International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-5 Issue-1, October 2015

Social Media Mining for Opinion Analysis

Saloni Bindra, Priyanka Karmarkar, Abhishek Kumar Verma, Laxmi Grover

Abstract- Here we describe a method which involves determining the sentiment of a review about Banks by extracting the phrases with a noun-adjective relationship, Identifying if the noun is present in the domain specific Ontology tree and then determining the polarity of the adjective, aggregating the polarity. The results so obtained are thus summarized and then categorized by characteristic feature pertaining to the Bank. This reduces the human efforts to go through them and a result specific to a particular Bank; sub-categorized by Peculiar features of it with polarity alongside each individual characteristic. Thus the fruits of the reviews are gained even without reading them.

Index Terms- Sentiment, Polarity, Domain Ontology, Opinion Mining.

I. INTRODUCTION

Sentiment is a thought, view, or attitude, especially one based mainly on emotion instead of reason. Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space. Opinions are central to almost all human activities and are key influencers of our behaviors. Opinions and its related concepts such as sentiments, evaluations, attitudes, and emotions are the subjects of study of sentiment analysis and opinion mining. Opinion mining can be defined as the sub discipline of computational linguistics that is concerned with the opinion that a document expresses. Sentiment classification is about determining the subjectivity, polarity (positive/negative) and polarity strength (weakly positive, mildly positive or strong positive) of an opinion text. [1]

The society as a whole is evolving, our outlook towards different things are changing. In the earlier days, one used to consult a neighbor or a close relative before buying, selling, investing anything. But today, we use the data available on social media to make such decisions. We don't believe in whispers anymore until it is displayed on a web page. Furthermore there are a lot of reviews available online, in hundreds and thousands, in different websites. Now it is not possible to go through all of them manually, and even if one does, it is very difficult to remember what each review showcased. Automated sentiment analysis systems are thus needed.

Manuscript published on 30 October 2015.
***** Correspondence Author (s)
Saloni Bindra, Department of Computer Engineering, University of Mumbai, Mumbai, (Maharashtra). India.
Priyanka Karmarkar, Department of Computer Engineering, University of Mumbai, Mumbai, (Maharashtra). India.
Abhishek Kumar, Department of Computer Engineering, Sir M. Visvesvaraya Institute of Technology, Bangaluru, (Karnataka). India.
Verma Laxmi Grover, Department of Computer Engineering, Dr. B. R. Ambedkar National Institute of Technology Jalandhar (Punjab). India.

To add to this we have another reason why we have a lot of corporate interest. The field of opinion mining and sentiment analysis is well-suited to various types of intelligence applications-Business Intelligence being one of them. Knowing the reputation of their organization and work or such information about their competitors is highly valuable for marketing and customer relationship management. Opinion analysis can thus replace the costly and time consuming customer surveys, also reducing the time and resources which would be eventually used post the surveys for interpretation. Sentiment-analysis technologies for extracting opinions from unstructured human-authored documents would be excellent tools for handling many business-intelligence. Consider a scenario, where the profits of a Bank are diminishing even when the market is burgeoning. In this case, knowing the root cause of this problem becomes very prominent internet and then display a converged version of individual reviews. This would save us from having to read potentially dozens or even hundreds of versions of the same reviews. Besides reputation management and public relations, one might perhaps hope that by tracking public viewpoints, one could perform trend prediction in growth or other relevant data. Each sentiment bearing sentence contains a topic and attitude towards it. In our research we define a methodology which involves using the relationship between parts of speech in a sentence to identify an individual feature of the subject, use ontology to find the relevance to the topic, finding the sentiment of the attitude using two different methods, aggregating their results and then displaying the them along with their sentiment grouped by Individual feature to a topic.

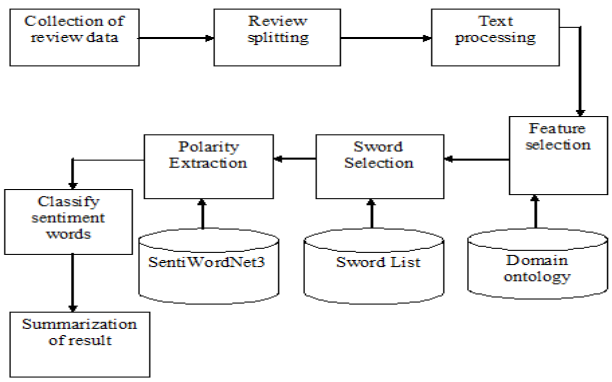

Figure 1: System Block Diagram

II. RESEARCH ELABORATIONS

Following is the Algorithm used to for Opinion Analysis. Given a domain for which we have to find the results.

Select a review and then select a sentence from the review, parse the sentence and then find features using ontology within that sentence. When a target word is found, find all type dependencies for sentence and filter for required Type dependencies. Use Mining Algorithm to find sentiment carrying word along with S-word list. Extract polarity using SentiWordNet3. Then assign net polarity to target word. Next, Classify and aggregate result.

The processing can be categorized into the following sections a)Parsing and Finding type dependencies b)Individual Feature identification c)Filtering the type dependencies d)Apply Mining Algorithm e) Assigning the polarity.

A. Parsing and Tokenization

The end of the sentence is identified by a de-limiter and is the sent to the parser. The output of the parser gives us the parts of speech for each word and also the relationship between them. This relationship between two words is later used to find the sentiment carrying word. This word is then used in the Senti Word Net to extract the polarity.

Consider the example below:

I found the bank manager to be very amiable man.

i	found	the	bank	manager	to	be	very	amiable	man
pronoun	verb	article	noun	noun	infinitive	verb	adjective	adjective	pronoun

Stanford dependency representation allows each word to have multiple governors and parsers may generate a different number of dependencies for each sentence. The dependency parsers require that the data is part-of-speech tagged. The Stanford typed dependencies representation was designed to provide a simple description of the grammatical relationships in a sentence that can easily be understood and used by people without linguistic expertise who want to extract textual relations. In particular, rather than the phrase structure representations that have long dominated in the computational linguistic community, it represents all sentence relationships uniformly as typed dependency relations.[7] We have to find the relationship between the target word and the attitude described towards it in the sentence. The parser type dependencies used are described as below:

Acomp: adjectival complement,

An adjectival complement of a verb is an adjectival phrase which functions as the complement Tina's behavior appears to be very bad. Acomp (appears, bad).

Advmod: adverbial modifier

An adverbial modifier of a word is a adverb or adverbial phrase that modifies the meaning of the word.

Call softly. Advmod (softly, call)

Agent: agent

An agent is the complement of a passive verb which is introduced by the preposition \by" and does the action.

The dog has been fed by the owner. Agent (owner, fed)

Amod: adjectival modifier

An adjectival modifier of a word is a adjective that modifies the meaning of the word.

The bird is well trained to speak. Amod (trained, well)

Ccomp: clausal complement

A clausal complement of a verb or adjective is a dependent clause with an internal subject which functions like an **object of the verb.**

I like the fact that you are honest" ccomp (fact, honest)

Neg: negation modifier

The negation modifier is the relation between a negation word and the word it modifies.

Toby is not ugly Neg (ugly, not)

Nn: noun compound modifier

A noun compound modifier of an NP is any noun that serves to modify the head noun.

We entered a well-lit room. Nn (well-lit, room)

Nsubj: nominal subject

The governor of this relation might not always be a verb: when the verb is a copular verb, the root of the clause is the complement of the copular verb, which can be an adjective or noun.

The employee is descent" nsubj (descent, employee)

Xcomp: open clausal complement

An open clausal complement (xcomp) of a verb or an adjective is a clausal complement without its own subject, whose reference is determined by an external subject.

He says that you like to swim" xcomp (like, swim)

B. Finding the feature bearing word using Ontology

Ontology is an explicit specification of a conceptualization [8]. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. The main aim of ontology is to provide knowledge about specific domains that are understandable by both the computers and developers. Ontology improves the process of information retrieval and reasoning thus results in making data interoperable between different applications.

Here each noun identified by the parser is checked with the Ontology and if a match is found, this noun becomes our target feature bearing word. This is then sent later forward to processing to find the sentiment related to this word.

Consider our domain in research to be Banking and Financial Institutions. Here we create an ontology tree for each feature that a bank represents. For aggregation purposes we categorize them into four main sections: Charges, Customer Service, Internet Banking and Miscellaneous where "Bank" is our main root node.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

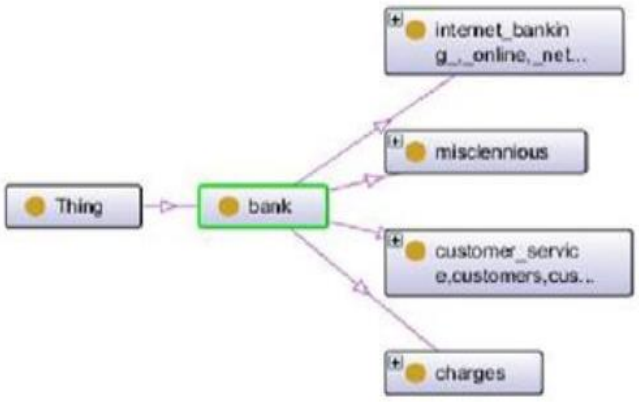

Figure 2: Ontology- Bank

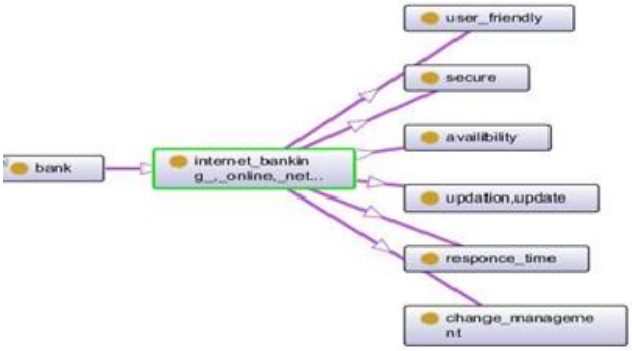

Figure 3: Ontology-internet banking [9]

C. Identifying the sentiment determining word

We have used to parallel ways to determine the sentiment carrying word and find its polarity. Firstly, we use the relationship between the target word and an adjective/noun/adverb. If it matches with the above mentioned dependencies type, then the word other than our feature bearing word is our winner. This word is then used to determine its polarity using Senti Word Net. SentiWordNet is a lexical resource for opinion mining. It assigns to each synset of WordNet three sentiment numerical scores, positivity, negativity and objectivity, describing how Positive, Negative and Objective the terms contained in the synset are. Each of the three scores ranges from 0.0 to 1.0, and their sum is 1.0 for each synset. [7]
The other parallel method is by using the Mining Algorithm [8] and the SentiWordList. The mining algorithm is described as below:

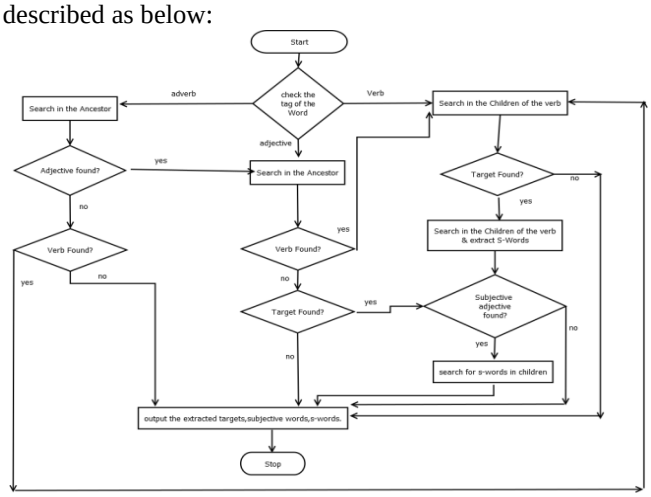

Figure 4: Mining Algorithm Flowchart

Sword List is used to find polarity of sentiment carrying word. Using Emotional Look-Up Table in SentiStrength, database of Sword List is created. Certain domain specific emotional words are been added up. SentiStrength estimates the strength of positive and negative words, even for informal language. It has human-level accuracy for short social web texts in English. SentiStrength reports two sentiment strengths:-1 (negative) to -5 (extremely negative) and 1 (positive) to 5 (extremely positive).[10] Senti Strength also provides Booster word list which changes polarity level of sentiment Word. Booster word list also has score ranging from positive to negative.
In case the sentence contains sentiment shifters, the entire polarity of the Feature and the sentiment carrying word is reversed. These are expressions that are used to change the sentiment orientations, e.g., from positive to negative or vice versa [1]. They are handled by using a negation word list.

Examples:

Input	Bank is good but manager is pathetic.
Output	feature : bank SentiWord : good score : 0.6337632198238539 Senti Strength : 2 feature : manager SentiWord : pathetic score : -0.6138181818181818 Senti Strength : -3

Input	Internet banking is very slow.
Output	feature : internet banking SentiWord : slow score : -0.08843537414965988 Senti Strength : -1

Input	Executives don't respond well.
Output	feature : executives SentiWord : well score : -0.36140805833892525 Senti Strength : -1

Input	They don't provide good service.
Output	feature : customer service SentiWord : good score : -0.6337632198238539 Senti Strength : -2

III. RESULTS

Opinion Mining is useful to the users to view the interpretation of the reviews without actually reading it. The system takes the review in the form of a text file and processes it and it is then passed to the parser where each word in the sentence is tagged. Based on the particular tag given to each word, nouns are sent to the Ontology where the feature determining word is found.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Later Sentence with the feature carrying word is processed and filtered using type dependencies and mining algorithm is run on the result so obtained. SentiWordNet and SentiStrength are used to determine the score of the sentiment carrying word. The result so obtained at document level obtained is aggregated and converged to a particular bank.

Our research is based upon two aspects. One being that the user does not have to read hundreds of review to make his/her judgment towards a bank, Instead can go through our analysis result, which is indeed based upon the sentimental analysis of these hundreds of review for that bank. Another aspect being that the user can enter his/her review and its scoring can be dynamically be generated and displayed respectively.

In process of applying sentimental analysis to these user generated reviews, we score each review based on the sentiment and in a categorized manner. Scoring is done and distributed in the review based on the feature that each bank provides, which in turn help the user to decide and make their analysis work easy. For scoring purpose, two approaches are used. One is being getting the score from SentiWordNet and other being SentiStrength. SentiStrength is our created domain specific list of sentiment carrying words with their score. The result through both these approaches is show in the diagrams:

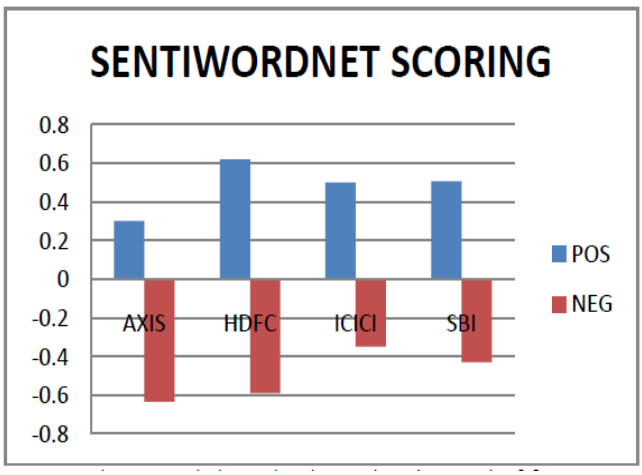

Figure 5: Analysis Result Using SentiWordNet Scoring [9]

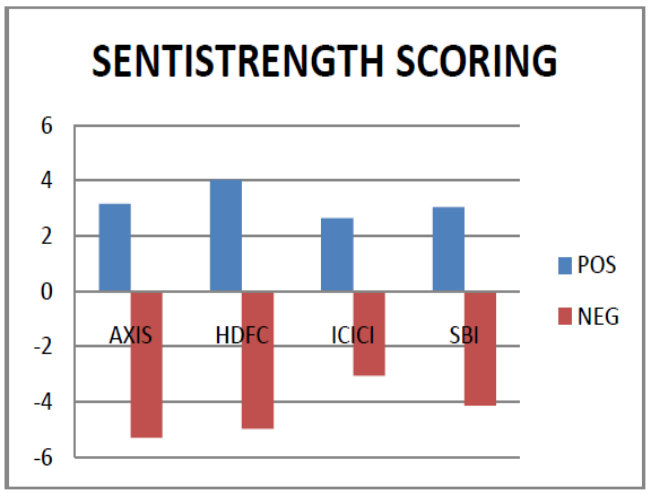

Figure 6: Analysis Result Using SentiStrength Scoring [9]

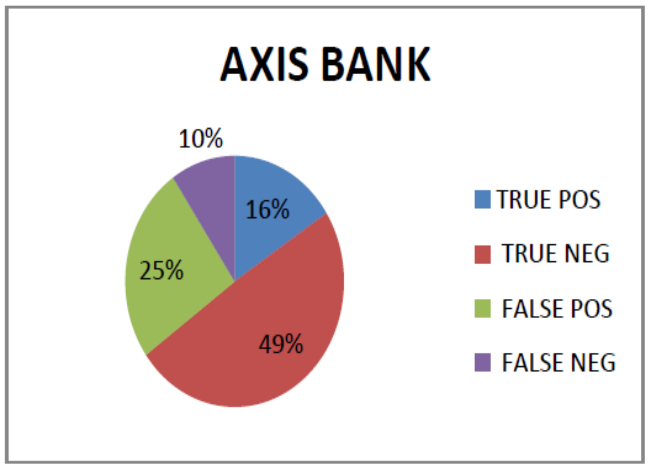

Figure 7: Analysis Result for Axis Bank

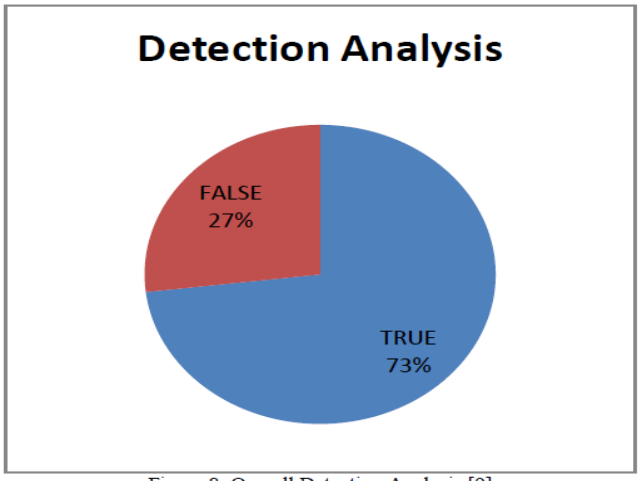

Figure 8: Overall Detection Analysis [9]

IV. CONCLUSION

We address the Sentiment Analysis problem from the end user's perspective. With hundreds of reviews about single entity, it is actually not feasible to go through all so as to get useful information. We proposed the combination approach of Parsing and Tokenizing, Domain Ontology intend to enhance the sentiment classification, by using this approach we can view the strength or weakness of the individual features of a Bank in more detail In future enhancements can be made so as to increase its features. The enhancements for future that could be added are a) Comparative review b) Scoring neutral reviews c) Finding the polarity of slang words used.

ACKNOWLEDGMENT

Our ideas on Sentiment Analysis developed when I was a student at Shan & Anchor Kuttchi Engineering College. We are indebted to Prof. Deepshikha Chaturvedi for her co-operation. We are grateful to her for her continuous support for the project, from initial advice & contacts in the early stages of conceptual inception & through ongoing advice.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

REFERENCES

1. Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.
2. WeiWei, Jon Atle Gulla," Sentiment Learning on Product Reviews via Sentiment Ontology Tree", Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 404–413, Uppsala, Sweden, 11-16 July 2010.
3. Vipin Kumar, Sonajharia Minz, " Mood Classification of Lyrics using SentiWordNet", 2013 International Conference on Computer Communication and Informatics (ICCCI - 2013), Jan. 04 – 06, 2013, Coimbatore, INDIA
4. "Stanford typed dependencies manual", Marie-Catherine de Marne_e and Christopher D. Manning, September 2008.
5. Zhongchao Fei,Xuanjing Huang, and Lide wu, "Mining the Relation between Sentiment Expression and Target Using Dependency of Words", Department of Computer Science and Technology, Fudan University, Shanghai, 200433, China.
6. "Reviews" [online] Available http://www.mouthshut.com/
7. SentiWordNet"[online]Availablehttp://sentiwordnet.isti.cnr.it
8. PDFforOWL"Available http://protege.stanford.edu/plugins/owl/publications/DL2004-protege.owl.pdf
9. "Customer sentiment on bank" Shalu Chopra, Deepshikha Chaturvedi.
10. Sentiment strength detection in Informal Texts, Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., & Kappas, A. (2010).

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)