
- In the postquery sentiment analysis is done on the understood short text by SVM classifier to classify intent of the user.
- Calculate the intensity (negativity) of the user input text in percentage on the basis of the sentiment analysis score.
- Priorities of the users for feedback given with the help of below cases. (Assume R1 and R2 are two requests are present.)

Case 1: Between R1 and R2 two requests, request which is more intensive will get priority.

Case 2: If $R1=R2$ (Percentage of Intensity) then the request which is inserted first will get priority.

Case 3: There might be cases where lower intensive request not get chance of feedback from administrator So, threshold time is decided when user reach threshold then priority will be assigned to that user.

V. CREATION OF DATASET

The static database for the system will be the questions inserted by the user and answers are given by the system. The dynamic database is the questions other than static database which will be handled by administrator.

Q.1. Who will be POC (Point Of Contact) for registering complaint about invasion on red road or Footpath ?

Ans. To PMC ward office's Assistant Commissioner Municipal Corporation in written or Email in detail with address.

Comments - This question is having ambiguous instance as red road, here it means danger area of road not colour of road.

Q.2. What is the plan of action against illegal invasion by PMC and when should we expect concrete action after submitting complaint?

Ans. Survey of the location followed by seizing their material. Action will be taken within 7 to 15 days according to type of invasion.

Comments – PMC means Pune Municipal Corporation. Here the Noisy short text is present.

Q.3. Can we put private ads or banners on PMC owned assets?

Ans. Prior approval is needed to put any private ads on PMC authorized assets.

Comments- ads means Advertisement. Noisy short text ambiguity is present.

Thus, above question has some type of ambiguity which will be handled by the proposed algorithm.

VI. CONCLUSION

In this paper we have proposed a framework for the Pune Municipal corporation where citizens get platform to register their queries and get immediate feedback. Sentiment analysis is done on the citizens inserted query to get intent of the citizens and to prioritize them on the basis of the sentiment analysis score. This system will minimize the time and effort of the citizen as well as PMC to better serve their citizens. We have prepared the static dataset and in the future scope, we are looking forward to implement this system for PMC.

ACKNOWLEDGMENT

I am profoundly grateful to Prof. Sankirti Shiravale for her expert guidance and continuous encouragement throughout completion of this paper.

REFERENCES

- [1] Wen Hua, Zhongyuan Wang, Haixun Wang, Kai Zheng and Xiaofang Zhou, "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge" IEEE transaction on knowledge and data engineering vol.29, no. 3, march 2017
- [2] Taesung Lee, Zhongyuan Wang, Haixun Wang, Seung-won Hwang, "Attribute Extraction and Scoring: A Probabilistic Approach" IEEE conference on data engineering 2013
- [3] Hari Krishna M, Rahamathulla K and Ali Akbar "A Feature Based Approach for Sentiment Analysis using SVM and Coreference Resolution" IEEE International Conference on Inventive Communication and Computational Technologies 2017
- [4] Monisha Kanakaraj and Ram Mohana Reddy Guddeti, "NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers" IEEE International Conference on Signal Processing, Communication and Networking (ICSCN) 2015
- [5] Abinash Tripathy, Ankit Agrawal, Santanu Kumar Rath, "Classification of sentiment reviews using n-gram machine learning approach" Elsevier 2016
- [6] Zhang, D., Xu, H., Su, Z., & Xu, Y. "Chinese comments sentiment classification based on word2vec and svm perf." Expert Systems with Applications. vol.42 (4), pp. 1857–1863, 2015
- [7] Neethu M S and Rajshree R "Sentiment Analysis in Twitter using Machine Learning Techniques" IEEE Computing, Communications and Networking Technologies (ICCCNT) IEEE 2013
- [8] Alessia D' Andrea, Fernando Ferri, Patrizia Grifoni and Tiziana Guzzo, "Approaches, Tools and Applications for Sentiment Analysis Implementation" International Journal of Computer Applications, vol. 125, No.3, pp. 0975 – 8887, September 2015

- [9] Liu, S. M., & Chen, J.-H., "A multi-label classification based approach for sentiment classification" *Expert Systems with Applications*, vol. 42 (3), pp.1083–1093,2015
- [10] Luo, B., Zeng, J., & Duan, J, "Emotion space model for classifying opinions in stock message board" *Expert Systems with Applications*, vol. 44, pp.138–146 ,2016
- [11] Niu, T., Zhu, S., Pang, L., & El Saddik, A," Sentiment analysis on multi-view social data", In *Multimedia modelling*, Springer, pp 15–27,2016
- [12] Kang Hanhoon, Yoo Seong Joon, Han Dongil., "Senti lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews", *Expert Syst Appl.*, vol.39, pp.6000–10, 2012
- [13] C. Anantaram and Amit Sangroya "Identifying Latent Beliefs in Customer Complaints to Trigger Epistemic Rules for Relevant Human-Bot Dialog", *IEEE International Conference on Control, Automation and Robotics* 2017
- [14] P Balaji, O Nagaraju, D Haritha "Levels of sentiment analysis and its challenges: A literature Review" *Big Data Analytics and Computational Intelligence (ICBDAC)*, 2017
- [15] Paramita Ray and Amlan Chakrabarti, "Twitter Sentiment Analysis for Product Review Using Lexicon Method" *International Conference on Data Management, Analytics and Innovation (ICDMAI)*,2017
- [16] Harpreet Kaur, Veenu Mangat and Nidhi "A Survey of Sentiment Analysis techniques" *IEEE International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)* 2017
- [17] Zimbra, David, M. Ghiassi, and Sean Lee. "Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks." *49th Hawaii International Conference on System Sciences (HICSS)*. IEEE, 2016.
- [18] A. Moreo, M. Romero, J.L. Castro, J.M. Zurita," Lexicon-based Comments-oriented News Sentiment Analyzer system", *Expert Syst Appl* ,vol. 39,pp.9166–80,2012
- [19] Hu Nan, Bose Indranil, Koh Noi Sian, Liu Ling. "Manipulation of online reviews: an analysis of ratings, readability, and sentiments".*Decision Support Syst* ,vol 52,pp 674–84, 2012.
- [20] Alvaro Ortigosa, José M. Martín, Rosa M. Carro "Sentiment analysis in Facebook and its application to e-learning" *Journal of Computers in Human Behavior* 2013
- [21] Walaa Medhat, Ahmed Hassan, Hoda Korashy "Sentiment analysis algorithms and applications: A survey" *Ain Shams Engineering Journal*, Elsevier 2014
- [22] A. Hassan, A. Abbasi, and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework," in *Social Computing (SocialCom)*, International Conference on IEEE, pp. 357–364,2013
- [23] P. T. Ngoc and M. Yoo, "The lexicon-based sentiment analysis for fan page ranking in facebook," in *Information Networking (ICOIN)*, International Conference on. IEEE, 2014, pp. 444–448.
- [24] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro,"Sentiment analysis of facebook statuses using naive bayes classifier for language learning," in *Information, Intelligence, Systems and Applications (IISA)*, pp. 1–6 ,2013
- [25] Sreshmi and Kannan Balakrishnan, "Implementation of an inquisitive chatbot for database supported knowledge bases" *Springer*, vol 41, pp 1173–1178,2016
- [26] Alexandra Balahur, Jesu's M. Hermida, and Andre's Montoyo," Building and Exploiting EmotiNet, a Knowledge Base for Emotion Detection Based on the Appraisal Theory Model" *IEEE transactions on affective computing*, vol. 3, no. 1, January march 2012