

Distortion Detection on Online Social Networks using Proficient Sentiment Analysis for Banking Institutions

¹N.Sumathi , ²Dr.T.Sheela

¹Research Scholar,(Part-time Ph.D Category-B),Department of Computer Science, Research & Development Centre,Bharathiar University, Coimbatore-46

²Research Supervisor,Dean Networking, HOD/IT, Sri Sai Ram Engineering College,Chennai -44.

Abstract:

Social media data has been extensively employed in the banking institutions in order to improve the product and service development, customer service, marketing, risk management and business performance. The online social networks such as Facebook, Twitter, etc are considered for analysis. To build the social media strategies, bank needs customer to drive trustworthiness, revenue and success is all about the customer experience. The approach employed Natural Language Processing (NLP) for social media intelligence retrieval. Then based on the concept of sentiment analysis, the customer opinions are categorized for efficient decision making. Thus the NLP & sentiment analysis can be employed for improving banking service for better customer satisfaction.

Keywords — Customers Opinions, Social Media Strategies,Sentiment Analysis,Natural Language Processing.

1.Introduction:

In the modern years a growing amount of individuals are fixed through an internet so the flowing information is always increasing. Public in all over the world post huge volume of opinionated information in various formats to various media services on the WWW. Some of the organization teams have immense interest in what customer's expression and opinion regarding their products. Computerizing of this process is helpful to know the customers thought and customer satisfaction. This is motivated to do the research in sentiment analysis. As numerous customers trust in social media, banks could influence it to congregate customer criticism regarding banking services and products .

In Social media commitment permits for crowds-sourcing of notions that could directs to customer-segmented altering of services and products, which provide unreliable requirements to customers. Grounded on customer behavior in social media, like likes and posts, banks could utilize analytics for producing primes from this information that could be utilized to carry out subsequently finest examination and build up next best actions . Banks could exploit social media to situate a feedback loop among the customer reactions and marketing department and tweak operations as well to turn out to be more cooperative and approachable, serving to maintain the bank on that customer's radar. Generally speaking, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event . The attitude may be a judgment or evaluation, affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication (that is to say, the emotional effect intended by the author or interlocutor).

Analyzing and reviewing the comments retrieved from the social media would be more complex due to varying reasons. One of the major issue found is that contents retrieved from the social media might consists more irrelevant information, noises, tags and so on. Processing this information during opinion mining would lead to inaccurate results. Like all opinions, sentiment is inherently subjective from person to person, and can even be outright irrational. It's critical to mine a large — and relevant — sample of data when attempting to measure sentiment. No particular data point is necessarily relevant. It's the aggregate that matters. An individual's sentiment toward a brand or product may be influenced by one or more indirect causes; someone might have a bad day and tweet a negative remark about something they otherwise had a pretty neutral opinion about. With a large enough sample, outliers are diluted in the aggregate. Also, since sentiment very likely changes over time according to a person's mood, world events, and so forth, it's usually important to look at data from the standpoint of time. In the proposed research, the issues mentioned above is concentrated to implement the effective framework which can analyze the sentiments retrieved from the social media accurately. Thus the bank customer opinions can be analyzed well based on which banking services can be promoted.

2. Literature review

In this section, different data mining techniques related to machine learning approaches that are used to perform sentiments analysis on the user expressed contents are discussed. This is done to find the merits and demerits of existing methods that are used for sentiments analysis in terms of accurate classification results.

In [13] authors compare machine learning approaches Support Vector Machine and Naïve Bayes (NB) with an ANN-based method in the context of document-level sentiment classification. In case of balanced data ANN outperformed SVM significantly in 13 tests, while SVM outperformed significantly in 2 tests. ANN has achieved best classification accuracy in all dataset. The Results indicated that SVM tends to less affected by Noisy terms the ANN when the data imbalances increase.

In [14] employed three Machine learning method Naïve Bayesian, Support Vector Machine and maximum entropy classification and data source taken from internet movies dataset(IDM).Rating were automatically extracted with three categorizations positive, negative and neutral. However author more concentrated on positive and negative categorization. They used the standard Bag of feature framework. In terms of relative performances Naïve Bayesian do the worst result and support vector machine results are best, however differences are not very large.

In 2015, (Øye, et al)[45] he proposed sentiment analysis on Norwegian Twitter messages. The aim of his work was to carry out typical sentiment analysis on three distinct datasets gotten from Twitter social media. The datasets were Norwegian general tweets, one about the prime minister of Norwegian, Erna Solberg, and another set of tweets about the Rosenborg, it is football team of Norwegian. Øye presented the analysis based on the two-step approach as described by (Pang and Lee, 2004) and found a precision up to 80% on the polarity classification, and up to 76% when merging subjectivity detection with polarity classification.

In 2012Pushpak Bhattacharyya, in this paper, we give the perception of NLP from the ambiguity processing and constraints of resource under computation. At all the levels of language is full with ambiguity such as words, phrases, sentences or paragraphs.

4. Proposed Methodology:

Text classification is the procedure of grouping archives into predefined classifications in view of their substance. It is the mechanized task of natural language texts to predefined classes. Content arrangement is the essential prerequisite of content recovery frameworks, which recover writings because of a client inquiry, and content comprehension frameworks, which change content somehow, for example, delivering synopses, noting addresses or separating information. Existing administered learning algorithms to consequently order content need adequate records to learn precisely. This paper shows another algorithm for content order utilizing information mining that requires less reports for preparing. Rather than utilizing words, word connection i.e. affiliation rules from these words is utilized to get highlight set from pre-ordered content records. A framework in view of the proposed algorithm has been actualized and tried. The trial results demonstrate that the proposed framework fills in as an effective content classifier.

➤ **Tokenization**

- ✓ Convert document to word counts = “bag of words”
- ✓ word token = “any non-empty sequence of characters”
- ✓ for HTML (etc) need to remove formatting

➤ **Canonical forms, Stopwords, Stemming**

- ✓ Remove capitalization
- ✓ Stopwords:
 - ✓ Remove very frequent words (a, the, and...) – can use standard list
 - ✓ Can also remove very rare words, e.g., words that only occur in k or fewer documents, e.g., k = 5

➤ **Data representation**

- ✓ “Bag of words” most commonly used: either counts or binary
- ✓ Can also use “phrases” for commonly occurring combinations of words.
- ✓ 6 months social Media data from sep 2015 – Feb 2016 with 700 observations
 - ✓ 700 Customer ID’s
 - ✓ 12 Cities
 - ✓ Customer E_Mail
 - ✓ Comments , Remarks

4. Models and Methodologies:

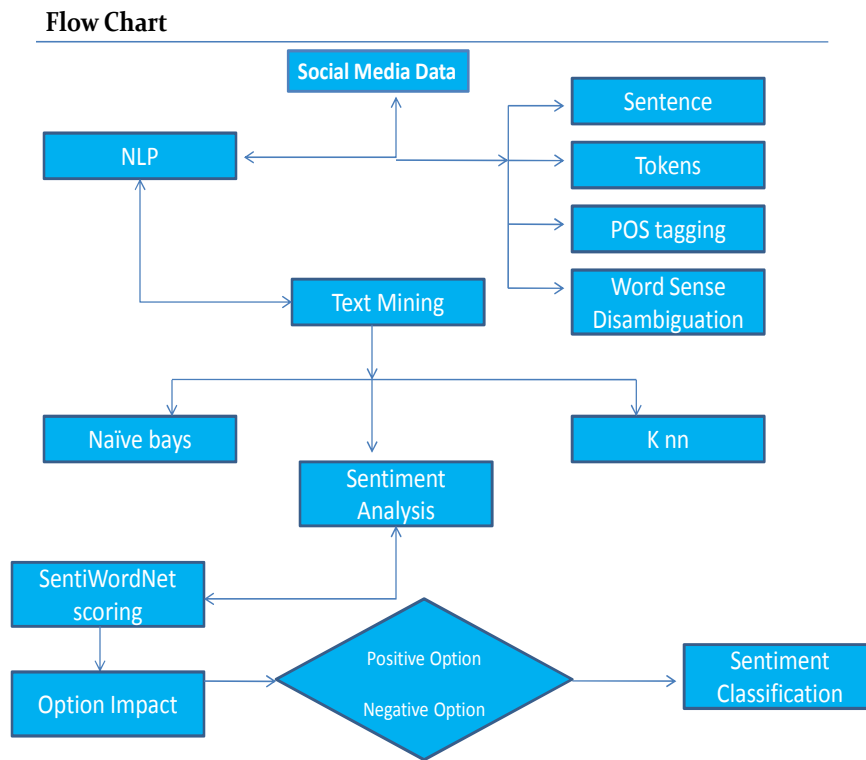


Fig 1. Proposed Metodology of Sentiment Classsification

Our proposed system is primarily combined of the following steps: data collection and cleaning, giving sentiment index value, generating the information system using roughset and analyzing the sentiment. Figure. 2 illustrates the diagram of our proposed method. There are three modules are defined to integrating opinion mining, such as data collection, rough set feature extraction and analyzing features attribute based result analyzed. First Module describing the getting content and data cleaning based on text sentiment analysis. The second module is lexicon-based approach, it helps to originates and analyzes positivity, negativity and uncertainty text posted in twitter social media about bank products. In Third module is extracting the feature from posting and compute the sentiment index value for each. Our proposed system will grant an integer value for each feature, with the sign showing its emotional polarity and the absolute value its emotional intensity.

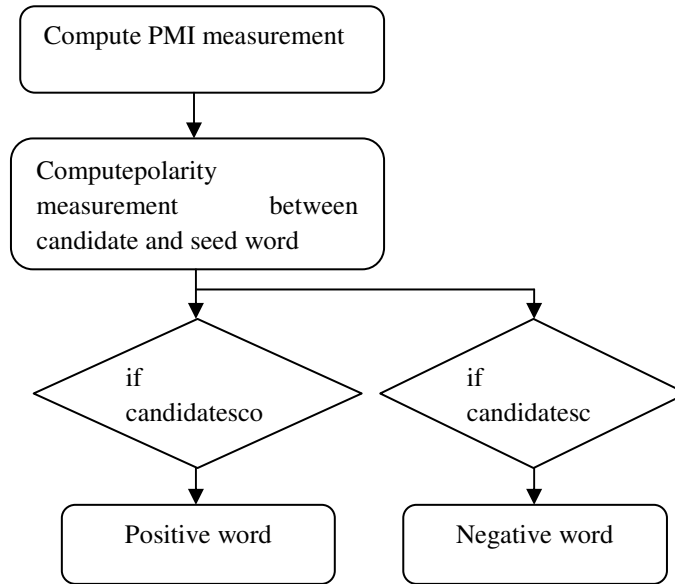


Fig:2 Polarity computation of Sentiment words

Algorithm 1 Text Sentiment Classification

```

    procedure ClassifySentiment(Data)
      Positives = 0
      Negatives = 0
      for every word T in data do
        if Lexicon(T) is positive then
          Positives = Positives + 1
        else if Lexicon(T) is negative then
          Negatives = Negatives + 1
        end if
      end for
      if Positives > Negatives then
        return "positive about products"
      else if Positives < Negatives then
        return "negative about products"
      else
        return "Uncertainty (Average Products)"
      end if
    end procedure
  
```

Positive	Good, Excellent, easy, Secure Transaction
Negative	Connection problem, less Response, Poor
Uncertainty	Average, Somewhat ok, not good not bad

Table 1: Example of Sentiment Words

The Banking dataset is prepared to examine banking assessments for the business intelligence. The numerous transactions are executed every day in a banking environment grounded on deposits and extracted to examine the course of the patterns for the customer contribution. A decision system is used for minimizing the attribute which is used to retrieve the text, those is called condition attributes. That is the set of feature attribute used to get exact results that particular attribute called as decision attribute.

A very simple information system is shown in Table. 2. There are six cases or objects, and two condition attributes

- Table 2: an example information system.

Term	Bank mission Sentiment	Feature lexicon
policy	37	+1
Financial	25	+1
Bank	36	0
That	8	0
Loan	85	+1
Wont	23	-1
Wanted	8	-1

- Any one posted data file will contain only a subset of all individual terms, and the rows corresponding to used terms and giving number of time it used by customer's. The third column feature lexicons for sentiment analysis such as positive represent as +1, negative represented as -1 and uncertainty represents as 0.
- This posteriori knowledge is stated by one separated variable called decision variable; the procedure is called as supervised learning. Information systems of this category are called decision systems.

5. Experimental Results

In this system, to the executed sentiment analysis technique have need of a great amount of text data. The purpose of this execution is to make a paradigm that exactly classifies any type of bank text as negative (-1), uncertainty (0) and positive (1). Appropriate to generate an effective model for this problem, all information is review of the bank product domain. The all reviews written by bank customers and all posted text scripted in English language.

The data set fetched from Facebook and Twitter. Facebook pages occasionally consists a review column as empty where individuals can drop reviews as blank. These posting are treated as

uncertainty sentiment. All Twitter posted text can be manually edited because of there is no rating of posted text.

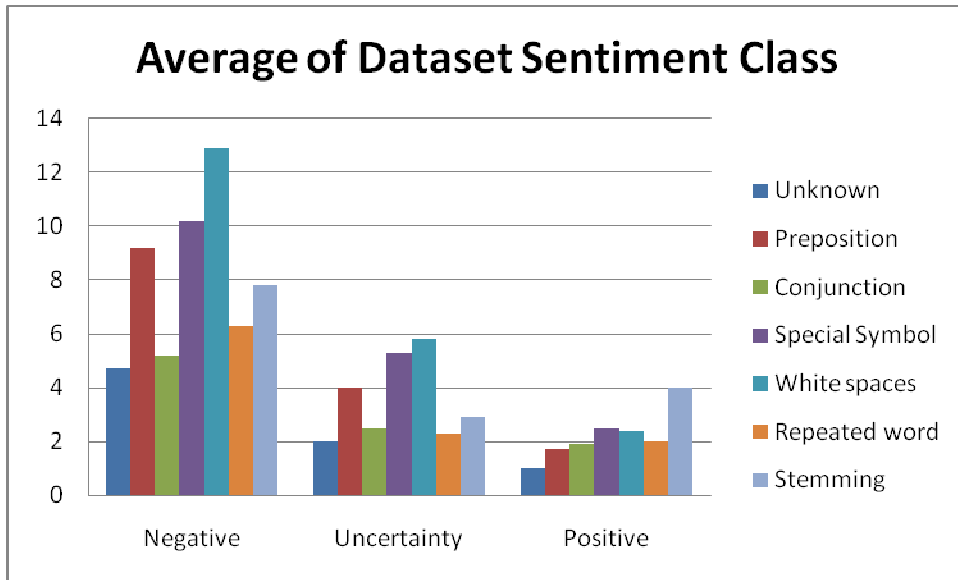


Figure 3. Dataset classification

Sentiment Analysis is the strategy for affecting whether a part of composing is positive, negative or neutral. It's otherwise called opinion mining, determining the judgment or disposition of a speaker. Feeling examination methods can develop to build up estimation on a scope of levels. Sentiment analysis is worried with the programmed annihilation of feeling related data from text. Even however the larger part of slant examination addresses, business undertakings, for example, destruction sentiments from development surveys, there is developing enthusiasm for the passionate estimation of the social web.

Term	Bank mission Sentiment	Feature lexicon
policy	37	+1
Financial	25	+1
Bank	36	0
That	8	0
Loan	85	+1
Wont	23	-1
Wanted	8	-1

Table 3. An Example Information System.

Any one posted data file will contain only a subset of all individual terms, and the rows corresponding to used terms and giving number of time it used by customer's. The third column feature lexicons for sentiment analysis such as positive represent as +1, negative represented as -1 and uncertainty represents as 0. This posteriori knowledge is stated by one separated variable called decision variable; the procedure is called as supervised learning.

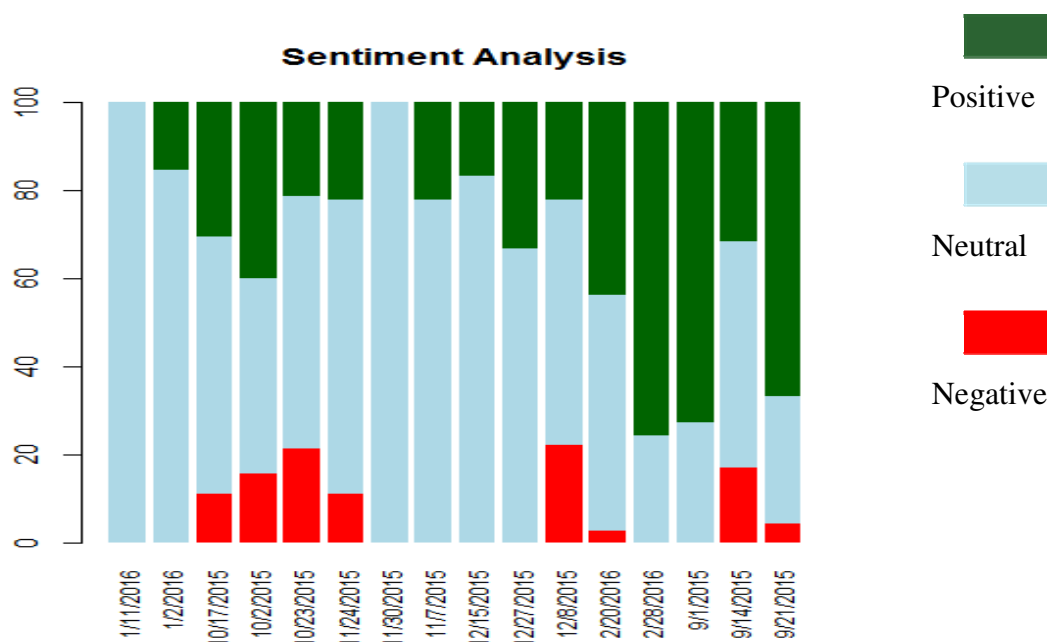


Fig . Sentiment Analysis On a range of levels

6. Conclusion:

To build the social media strategies, bank needs customer to drive trustworthiness, revenue and success is all about the customer experience. The proposed system introduced an efficient sentiment analysis to build the social media strategies, bank needs customer to drive trustworthiness, revenue and success is all about the customer experience. The proposed system introduced an efficient sentiment analysis. Use of the Social Media Intelligence by using Natural Language Processing (NLP) are proving truly a game changer for the banking institutions. Undoubtedly, the advancement in the technological field and wide use of the NLP will boom the social media intelligence in banking institution. To tackle the challenge that is how to create a new social science of online media, banking institutions are more frequently moving towards NLP tools which are integrated with the social media intelligence and with the science as well.

In future, we use various classification methods for sentiment lexicons in bank with rough set feature selection technique, and comparing another technique which is give the more accuracy result.

7. References :

- [1] Mayfield, J., McNamee, P. and Piatko, C.(2000), The JHU/APL HAIRCUT System at TREC-8, In Proceedings of TREC-8, Gaithersburg, Maryland, USA.
- [2] Kamps, J., Monz, C., de Rijke, M. and Sigurbjörnsson, B.(2003) The University of Amsterdam at CLEF-2003, In Results of the CLEF 2003 Cross-Language System Evaluation Campaign, pages 71–78, Trondheim, Norway.

- [3] Chen, A.(2002), Cross-Language Retrieval Experiments at CLEF 2002, In Proceedings of CLEF-2002, Rome, Italy.
- [4] Baldwin, T.(2001), Low-cost, High-performance Translation Retrieval: Dumber is Better, In Proceedings of ACL-01, Toulouse, France.
- [5] Zhang, T. 2002. Covering Number Bounds of Certain Regularized Linear Function Classes. In Journal of Machine Learning Research, 2(Mar):527-550, 2002.
- [6] **Pang, B., Lee, L. & Vaithyanathan, S, 2002.** Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing. Association for Computational Linguistics. 10: 79-86.
- [7] **Mukherjee, A., Liu, B. & Gance, N, 2012.** Spotting fake reviewer groups in consumer reviews. In Proceedings of the 21st international conference on World Wide Web, 191-200.
- [8] **Fu, G. & Wang, X, 2010.** Chinese sentence-level sentiment classification based on fuzzy sets. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, Pp. 312-319.
- [9] **Mohandas, N., Nair, J.P. & Govindaru, V, 2012.** Domain specific sentence level mood extraction from malayalam text. In International Conference on Advances in Computing and Communications (ICACC), Pp. 78-81