SENTIMENT ANALYSIS USING FEATURE SELECTION AND CLASSIFICATION ALGORITHMS – A SURVEY

SHRUTI PANT
Department of computer science and engineering, MPUAT University/College of technology and engineering, Udaipur, India, pant.shruti5@gmail.com

MRS. KALPANA JAIN
Assistant professor of CSE, College of technology and engineering, Udaipur, India,kalpana_jain2@rediffmail.com

ABSTRACT

In today's techno-savvy world people post their day to day emotions on the internet. With the commencement of social media platforms companies have started to use social media as fertile ground for sentiments and opinions. However the challenge is to make this influx of opinion into valuable information. Sentiment analysis is computationally identifying the opinions by categorizing a piece of text in positive and negative classes. Sentiment analysis is a popular area of research in data mining. This paper discusses various techniques, methods and viewpoint of various authors.

INTRODUCTION

With the advance of World Wide Web more and more people are expressing their views, comments, feedbacks reviews online. So we need to find a way to extract and analyse heaps of content into simple meaningful forms which can help in decision making. Today many companies have started using sentiment analysis as the basis of their marketing game plan. Researchers have developed automatic tools that hide the tiny intricate details of Sentiment analysis from the users and help them to find sentiments of users in a much convenient way.

The term sentiment analysis perhaps first coined by Nasukawa and Yi in 2003[1], and opinion mining was first used by Dave, Lawrence and Pennock, in 2003[2]. Initially text mining [3], [4] focused on extracting factsfrom documents. Today, focus is opinion mining or sentiment analysis becauseof the availability of opinionated text in the form of reviews, blog posts, social media comments and more recently, tweets, also called User Generated Content (UGC).
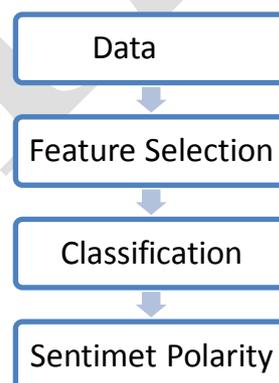
```
┌─────────────────┐
│      Data       │
└─────────────────┘
         ↓
┌─────────────────┐
│ Feature Selection │
└─────────────────┘
         ↓
┌─────────────────┐
│  Classification  │
└─────────────────┘
         ↓
┌─────────────────┐
│ Sentimet Polarity │
└─────────────────┘
```

Fig 1 Process of sentiment analysis

REVIEW ON TYPES OF SENTIMENT ANALYSIS

Sentiment analysis can be categorised into three parts as discussed below

Document level

It is the simplest form of analysis as it considers the whole document as its fundamental unit of information which is then classified for positives and negatives.Peter D. Turney, 2002[6]; Yan Zhao *et al*, 2014[7]; Richa Sharma *et al*, 2014[5] has worked a lot in document based sentiment analysis.

Sentence Level

Analysis uses sentences for classification of sentences. Not more than often document level analysis and sentence level analysis is used interchangeably as a document is a collation of sentences. Sentence level sentiment analysis is the fine- grained analysis of the document and can be categorised in two tasks.

- Subjectivity Classification: A sentence can be either Objective sentence which contains the facts. It has no judgement entity while subjective sentence are driven by emotions or opinions (e.g.), Udaipur's economy depend on tourism. It is a beautiful city.The first sentence is a factual one and does not convey any sentiment towards Udaipur. Hence this does not play any role in deciding on the polarity of the review and must be filtered out [8]. The advantage of sentence level analysis is in the subjectivity/ objectivity classification.

- Sentiment Classification: Document level classification methods can be used as the sentence level classification problem. A number of researches have focused on finding ways to classify the text effectively [9][10][11]. A number of different methods,under supervised machine learning approach, are discussed and compared by [12]. A rule based domain independent and lexical approach for classification of objective and subjective sentenceshas been proposed by [13] where the semantic score is extracted from SentiwordNet lexical resource.
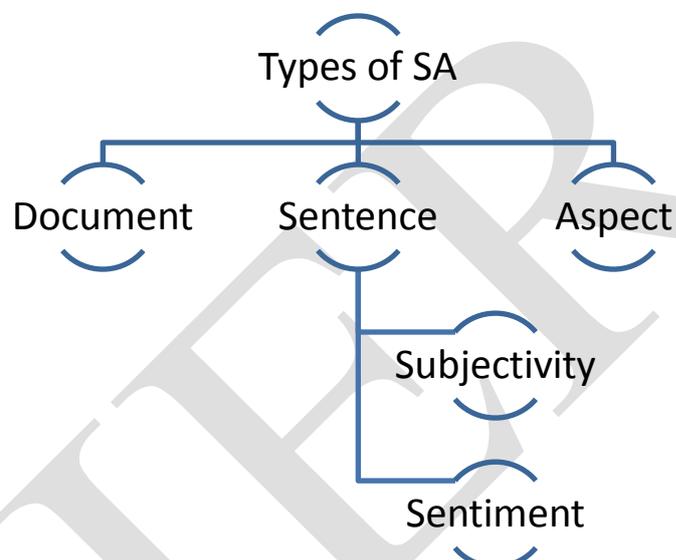

Fig 2 Types of Sentiment Analysis

Aspect level

Aspect level analysis deals with a particular facet or characteristic of the product. For example a mobile phone review can have different aspect such as battery life, camera, screen size and resolution, RAM. Minqing Hu and Bing Liu's [14] work is standard in this field. They used different techniques of data mining and NLPand these techniques have been further outperformed by frequent pattern mining algorithm called H-mine presented in[15]. A new approach is proposed in [16] uses feature oriented words lexicon which review categorization based on polarity of the adjectival words for the frequent features of the product.

REVIEW ON TECHNIQUES OF SENTIMENT ANALYSIS

Sentiment Classification can be done with three techniques machine learning (ML) approach, lexicon based approach and hybrid approach [17]. The text categorisation methods using machine learning can be divided into supervised and unsupervised learning methods. The supervised methods are used when large number of labelled training data is available. The unsupervised methods are used when it is difficult to find labelled training data.

Supervised Methods

Sentiment classification can be composed as a supervised learning problem with four classes, positive, negative, neutral and constructive where user generated contents are mostly used as training and testing data [18], [19], [20], [21]. Many existing supervised learning techniques areused as sentiment classification, such as Naïve Bayes and Support Vector Machines (SVM). One of the very first to perform sentiment analysis on

online movie reviews were Pang et.al. [19]. They tested machine learning techniques, namely, Support Vector Machine, Maximum Entropy and Naïve Bayes classifiers, and trained them on various feature sets including unigrams. Their results showed that an SVM trained on a unigram feature set, outperforms all other approaches. Drawbacks of standard machine learning approaches are that they are both domain and temporally dependent [22]. In most of the cases, SVMs outperformed Naïve Bayes classifiers but whenever the set of training data is small, a Naïve Bayes classifier was more appropriate. One of the most fundamental tasks in sentiment classification is selecting an appropriate set of features. Some of the important features are:

- Terms and their Frequency: These features are individual words (unigram) and their n-grams which are associated with their frequency counts
- Part of speech: Words representing different parts of speech (POS) may be treated differently like adjectives can carry a large amount of information about a document's sentiment.
- Sentiment words and phrase: Sentiment words are words that express positive or negative sentiments. For example, good, awesome, and nice are positive sentiment words, and poor, and risky are negative sentiment words.

Unsupervised Methods

Sentiment words and phrases are indicators for sentiment classification and using unsupervised learning on such words and phrases would be quite natural. For example, the method in [23] uses known opinion words for classification, while [14] defines some phrases that are likely to be opinionated.

Turney[24] displayed an unsupervised learning algorithm for characterizing a review as suggested or not suggested. He figured out whether the words are positive or negative and how correct the assessment is by finding the words' point wise mutual information (PMI) for their co-occurrence with a positive seed word "excellent" and a negative seed word "poor" and called this value the word's semantic orientation.He accomplished 74% accuracy classifying a corpus of item reviews.

Harb et al. [25] performed blog classification with two sets of seed words with positive and negative semantic introductions separately same as in [21] and used Google's web search engine to make association rule. They counted the number of positive versus negative adjective in a document to classify the documents.Taboada et al. [26]presented a lexicon-based method for sentiment classification where they used dictionaries of positive or negative polarized words for classification task.
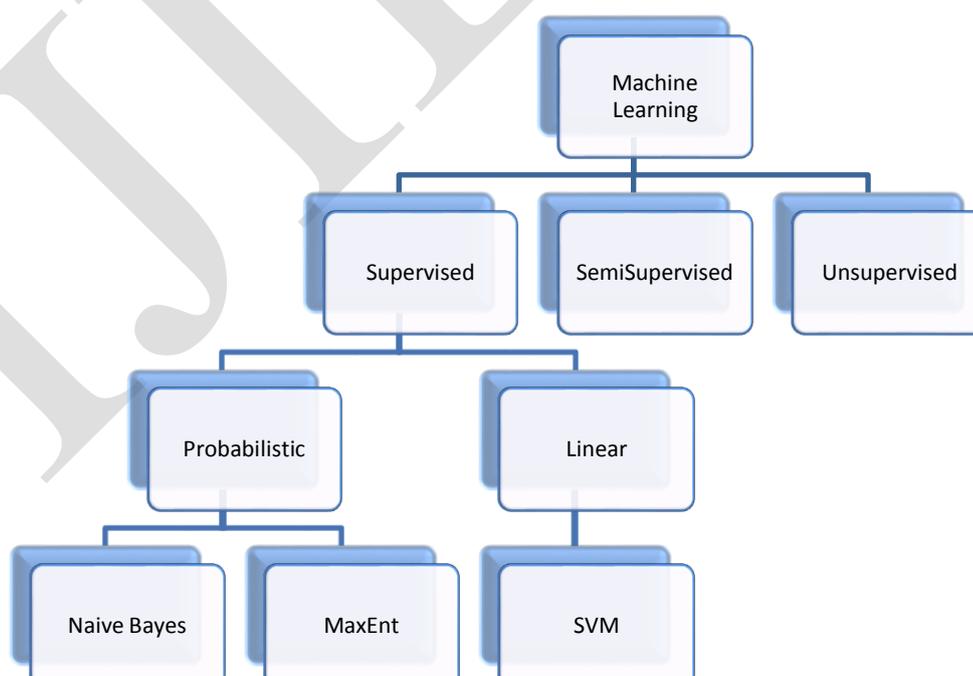

Fig 3 Classification of Machine Learning

CONCLUSION

Sentiment analysis is a hot area of research in the field of data mining and has a wide variety of applications in e-commerce. It helps in classifying and summarizing reviews in real time operations. This paper presents a survey about types of sentiment analysis and techniques of sentiment classification discussed by various authors.

REFERENCE

1) T. Nasukawa, "*Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions,*" pp. 70–77, 2003.

2) K. Dave, I. Way, S. Lawrence, and D. M. Pennock, "*Mining the Peanut Gallery□: Opinion Extraction and Semantic Classification of Product Reviews,*" 2003.

3) D. M. Blei, A. Y. Ng, and M. I. Jordan, "*Latent Dirichlet Allocation," vol. 3, pp. 993–1022, 2003.*

4) T. Hofmann. P. latent, " *semantic indexing," In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in inFmnation retrieval, SIGIR '99, pages 50-57, New York, NY, USA, 1999. ACM*

5) Richa Sharma, Shweta Nigam and Rekha Jain, (2014), *Opinion Mining Of Movie Reviews At Document Level, International Journal on Information Theory (IJIT), Vol.3, No.3*

6) Peter D. Turney, (2002), *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp. 417-424*

7) Yan Zhao, Suyu Dong and Leixiao Li, (2014), *Sentiment Analysis on News Comments Based on Supervised Learning Method, International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.7 pp.333-346*

8) S.ChandraKala and C. Sindhu, (2012), *Opinion Mining And Sentiment Classification: A Survey, ICTACT Journal on Soft Computing, Vol- 03, ISSUE: 01, ISSN: 2229-6956*

9) V. S. Jagtap, KarishmaPawar, (2013), *Analysis of different approaches to Sentence-Level Sentiment Classification, International Journal of Scientific Engineering and Technology, PP : 164-170.*

10) Raisa Varghese, Jayasree M, (2013) *A Survey on Sentiment Analysis and Opinion Mining, International Journal of Research in Engineering and Technology (IJRET), eISSN: 2319-1163 pISSN: 2321-7308*

11) GizemGezici, BerrinYanikoglu, DilekTapucu, and YucelSaygın, (2012), *New Features for Sentiment Analysis: Do Sentences Matter? First International Workshop on Sentiment Discovery from Affective Data (SDAD)*

12) S Padmaja and Prof. S Sameen Fatima, (2013*), Opinion Mining and Sentiment Analysis –An Assessment of Peoples' Belief: A Survey, International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC), Vol.4, No.1*

13) Aurangzeb Khan, Baharum Baharudin, (2011), *Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs, Int. J Comp Sci. Emerging Tech, Vol-2 No 4*

14) Minqing Hu and Bing Liu, (2004), *Mining and Summarizing Customer Review, ACM*

15) Seyed Hamid Ghorashi, Roliana Ibrahim, Shirin Noekhah and Niloufar Salehi Dastjerdi, (2012),*A Frequent Pattern Mining Algorithm for Feature Extraction of Customer Reviews International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, ISSN: 1694-0814*

16) D DChaudhari, R A Deshmukh , A B Bagwan, P K Deshmukh, (2013), *Feature based approach for Review Mining Using Appraisal Words, IEEE, ISBN 978-1-4799-1082-3*

17) W. Medhat, A. Hassan, and H. Korashy, *"Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J., May 2014.*

18) B. Pang and L. Lee, *"Opinion Mining and Sentiment Analysis," Found. Trends® Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.*

19) B. Pang, L. Lee, and S. Vaithyanathan, "*Thumbs up?: sentiment classification using machine learning techniques," presented at the Proceedings of the ACL-02 conferenceon Empirical methods in natural language processing - Volume 10, 2002.*

20) D. Bespalov, B. Bai, A. Shokoufandeh, and Y. Qi, *"Sentiment Classification Based on Supervised Latent n-gram Analysis," pp. 375–382, 2011.*

21) M. Gamon, *"Sentiment classification on customer feedback data: noisy data, large feature vectors , and the role of linguistic analysis."*

22) J. Read, *"Using emoticons to reduce dependency in machine learning techniques for sentiment classification," In Proceedings of the ACL Student Research Workshop, 2005.*

23) M. Taboada, J. Brooke, M. To_loski, K. Voll, and M. Stede, *"Lexicon-based methods for sentiment analysis". Computational Intelligence, 2010.*

24) P. D. Turney, *"Thumbs Up or Thumbs Down Semantic Orientation Applied to Unsupervised Classification of Reviews," no. July 2002, 2001.*

25) A.Harb, M. Planti, M. Roche, A. Harb, M. Planti, M. Roche, N. Cedex, and A. Harb, *"Web Opinion Mining: How to extract opinions from blogs To cite this version Web Opinion Mining How to extract opinions from blogs Categories and Subject Descriptors."*

26) M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "*Lexicon-based methods for sentiment analysis," Comput. Linguist. vol. 37, pp. 267-307, 2011.*