

group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters.

4.3. Case Based Reasoning

Case based reasoning is an emerging Artificial Intelligence supervised technique used to find the solution of a new problem on the basis of past similar problems. CBR is a powerful tool of computer reasoning and solve the problems (cases) in such a way which is closest to real time scenario. It is a recent problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical rules. The past problem's solutions are stored in CBR repository called Knowledge base or Case base. Instead of solving the new problem by "first principal" reasoning, CBR use the knowledge base to reuse the The solution of past similar problem if needed to the In case base repository as a new solution instance in CBR cycle consists of four R's. Nowadays it is the most emerging technique used in opinion mining systems. Statistical methods are combined with knowledge extracting techniques in to enhance case searching, browsing and Reuse it for the problem solving methods semantic analysis of a sentence in natural language that can be easily used and manipulated in a textual data mining process. This sentence analysis uses and depends on several types of knowledge that are: a lexicon, a case base and hierarchy of index. In this methodology a case based reasoning model is adopted that is based on the classification rules and course of similarity for the assurance of the compliance.

5. Opinion Mining and Summarization Process

Opinion Mining also called sentiment analysis is a process of finding user's opinion towards a topic. Opinion mining concludes whether user's view is positive, negative, or neutral about product, topic, event etc. Opinion mining involves analyzing user's opinion, attitude, and emotion towards particular topic. This consists of first categories text into subjective and objective information, and then finding polarity in subjective text. Opinion mining can be performed word, sentence or document level. Opinion mining and summarization process involve three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization.

Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction [4] and opinion summarization are key issues. Many researchers worked on summarization product reviews [2]. The opinion summarization process mainly involve following two approaches. One is Feature based summarization another one is Term Frequency based summarization.

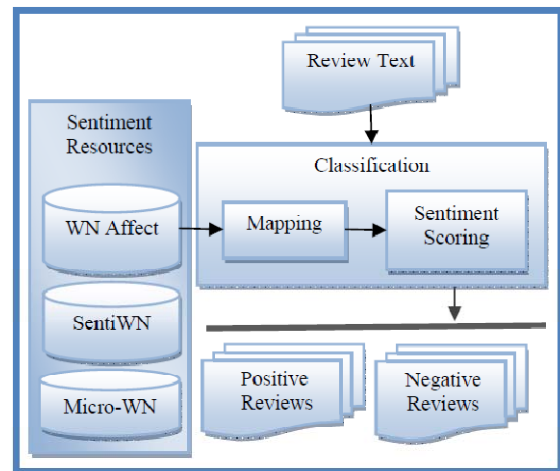


Figure 3: Opinion Mining Summarization

5.1 Opinion Retrieval

Opinion retrieval is a process of collecting reviews text from review websites. Different review websites contain reviews for products, movies, hotels, news etc. Information retrieval techniques such as web crawler can be applied to collect review text data from many sources and store them in database. This step involves retrieval of reviews, micro-blogs, comments etc of user. We should only consider the data which contain subjective data but not the objective data. Reviews are retrieved by query based information retrieval techniques.

5.2. Opinion Classification

Primary step in sentiment analysis is classification of review text. Given a review document $D = \{d_1..d_l\}$ and predefined categories set $C = \{positive, negative\}$, sentiment classification is to classify each d_i in D , with expressed in C . There are many approaches for sentiment classification in opinion text. Machine learning and lexicon based approach is more popular.

5.2.1. Machine learning approach for opinion classification:

The machine learning approach uses supervised learning method for classification of review text. The first step is to train a classifier using sample of reviews with its class (positive/negative). Then the built model of trained classifier is used to predict category of new text reviews. Popular machine learning classifiers for text categorization are Support Vector Machines (SVM) and Naive Bayes(NB).

5.2.2. Lexicon based approach for opinion classification:

The lexicon Approach predicts sentiment of review text using databases which contain word polarity values e.g. SentiWordNet [10]. Review text is classified by calculating and averaging polarity score of individual words in sentences. Many factors such as word position, word relationships, negation handling should be considered while sentiment classification using lexicon based approach.

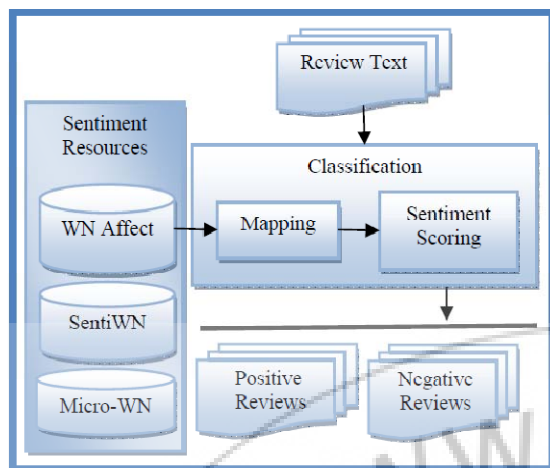


Figure 4: Lexicon based Opinion Classification

5.2.3 Opinion Summarization

Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction and opinion summarization are key issues. Many researchers worked on summarization product reviews. The opinion summarization process mainly involve following two approaches.

1) Feature based summarization:

This type summarization involve finding of frequent terms (features) that are appearing in many reviews. The summary is presented by selecting sentences that contain particular feature information. Features present in review text can be identified using Latent Semantic Analysis (LSA) method. For a short summary of product reviews, product features and opinion words associated.

2) Term Frequency based summarization:

Term frequency is count of term occurrences in a document. If a term has higher frequency it means that term is more import for summary presentation. In many product reviews certain product features appear frequently and associated with user opinions about it. In this method sentences are scored by term frequency. The summary is presented by selecting sentences that are relevant and which contain highest frequency terms. Opinion Summarization process is shown in Fig.5 It shows review text is preprocessed which involve sentence segmentation and tokenization of sentence in terms. After calculating term frequency of each term, each sentence score and relevance is calculated. As per the compression rate highest scoring and relevant sentences are presented in summary.

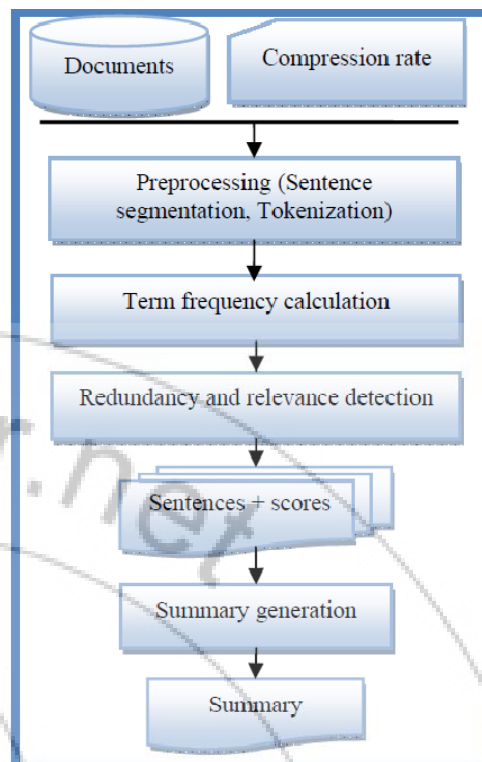


Figure 5: Opinion summarization

6. Research Areas in Opinion Mining

Current research is focusing on

- 1) Customer Reviews for Individual Product Feature Based Ranking.
- 2) Overall positive and negative polarity at paragraph level.
- 3) Ranking of best paragraph or sentence based on best feature and their polarity involved.
- 4) Continuous Improvement of the algorithms for opinion detection.
- 5) Decrease the human effort needed to analyze contents.
- 6) Semantic analysis through lexicon/corpus of words with known sentiment for sentiment classification.
- 7) Identification of highly rated experts.

6.1 Future Research: long term and short term issues

Short-term:

- Enhanced discoverability of content through Linked Data
- Visual representation
- Audiovisual opinion mining
- Real-time opinion mining
- Machine learning algorithms
- SNA applied to opinion and expertise
- Bipolar assessment of opinions
- Multilingual reference corpora
- Comment and opinion recommendation algorithm
- Cross-platform opinion mining
- Collaborative sharing of annotating/labeling resources

Long-term:

- Autonomous machine learning and artificial intelligence
- Usable, peer-to-peer opinion mining tools for citizens
- Non-bipolar assessment of opinion
- Automatic irony detection

Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications. There are likely to be many other applications that are not discussed. It is found that sentiment classifiers are severely dependent on domains or topics. From the above work it is evident that neither classification model consistently outperforms the other, different types of features have distinct distributions. It is also found that different types of features and classification algorithms are combined in an efficient way in order to overcome their individual drawbacks and benefit from each other's merits, and finally enhance the sentiment classification.

7. Conclusion

Due to web and social network, large amount of data are generated on Internet every day. This web data can be mined and useful knowledge information can be fetched through opinion mining process. This paper discussed different opinion classification and summarization approaches, and their outcomes. This study shows that machine learning approach works well for sentiment analysis of data in particular domain such as movie, product, hotel etc., while lexicon based approach is suitable for short text in micro-blogs, tweets, and comments data on web.

References

- [1] N. Au, R. Law, and D. Buhalis. The impact of culture on ecomplaints: Evidence from the chinese consumers in hospitality organization. In U. Gretzel, R. Law, and M. Fuchs, editors, *Information and Communication Technologies in Tourism 2010*, pages 285–296. Springer Verlag Wien, 2010.
- [2] C. Chen, F. Ibekwe-SanJuan, E. SanJuan, and C. Weaver. Visual analysis of conflicting opinions. In *IEEE Symposium on Visual Analytics Science and Technology*, pages 35–42, 2006.
- [3] Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, “Sentiment classification of Internet restaurant reviews written in Cantonese”, *Expert System with applications*, 2011.
- [4] Padmaja.S and SameenFatima, “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, February 2013*
- [5] KaiquanXu, Stephen Shaoyi Liao, Jiexun Li, Yuxia Song, “Mining comparative opinions from customer reviews for Competitive Intelligence”, *Decision Support Systems* 50, 743–754, (2011).
- [6] G. Jaganadh 2012. Opinion mining and Sentiment analysis CSI communication.
- [7] Yingcai Wu, Furu Wei, Shixia Liu, Norman Au, Weiwei Cui, Hong Zhou, and Huamin Qu, *Member, IEEE* “Opinion Seer: Interactive Visualization of Hotel Customer Feedback,” *IEEE transaction on visualization and computer graphics*, vol.16, no.6, november/december 2010.
- [8] Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.
- [9] C.L. Liu, W.H. Hsaio, C.H. Lee, G.C. Lu and E. Jou, “Movie Rating and Review Summarization in Mobile Environment”, *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, Vol. 42, No. 3, pp. 397-407, 2012.
- [10] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques,” in *Proc. ACL-02 Conf. Empirical Methods Natural Lang. Process.*, 2002, pp. 79–86.
- [11] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, “Movie Rating and Review Summarization in Mobile Environment”, *IEEE VOL. 42, NO. 3, MAY 2012*.
- [12] Pang, B., L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)*, 2002.
- [13] Ku, L.-W., Liang, Y.-T., & Chen, H, “Opinion extraction, summarization and tracking in news and blog corpora”. In *AAAI-CAAW’06*.
- [14] Melville, Wojciech Gryc, “Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification”, *KDD09*, June 28–July 1, 2009, Paris, France. Copyright 2009 ACM 978-1-60558-495-9/09/06.
- [15] Titov, I., McDonald, R.: A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In: *Proceedings of ACL-2008: HLT*, pp. 308–316 (2008).
- [16] Nilesh M. Shelke, Shriniwas Deshpande, PhD. and Vilas Thakre, PhD., *Survey of Techniques for Opinion Mining, International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012*.
- [17] Xiaohui Yu, Member, IEEE, Yang Liu, Member, IEEE, Jimmy Xiangji Huang, Member, IEEE, and Aijun An, Member, IEEE, “Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 24, NO. 4, APRIL 2012.
- [18] Christopher C. Yang and Tobun Dorbin Ng, *Member, IEEE*, “Analyzing and Visualizing Web Opinion Development and Social Interactions With Density-Based Clustering,” *IEEE Transactions on Systems, man, and cybernetics—part a: systems and humans*, vol. 41, no. 6, november 2011.
- [19] Ainur Yessenalina, Yisong Yue and Claire Cardie, *Multi-level Structured Models for Document-level Sentiment Classification*, *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1046–1056, MIT, Massachusetts, USA, 9-11 October 2010. Association for Computational Linguistics.
- [20] Hanhoon Kang, SeongJoon Yoo, Dongil Han, “Sentilexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews”. *Expert Systems with Applications* 39 (2012) 6000–6010.
- [21] Qiang Ye, Ziqiong Zhang, Rob Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches”, *Expert Systems with Applications* 36 (2009) 6527–6535.

Author Profile



S. Veeramani received the B.E. degree in Computer science and Engineering from Nandha engineering college in 2012. He is currently doing his M.E Computer science and Engineering in Nandha engineering college, Erode, India