A Survey Paper: Areas, Techniques and Challenges of Opinion Mining

Ayesha Rashid[1], Naveed Anwer[2], Dr. Muddaser Iqbal[3], Dr. Muhammad Sher[4]

[1]Department of Computer Science and Software Engineering
University of Gujrat, Pakistan

[2]Department of Computer Science and Software Engineering
University of Gujrat, Pakistan

[3]Department of Computer Science and Software Engineering
University of Gujrat, Pakistan

[4]Department of Computer Science and Software Engineering
International Islamic University Islamabad, Pakistan

Abstract

Opinion Mining is a promising discipline, defined as an intersection of information retrieval and computational linguistic techniques to deal with the opinions expressed in a document. The field aims at solving the problems related to opinions about products, Politian in newsgroup posts, review sites, comments on Facebook posts and twitter etc. This paper is about to covers the techniques, applications, long and short future research areas, Research gaps and future challenges in opinion mining. Further, an attempt has been made to discuss in detail the use of supervised, unsupervised, machine learning and case based reasoning techniques in opinion mining to perform computational treatment of sentiments.

Keywords: *Data mining, web mining, opinion mining, sentiment classification, supervised learning, unsupervised learning, CBR, machine learning, knowledge discovery in database, classification, clustering, sentiment analysis*

1. Introduction:

This era is of automated systems [1] and digital information [2]; every field of life is evolving rapidly and generating data. As a result huge volumes of data produce in field of science, engineering, medical, marketing, finance, demographic etc [3]. Automated systems are needed to automate analysis, summarization, and classification of data. It also helps at enterprise level to take related decisions. Multiple research fields like statistics, machine learning, artificial intelligence and visualization are involved to develop such automated systems [3-7].

A number of efficient ways are available [4] to store the huge volumes of data, computational techniques and models [8] are required to extract the hidden patterns and knowledge. These techniques and tools are used to transform the data into useful information, to make market analysis, fraud detection and find the customer intentions etc. Such computational tools and techniques are the subject of *Knowledge Discovery in Database and Data Mining* [4-5, 9, 10].Text mining is an interdisciplinary method used in different fields like machine learning, information retrieval, statistics, computational linguistic and data mining [9]to form mining algorithms[11]. Some researchers defined text mining as tool to discover the new knowledge from huge volume of natural language text using computationl algorithms. Web mining is a sub discipline of text mining used to mine the semi structured web data in form of web content mining, web usage mining and wed structre mining.

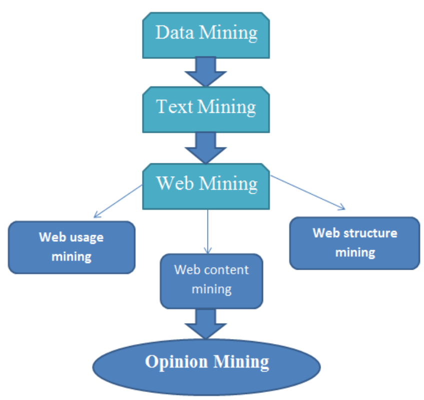

Fig. 1.Data Mining Hierarchy

This paper is organized as follows: section 2 covers opinion mining. Section 3 is about the levels of sentiment classification. Section 4 describes current and future research areas, research gaps and key challenges of sentiment classification. Section 5 defines the techniques of opinion mining i-e supervised machine learning, Unsupervised Learning, and Case based reasoning.

2. Opinion Mining (O.M)

Opinion Mining (OM), a promising discipline, defined as combination of information retrieval and computational linguistic techniques deals with the opinions expressed in a document [12]. The field aims at solving the problems related to opinions about products, politics in newsgroup posts, review sites, etc [13]. There are different techniques for summarizing customer reviews like Data Mining, Information Retrieval, Text Classification and Text Summarization [3, 12, 13]. Before World Wide Web users asked the opinions of his family and friends to purchase the product. In the same way when organizations needed to take the decision about their products they had to conduct the surveys to the focused groups or they had to hire the external consultants [5, 13]. Web 2.0 [14], facilitate the customers to take decision to purchase the product by reviewing the posted comments. Customers can post reviews on web communities, blogs, discussion forums, twitters, product's web site these comments are called user generated contents [12]. Web2.0 is playing a vital role in data extracting source in opinion mining. It facilitates users to know about the product from other customer's reviews who have already used it instead of asking friends and families. Companies, instead of conducting surveys and hiring the external consultants to know about the consumers opinions, extract opinionated text from product web site [13,15]. An automated opinion summarization model is needed to perform these tasks. **Opinion Mining or Sentiment Analysis** is the field to extract the opinionated text datasets and summarize in understandable form for end user [15]. Opinion mining is to extract the positive, negative or neutral opinion summery from unstructured data. It involves computational management of opinion and subjectivity in text. It is the sub-discipline of web content mining, involves

Natural Language Processing and opinion extraction task to find out the polarity of any product consumers feedback [5]. Figure 2 describes the model of O.M.

Fig. 2. Opinion mining Model [16]

Opinion: is the sentiment, views or judgment about any object based on knowledge or experience.
Opinion Holder: is the person, organization that expresses its views or sentiments about any object.
Object: an entity (person, topic, product or organization) about which the opinion expressed.

3. Sentiment Classification

3.1 Document level

Document level sentiment classification executed on the overall sentiments expressed by authors [3]. Documents classified according to the sentiments instead of topic. It is to summarize the whole document as positive or negative polarity about any object (mobile, car, movie, and politician).

In [18] authors proposed a new approach "classification of opinion documents by a vote system" based on combining text representations using key-words related to bigrams. Sentiment Classification Using Phrase Patterns in [19] used Special tags opinion words. System constructed some phrase patterns and compute sentiment orientation using unsupervised learning algorithm. Proposed system achieved 86% accuracy.

[20] Investigated perspective from which a document was written. They build Naïve Bayes based model and test on Israeli-Palestinian conflict. Their corpus consists of articles published on the bitter lemons website. They used NB-B (full Bayesian inference) and NB-M (Maximum a posteriori).

For automatic analysis of the audio reviews about telephone [21] find the sentiment orientation from fragments of the messages to facilitate the classification task. A new approach **AMOD** in [22] automatically extracts positive or negative adjectives in the context of relevant domain.

3.2 Sentence level

Sentence level sentiment classification models extract the sentences contains opinionated terms, opinion holder and opinionated object [15,16]. It is one level deep to document level and just concerns to the opinionated words but not the features [12]. Number of positive and negative words counted from sentences if positive words are maximum then opinion about object is positive and if the negative words are more than opinion is negative otherwise neutral [1].

To mine the customer reviews on a product [2] proposed unsupervised algorithm. They find frequent features using Apriori algorithm. Chinese WordNet set classify opinion words in clauses (pos, neg or neutral) to summarize the comments.

Sentence level opinion mining Using learnt patterns Rilloff and Wiebe [23], Subjectivity and polarity (orientation) Yu and Hatzivassiloglou [24], and Finding strength of opinions at the clause level Wilson et al.[25], all these are a notable work in this regard.

To find the strength of opinions [26] used a new idea of syntactic clues. They use a wide range of features to find the strength of opinions. The system is about to provide tools and support for information analysts in government, commercial, and political domains, who want to be able to automatically track attitudes and feelings in the news and on-line forums.

[27] Proposed an IE system to improve subjectivity classification by using AutoSlog-TS extraction pattern learning algorithm. Opinion Analysis based on Lexical

Clues and their Expansion to improve combination of rule-based algorithms and machine learning techniques.

[28] proposed semi-supervised learning method based on highly precise seed rules. Subjectivity discovered at the sentence level. Polarity of the sentence defined as Positive, Negative or Neutral as well as opinion holders identified. The experimental results demonstrate that system achieved 45% Accuracy to extract opinionated sentences and 35% Accuracy to identify opinion holders.

[29]Presented a system based on conditional probabilities to identify an opinion holder based on an anaphor resolution technique to improve the performance of syntactic features. [31] Proposed a learning method to distinguish between subjective and objective sentences. They used only un-annotated text as their training data. It used Naive Bayes algorithm to classify the sentence.

3.3 Feature based level

In customer reviews document, reviewer express positive, negative or both sentiments about the object and attributes. Document level and sentence level classification does not tell the likes and dislikes of consumer about particular attributes of object [33]. When consumer comment on object (product, person, and topic, organization) he comment on the features of object[13].For example, if users commented on a Mobile Phone they basically comment on Camera result, LCD size, speaker, weight etc. On camera output 125 comments express the positive opinions and 25 comments may be negative. If a new customer is interested in camera quality of mobile he can take decision easily to purchase the product or not [34]. To explore the detailed opinion on product or any topic, a detailed opinion mining study is required that is called feature based opinion mining [35].

Statistical Opinion Analyzer (SOA) [13] extract the polarity of online customer reviews using Bayesian probability and frequency distribution. The proposed system helps the new customer to purchase the product and manufacturer to enhance the product's functionality. Reviews crawled, preprocess, tagged (GO tagger) and insert in SOA to find the positive and negative opinion probability and frequency distribution as well. The proposed system originated the very promising results.

In [37] a web based system SUMView crawled reviews from Amazone.com, decompose into sentences and tagged to find the nouns and noun phrases. Product features extracted using Hu and Liu (2004) method and top five extracted features were suggested to the users on the basis of frequency. The proposed Feature-Based Weighted Non-Negative Matrix Factorization algorithm [38] grouped the feature related highest probability sentences into clusters on the basis of Term-Sentence matrix and user selected feature.

[39] Presented a system to help the tourist to choice the hotel by visualizing on Google map. Burst technique used to find the change in opinion behavior and visualization used to represent "good" and "bad" hotels in graphical form. Linear regression with rule based approach used in [40] to rank the product features according to their importance. Proposed technique divide the overall product's rating into product features rating and generate the customer's survey automatically. Product features and their along sentiment words extracted using Double Propagation Algorithm. [41]Introduced an unsupervised information extraction system OPINE for extracting the frequent features and polarity from customer reviews. OPINE works on PMI feature assessment and relaxation labeling to extract high-precision features and find the semantic orientation of potential opinion words respectively. [42] And [43] both had presented systems to extract product features but OPINE achieve 22% high precision than other systems.

Opinion Observer system [44] analyze and compare customer's opinion on web which is very helpful for manufacturer and new customers. Opinion observer extracted all the explicit and implicit features of product, create opinion segments of features. They also proposed a supervised rule mining technique based on language pattern to automatically extract the product features from Pros and Cons format of customer reviews.

[48] Presented Swarm Based Features Selection for Text Summarization. It finds the weights of each feature based on selection score. Swarm optimization selects the features and their weights to assign score to sentences and selected top scored sentences as summary.

Table 1 describes shortcomings of sentiment classification at documents, sentence and feature level. To overcome these drawbacks could be the future research directions in opinion mining.

Table 1: drawbacks at different sentiment level

Document level classification	It does not give details of what people likes or dislikes because writer comments only the specific aspects of product [5].
	It is not applicable on forums and blogs as they contains only few opinionated sentences on features of object.
	Document level opinion mining defines the polarity of document. But a positive phrase dose not indicates that the user likes everything and similarly a negative phrase does not indicate that the opinion holder

	dislikes everything. Just imagine for a moment, if a document containing review in which holder has likes the story of movie but dislikes its sound and print. The overall sentiment of the document is negative but the holder still likes the movie. So such kind of reviews shows the wrong classification. Similarly if some user dislikes the movie but likes everything else, again the review will be classified as positive due to the average orientation of the positive phrases [50,51].
Sentence level classification	A user can express different views in a single sentence. If a user expresses his likeness of picture quality and dislike the sound of the movie, the review will be ranked as neutral at sentence level. As such kind of the sentences the average orientation of positive and negative phrases will be equal and one cannot find out what user wants to convey [1].
	Similarly if user expresses his opinion about the likeness of the movie in just one sentence and the rest of the sentences are expressing the user dislike some feature then the classification of the document could be wrong and will create negative impact [50].
NLP	A drawback of the NLP approach is that it could really cut very badly if they are used grammatically incorrect text. Currently a large part of the web based sentiment data fall into this category, methods to detect and correct bad English, if any, would be necessary before using them on a larger scale [28].

4. Research Areas in Opinion Mining:

Current research is focusing on

1. Customer Reviews for Individual Product Feature Based Ranking [48]
2. Overall positive and negative polarity at paragraph level [2].
3. Ranking of best paragraph or sentence based on best feature and their polarity involved [50].
4. Continuous Improvement of the algorithms for opinion detection [50, 51].
5. Decrease the human effort needed to analyze contents [51].
6. Semantic analysis through lexicon/corpus of words with known sentiment for sentiment classification [28].
7. Identification of highly rated experts [28].

Table 2 describes the key challenges and gaps found at the different levels and applications of opinion mining. The gaps could be the research directions for researchers.

5. Key challenges and Research Questions of Opinion Mining

1. Product reviews, comments and feedback could be in different languages (English, Urdu, Arabic etc), therefore to tackle each language according to its orientation is a challenging task.
2. As noun words are considered as feature words but Verbs and adjectives can also be used as feature words which are difficult to identify [2].
3. If a customer-A comments on mobile phone, “the voice quality is excellent” and customer-B comments, “sound quality of phone is very good”. Both are talking about same feature but with different wording. To group the synonym words is also a challenging task [39].
4. Orientation of opinion words could be different according to situation. For example “camera size of mobile phone is small”. Here adjective small used in positive sense but if customer parallel said that “the battery time is also small”. Here small represent negative orientation to battery of phone. To identify the polarity of same adjective words in different situation is also a challenging task [40].
5. As the customer comment in free format, he can use abbreviation, short words, and roman language in reviews. For example *cam for camera, pic for picture, f9 for fine, gud for good etc.* To deal with such type of language need a lot of work to mine opinion [17].
6. Different people have different writing styles, same sentence may contain positive as well as negative opinion, so it is difficult to parse sentence as positive or negative in case of sentence level opinion mining [17].
7. In Bing Liu approach opinion always classified only in two categories positive and negative but Neutral opinion also expressed sometimes. Liu considers only adjective as opinion words but opinion can also expressed as adverb, adjectives and verb. For example “like” is a verb but also an opinion word. His approach finds the implicit features because it extracts the sentences contain at least one feature word. So the features commented by customer indirectly are ignored [16, 17].
8. Lexicon based methods use for opinion mining has not an effective method to deal with context dependent words. For example the word “small” can express the either positive or negative opinion on the product features. For a mobile phone if customer comments that “size of mobile phone is small” this sentence does not show

either size is positively opinioned or negatively [54].

9. To finding of spam and fake reviews, mainly through the identification of duplicates.
10. The comparison of qualitative with summary reviews and the detection of outliers, and the reputation of the reviewer [55].
11. The combination of opinion with behavior to validate data and provide further analysis into the data ahead of opinion expressed [56].
12. The continuous need for better usability and user-friendliness of the mining systems [57].

6. Techniques used in Opinion Mining:

Database contains the important hidden information used for decision making. Different databases like relational, object oriented, transactional and spatial databases consist on the complex dataset [5]. The rapid growth in databases has created the need to develop such technologies to extract the nuggets of knowledge and information intelligently [4,5]. Data mining techniques are most suitable for this purpose, these techniques directly refers Artificial Intelligence. Major data mining techniques used to extract the knowledge and information are: generalization, classification, clustering, association rule mining, data visualization, neural networks, fuzzy logic, Bayesian networks, genetic algorithm, decision tree, multi agent systems, CRISP-DM model, churn prediction, Case Based Reasoning and many more [57].

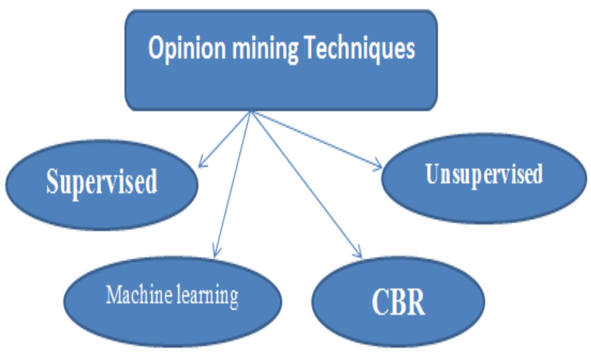

Fig. 5: Techniques of Opinion Mining

6.1 Supervised Machine Learning:

Classification is most frequently used and popular data mining technique. Classification used to predict the possible outcome from given data set on the basis of defined set of attributes and a given predictive attributes. The given dataset is called training dataset consist on independent variables (dataset related properties) and a dependent attribute (predicted attribute) [58]. A training dataset created model test on test corpora contains the same attributes but no predicted attribute. Accuracy of model checked that how accurate it is to make prediction [59]. Classification is a supervised learning used to find the relationship among attributes. Classification model generate the rules in the form of IF-THEN as result from training data. For example a classified model executed on patient database and extracted many rule, one rule could be

"IF (Age=65 AND Heart rate>70) OR (Age>60 AND Blood pressure>140/70) THEN Heart problem=yes" [60]

Prediction hit rate is used to measure the accuracy of extracted rules that how true they are to make prediction by applying on test data. Minimum hit rate is 80%, 100% is not impossible but hard to achieve [60, 61]. Different systems and models used supervised machine learning techniques are as follows:

Statistical Opinion Analyzer (SOA) [13] extract the polarity of online customer reviews using Bayesian probability and frequency distribution. The proposed system helps the new customer to purchase the product and manufacturer to enhance the product's functionality. Reviews crawled, preprocess, tagged (GO tagger) and insert in SOA to find the positive and negative opinion probability and frequency distribution as well. The proposed system originated the very promising results.

Linear regression with rule based approach used in **[40]** to rank the product features according to their importance. Proposed technique divide the overall product's rating into product features rating and generate the customer's survey automatically. Product features and their along sentiment words extracted using **Double Propagation** Algorithm. Proposed system (DPLR-R) is a two-stage method, first extract product features and opinion using the **state-of-the-art product feature extraction algorithm**. At the second stage, degenerate on the extracted product features and opinion units by exploiting overall ratings of reviews.

Opinion summarization model construct in [58], product specific data extract web and extract customer reviews in English and Spanish. Features extracted by using WordNet and ConceptNet along opinion. JavaRAP for English and SUPAR for Spanish used to replace the anaphoric references with corresponding references contain product features. LingPipe is used to split reviews into sentences. Minipar for English and FreeLing for Spanish used to parse the sentences. SMO SVM machine learning and the Normalized Google Distance are used to assign the polarity to features. Calculate the positive and negative polarity percentage probability. Experimental results show that the proposed method is highly effective.

[59] Presented a model that extracted data from NTCIR-6 Chinese opinion corpus. Chi- Square is used to extract the subjective cues from customer reviews. To find the subjectivity density from training data extracted subjective cues were used. Naïve Bayesian classifier applied with sentiment density subintervals for subjectivity classification. Proposed system achieved high recall of multi subjective term integration and proved effective to distinguish objective and subjective features.

To summarize customer feelings about any product, politician or any other object [60] used the ISREA and SemEval 2007 emotion wordlists based on WNA as training dataset. Opinion words were extracted in six different concepts such as ANGER, DISGUST, FEAR, JOY, and SADNESS. Concepts are divided into 6 clusters using Fuzzy-C clustering technique and then automatically assign the WordNet-Affect (WNA) scores to extracted opinions. Two steps process is used to classify the assignment of single WNA to each concept.

An upgrade FOMS model on Vietnamese reviews of mobile phone proposed in [61]. Synonym feature words were grouped by using HAC clustering and semi-SVM-kNN classification. Implicit feature words extracted using pre-constructed Adjective words and VietSentiWordNet orient the opinion words along with weights. Then opinion orientation for the feature in positive/ negative or neutral polarity based on the weight.

Syntax based pattern (SBP) used in [62] to enhance the subjective feature extraction from text. Syntax based pattern based on four type of words adjective, adverb, verbs and nouns. Linguistic features are extracted using syntactic information and apply Maximum Entropy Model (MEM) Classifier to find out objective and subjective sentences. They compare the results with Pangand Lee research [63] and evaluate that their proposed system achieved 92.1% accuracy.

To extract the features only from opinionated sentences [64] proposed an approach. Opinionated sentences extracted from reviews using SentiWordNet based algorithm. Fuzzy matching is used to handle the misspelling problem. Irrelevant features removed in sentence preprocessing step. Then extract the relevant explicit features from opinionated sentences using probability based algorithm and linguistic patterns. The proposed system achieved 0.908 averages Recall and 0.927 averages Precision.

Four machine learning techniques used for sentiment classification compared in [65]. Textual reviews and numerical satisfaction score reviews were extracted. Reviews were preprocessed, parsed and a dictionary based Naïve Count Method used to calculate polarity. Word based Unigram, Bigram, Chines Character Based Bigram and Trigram are used to extract the features. WBB and CBB are good in text classification using tfidf-c as feature weighting but on the whole Naïve Bayes Classifier achieved best performance under all conditions.

[89] Proposed a method to evaluate the quality of information from customer reviews using Support vector machine algorithm.

[90] Used SVM to explore the different domains of datasets. [93] Compared SVM, Bayesian classifier and content based N-gram modeling to classify the online customer reviews. They derived that SVM and Content based N-gram model overtake the Bayesian classifier.

[98] is about the comparison of Artificial Neural Network and SVM in sentiment classification of customer reviews. Experimented results showed that ANN is more promising to apply on movie reviews than SVM algorithm.

[68]Proposed a novel idea to extract the people habits, ideas and sentiments from twitter data. They applied the content analysis framework of Cheong and Lee’s to extract the hidden properties. Users based and message based extracted patterns are corpuses and clusters using Self-Organizing Map (SOM). Cluster visualization helped to find the distinguished attributes of a cluster in term of user contribution in specified topics. Then compared the SOM extracted results with k-means using minimal Euclidean distance results, they evaluate that their proposed framework is much efficient.

A system [74] is about to mine polarity of user comments and preserve information during text mining process. Reviews crawled, preprocess and parse using Stanford tagger and Parser. Domain ontology applied and using Bayesian Classifier finally classifies the polarity of reviews as positive, negative or neutral.

[75] Proposed a method to classify the overall polarity of call center services. They divided the call into three segments and find the polarity of each segment, a binary classifier considered to extract emotions as feature. Normalized Emotions Scores (NES), Normalized Emotion Category Scores (NECS), Last-K Emotions (LKE), and Dominant Emotions across M Intervals (DEMI) algorithms are used to classify the calls as positive or negative.

[76] Proposed a kernel-based machine learning approach to extract sentiment at sentence level by integrated multiple features from syntactic and lexical levels. Sentiment word, adjective and verb union used as input sentences. Then kernel approaches based on SVM SMO technique applied to measure similarity between instances.

Table 3 defines the different models and systems based on supervised machine learning techniques. Each system is defined in detail by describing the used datasets, techniques and achieved results.

Table 2: Supervised Machine Learning Opinion Mining Models

Model	Dataset	Technique	Results
A Feature Dependent Method for Opinion Mining and Classification	amazon.com,newegg.com, dealsdirect.com, ciao.es, shopmania.es, and quesabesde.com	SMO SVM machine learning Normalized Google Distance	Combined Precision: 0.79 Combined Recall: 0.79
Chinese subjectivity detection using a sentiment Density-based naive Bayesian classifier	NTCIR-6 Chinese opinion data	chi-square Naïve Bayesian classifier with sentiment density subintervals	Precision: 0.242 Recall: 0.950 f-score: 0.386
Enriching SenticNet Polarity Scores through Semi-Supervised Fuzzy Clustering	ISREA and SemEval 2007 emotion wordlists based on WNA	WordNet-Affect scores, Fuzzy-C clustering	Accuracy 92.15%
An Upgrading Feature-Based Opinion Mining Model on Vietnamese Product Reviews	Vietnamese mobile phone product reviews	HAC clustering semi-SVM-kNN classification	Purity = 0.7 Accuracy = 0.72 Entropy = 0.77
A Study of Opinion Mining and Visualization of Hotel Reviews	booking.com, TripAdvisor.com	Double Propagation Algorithm. (DPLR-R) state-of-the-art product feature extraction algorithm	Trip advisor: naïve Bayes: 57%, Dyn. LM Classifier: 57%, Booking: naïve Bayes: 44%, Dyn. LM Classifier: 48%
Linguistic Features for Subjectivity classification	movie review data	Syntax based pattern Maximum Entropy Model (MEM) Classifier	92.1% accuracy
Mining Product Features from Online Reviews	reviews for 5 product classes from Hu and Liu data	SentiWordNet based algorithm Fuzzy matching	Recall: 0.903 Precision: 0.927
Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics	Amazone.com	Bayesian probability Frequency Distribution	Accuracy: 72%
A Kernel-based Sentiment Classification Approach for Chinese Sentences	China cars reviews	Word Kernel Path Kernel N-gram Kernel	Precision: 80.93% Recall: 62.03% F-score: 70.23%
A Study on Detecting Patterns in Twitter Intra-topic User and Message Clustering	twitter data on 2009 Iran Election, iPhone OS 3.0 software launch	Self-Organizing Map k-means	Self-organizing feature map
Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian classification	Indian Hotel Reviews	NLP and Bayesian Classification	Accuracy 96.09%
Sentiment classification of online reviews to travel destinations by supervised machine learning approaches	Naïve Bayes, SVM and the character based N-gram model	travel blogs	Accuracy 80%
Web Opinion Mining: How to extract opinions from blogs?	Movie Reviews	AMOD approach	positive: 82.6% negative: 52.4%
Sentiment Identification Using Maximum Entropy Analysis of Movie Review	Movie Review	Maximum Entropy method	Accuracy 85%
Learning Extraction Patterns for Subjective Expressions”	U.S. Foreign Broadcast Information Service data	High precision Classifier	Recall 32.9Precision91.3 Recall 40.1 Precision 90.2
Document-level sentiment classification: An empirical comparison between SVM and ANN	Movies, GPS, Books and Cameras review dataset	SVM classifier ANN classifier	SVM Accuracy : 0.92% Ann Accuracy: 0.88%
Experiments with SVM to classify opinions in different domains	Pang corpus, SINAI corpus, Taboada corpus, Pang corpus	Support Vector Machines	Accuracy: 82.90% SVM Unigam 3-fold cross validation accuracy 91.51% with TFIDF

Emotions Analysis of Speech for Call Classification	Call center data	Normalized Emotions Scores , Normalized Emotion Category Scores Last-K Emotions and Dominant Emotions across M Intervals	Accuracy 79%
Which Side are You on? Identifying Perspectives at the Document and Sentence Levels	Bitter lemons articles	Naıve Bayes Model Latent Sentence Perspective Model SVM	SVM Editors 0.9724 NB-M Editors 0.9895 NB-B Editors 0.9909 SVM Guests 0.8621 NB-M Guests 0.8789 NB-B Guests 0.8859

6.2 Unsupervised Learning:

In contrast of supervised learning, unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so unsupervised learning is about to learn by observation instead of learn by example [61]. Clustering is a technique used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering [60]. Objects in one cluster are dissimilar to the objects in other clusters [62]. In opinion mining different clustering techniques are used as follows

To summarize customer feelings about any product, politician or any other object [60] used the ISREA and SemEval 2007 emotion wordlists based on WNA as training dataset. Opinion words were extracted in six different concepts such as ANGER, DISGUST, FEAR, JOY, and SADNESS. Concepts are divided into 6 clusters using Fuzzy-C clustering technique and then automatically assign the WordNet-Affect (WNA) scores to extracted opinions. Two steps process is used to classify the assignment of single WNA to each concept.

An upgrade FOMS model on Vietnamese reviews of mobile phone proposed in [61]. Synonym feature words are grouped by using HAC clustering and semi-SVM-kNN classification. Implicit feature words extracted using pre-constructed Adjective words and VietSentiWordNet orient the opinion words along with weights. Then opinion orientation for the feature in positive/ negative or neutral polarity based on the weight.

Bayes classifier and K-Means algorithm is used to clusters the feature words along opinions from sentence group in [66]. They proposed a method to conduct the opinion type answer questions on Web Pages The final extracted answers from sentences are presented in form of quaternion.

An unsupervised twice- clustering technique proposed in [67] to categories the product features. They used three types of context information, full context information, opinion words, group information of opinion words. They considered top 1/3 features as active product feature from the group of associated opinion word features, cluster them using COP K-Means algorithm and convert into “must link” and “cannot link” constraints. The experiment results showed that proposed model gives better quality results by considering the opinion word context instead of full words context.

A novel idea proposed in [68] to extract the people habits, ideas and sentiments from twitter data. They applied the content analysis framework of Cheong and Lee’s to extract the hidden properties. Users based and message based extracted patterns are corpuses and clusters using Self-Organizing Map (SOM). Cluster visualization helped to find the distinguished attributes of a cluster in term of user contribution in specified topics. Then compared the SOM extracted results with k-means using minimal Euclidean distance results, they evaluate that their proposed framework is much efficient.

[69] Proposed a method to relatively evaluate the video quality. Multiple attributes like motion, saliency and coding artifacts are measured to evaluate the video quality. The artifact based signatures of video are created and K-Means cluster technique used to cluster the videos according to created signatures. To identify the cluster’s center closet videos Mean Opinion score calculated. Total 34 representative clusters out of 100 were used for subjective assessment using Absolute Category Rating.

Sentiment based clustering by reducing object features [70] crawled the customer review data from web and sum up the review sentences of the same product. To tackle the implicit features Incomplete Information System established. As the web data consist on the high feature dimensions, they adopt the feature dimension reduction algorithm based on discernibility matrix. They not only extract the product opinion along feature but predict rank to each aspect of product. Aggregate product sentiment using K-Means clustering technique and predict a rank to each aspect of the product. Proposed method evaluated the effective clustering results. **[71]** Proposed a novel method to feature level opinion mining. Part of Speech dictionary and opinion words are used to extract the implicit and explicit features. Feature clusters created using K-means are based on three aspects: related opinion words, similarity of features and structure of features. Remove low frequency features (noise). Final clusters represented object’s polarity in one sight. Clusters are enhanced using background knowledge comes from

COP-KMeans. Experimental results show that the proposed method is highly effective.

A feature based opinion mining technique [72] utilize the syntactic dependency knowledge by differentiated nominal (nouns) and non-nominal (opinion words) terms. Data crawl from "dpreview.com", tagged data using OpenNLP tagger and parse with Malt parser. Then clusters of nominal terms are created by factorization method and refined clusters of nouns are obtained. Experimental results are compared with baseline technique and showed the high accuracy.

[73] Proposed a model to predict the subjective speech quality. Objective based Distance measured to represent perceptually-based parameter vectors to represent the voice part of speech. Calculated distance measures were mapped into equivalent Mean Opinion scores using regression. Self-Organizing Map a neural network algorithm was used to clusters the speech parameters. Wong et. al. [83] presented an unsupervised learning frame work to extract information from customer reviews as well as extract the knowledge based on feature based opinion mining.

Table 4 defines the different models and systems based on unsupervised learning techniques. Each system is defined in detail by describing the used datasets, techniques and achieved results. A comprehensive overview of opinion mining models based on unsupervised learning techniques is described in the following table

Table 3: Unsupervised Learning Models Of Opinion Mining

Model	Dataset	Technique	Results
A Framework To Answer Questions Of Opinion Type	Google, Baidu ,Yahoo, Bing	Bayes classifier k-means clustering	Accuracy: 71%
An Upgrading Feature-Based Opinion Mining Model On Vietnamese Product Reviews	Vietnamese Mobile Phone Product Reviews	HAC clustering semi-SVM-kNN classification	Purity = 0.7 Accuracy = 0.72% Entropy = 0.77
A Novel Product Features Categorize Method Based On Twice-Clustering	Www.360buy.Com.	K-Means COP K-Means	Accuracy: 66%
A Study On Detecting Patterns In Twitter Intra-Topic User And Message Clustering	Twitter Data On 2009 Iran Election, Iphone OS 3.0 Software Launch	Self-Organizing Map k-means	MAP Visualization
Enriching Senticnet Polarity Scores Through Semi-Supervised Fuzzy Clustering	ISREA And Semeval 2007 Emotion Wordlists Based On WNA	WordNet-Affect (WNA) scores Fuzzy-C clustering	Accuracy: 92.15%
Mining Web Videos For Video Quality Assessment	2,107 Youtube Videos	Mean Opinion score K-Means cluster	Accuracy: 82%
Sentiment Clustering Of Product Object Based On Feature Reduction	742 Reviews Documents About 13 Manufacturers And 68 Products	feature dimension reduction algorithm K-Means clustering technique	runtime reduced: 41.38%, sparsity reduced:15.69%,
Opinion Mining Based On Feature-Level	Chinese Reviews From Web	K-means COP-KMeans	Precision: 71% Recall: 61%
Product Feature Mining With Nominal Semantic Structure	Dpreview.Com	OpenNLP tagger Malt parser factorization method	More clusters with lower entropies
An Unsupervised Method For Joint Information Extraction And Feature Mining Across Different Web Sites	Digital Camera Domain, MP3 Player	undirected graphical model	Recall: 81.9% Precision: 81.2%

6.3 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 solution of past similar problem. If needed to little bit modify the solution according to new problem parameters, change the solution and stored in case base repository as a new solution instance. CBR cycle consists of four R's. Now days it is the most emerging technique used in opinion mining systems

[77] Presented the work in textual Case-Based Reasoning within jCOLIBRI. First phrase are detected and features are identified then a glossary is maintained which contains synonyms of the context. Then domain structure layer is used to assign identifiers to the data through rules. On the basis of rules local and global similarity calculated to find the cases polarity.

In [3] a sentence analysis is anticipated that depends on case-based reasoning rules. This methodology offers a 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.

To find the product features Fuzzy CBR technique used in [78]. Product database is based on the 100 attributes, 87 are used for use-scenario and 13 used for manufacturing scenario. CBR developed on the basis of used- scenario attributes retrieve the product ideas and enhanced its functionality. Manufacturing features based CBR monitor the retrieved product ideas to obtain the best ideas. Experiments show that the proposed system generates the higher percentage of good product ideas.

[79] Summarized a selection of techniques for software quality prediction using Case-Based Reasoning and Fuzzy logic. A software quality prediction CBR is based on analogy consisting on different parameters like no. of variables, LOC, No. of functions, difficulty level of software experience of programmer and mechanism to update the knowledge database when new case occurs etc. They used 70% training and 30% test data. For training data Prediction is 95.4% within 10% error and for testing data 91.3% within 10% error.

A system combination of data mining and CBR [81] is about to predict the retinopathy frequency in diabetes patients in Malaysia. System executed on three phases, first is to create the knowledge base by interviewing the medical experts and by literature review. Total 16 variables were determined from the knowledge base and 140 diabetes patient's data used as training files. C5.0 algorithm and CBR were used to develop the inference engine. KNN algorithm used in CBR to find similar cases and voting mechanism is used to find the final prediction. C5.0's accuracy 76% and CBR's showed 73 % accuracy but the integration of both techniques showed 85% accuracy.

To eliminate the issues of CBR tools such as outdated cases, stagnant case growth, lack of new cases, case updating, user's participation and user engagement [84] use the text mining and web 2.0 techniques to enhance the CBR's user experience and activities. Text mining used to extract metadata identifies the indexing terms and tags for case base. KITE CBR helps the teacher to integrate technologies into their teaching methodology. They take 50 special education related files preprocess the data and apply the text mining techniques and tools like concept extract, categorization and clustering. Then evaluate the generated results to find the patterns, issues, trends and models.

Statistical methods are combined with knowledge extracting techniques in [82] to enhance case searching, browsing and Reuse it for the problem solving methods using CBR from large scattered data set. NOEMIE concept created the problem domain knowledge base, establish parameters and conventions for relevant cases retrieval. Compare the new solution to the past case then testing and improving the measure of likeness through a learning process and at end Retained and generalized a new case as they are solved.

[83] Presented data mining technology into CBR system and GHSOM (Growing Hierarchical Self Organizing Map) using artificial neural network. The proposed system selects the features as case; GHMOS organized those cases as case base, past cases as sub-case base and for case retrieval new case guided into related sub-case base which improves the system accuracy. They used data mining techniques for case selection and retrieval. The experimented results show that the proposed CBR system can help designer to gain more accurate case and improve the efficiency of the design.

[107] presented a CBR system used to give the best strategy to local government to get approval for public sector small projects. The proposed system takes the past cases from cases repository as solution for new projects that helps in decision making. CBR takes the case from repository, prepare the case and retrieve it by Manhattan or City distance method. Reuse & revise the case and apply the final solution. The proposed system provide the efficient and quick decision making with high quality, impartial and strong decision by achieving 90% high accuracy. Table 5 defines the different models and systems based on Case Based Reasoning techniques. Each system is defined in detail by describing the used datasets, techniques and achieved results. A comprehensive overview of opinion mining models based on Case Based Reasoning technique is described in the described table.

Table 4: Case Based Reasoning Opinion Mining Models

Model	Dataset	Technique	Results
A Survey In The Area Of Machine Learning And Its Application For Software Quality Prediction	University Student's Data	Case-Based Reasoning And Fuzzy Logic	Accuracy: 91.3% Within 10% Error
A Fuzzy CBR Technique For Generating Product Ideas	1600 Products Data	Fuzzy CBR	Cell Phone Retrieving: 91.86% Cell Phone Retrieving+ Filtering: 47.78% Ball Retrieving: 82.64% Ball Phone Retrieving+ Filtering: 44.45%
Intelligent Project Approval Cycle For Local Government Case-Based Reasoning Approach	Health Care Project Data	Manhattan Or City Distance, CBR	Accuracy: 90%
Predictions Using Data Mining And Case-Based Reasoning: A Case Study For Retinopathy	Diabetes Patients Data In Malaysia	C5.0 Algorithm And CBR	Accuracy :85%
Combining Case Based Reasoning And Data Mining – A Way Of Revealing And Reusing Rams Experience	Schlumberger And Norsk Hydro Data	Data Mining + CBR	Accuracy: 92%
Research On CBR System Based On Data Mining	GHMOS, SOM data	Ghsom (Growing Hierarchical Self Organizing Map), CBR	TCBR Accuracy: 78% Ga-CBR : 84 % Proposed CBR: 94%
Improving User Experience With Case-Based Reasoning Systems Using Text Mining And Web 2.0	50 Special Education Related Files	Kite CBR	KITE CBR reduce effort time in blog site case retrieval
Extending Jcolibri For Textual CBR	Restaurant Data	Information Extraction Information Retrieval	IR+IE Accuracy: 75%

Conclusion:

Opinion miming is an emerging field of data mining used to extract the pearl knowledge from huge volume of customer comments, feedback and reviews on any product or topic etc. A lot of work has been conducted to mine opinions in form of document, sentence l and feature level sentiment analysis. Although many drawbacks exist in these sentiment classifications which are discussed in table 1 in detail. Many key challenges and gaps like infrequent feature extraction, opinion word strength, language orientation, writing styles, translate roman languages, detect fake reviews and many more are discussed as research directions in table 2. Current and future research areas clear a complete picture of opinion mining research scope. It is examined that now opinion mining trend is moving to the sentimental reviews of twitter data, comments used in Facebook on pictures, videos or Facebook status. A comprehensive literature review of multiple opinion mining models based on Supervised, unsupervised, and case based reasoning techniques is conducted to give a better understanding to future researchers.

Acknowledgement:

Authors are very thankful to Dr. Javed Anjum Sheikh (Associate Director, Faculty of Computer Science and Information Technology) for his value able research advices and encouragement. A very special thanks to my family and friends for their moral support.

References:

1. T. Khushboo "Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm" ,Int. Journal. Computer Technology & Applications, Vol 3 IJCTA | MAY-JUNE 2012.
2. J. Han and M. Kamber, "Data Mining: Concepts and Techniques", third edition Publication Date: July 6, 2011.
3. N, Anwer and A, Rashid "Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics" Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug. 2010.
4. B. Seerat and F.Azam "Opinion Mining: Issues and Challenges (A survey)" International Journal of Computer Applications (0975 – 8887) Volume 49– No.9, July 2012.
5. N. M. Shelke, S. Deshpande and V. Thakre "Survey of Techniques for Opinion Mining" International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
6. D. S. Deshpande " A Survey on Web Data Mining Applications" Emerging Trends in Computer Science and Information Technology -2012(ETCSIT2012) Proceedings published in International Journal of Computer Applications® (IJCA).
7. R. Kohavi and F. Provost "A research on web data mining and its application in electronic commerce". Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference onDate of Conference: 11-13 Dec. 2009.
8. A. Aamodt, H. A. Sandtorv, O. M. Winnem "Combining case base Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience" Safety and Reliability; Proceedings of ESREL '98, Trondheim, June 16-19, 1998.
9. D. Thain, T, Tannenbaum and M. Livny "Distributed Computing in Practice: The Condor Experience" Concurrency and Computation: Practice & Experience - Grid Performance Volume 17 Issue 2-4, February 2005 Pages 323 – 356.
10. I. Foster , C. Kesselman "A Metacomputing Infrastructure Toolkit" nternational Journal of High Performance Computing ApplicationsNovember 1, 2001.
11. Zakim J, Pan Y, "Introduction: recent developments in parallel and distributed data mining", Journal of Distributed Parallel Database, Vol. 11, pp. 123-127, 2002.
12. Tim Haithcoat "Relational Database Management Systems, Database Design, and GIS".
13. M Fan, G WU "Opinion Summarization of Customer comments" International conference on Applied Physics and Industrial Engineering in 2012.
14. Pang, B. and Lee, L. 2004, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", Proceedings of the Association for Computational Linguistics ACL 2004.
15. Hulth, A. and Megyesi, B.B., "A Study on Automatically Extracted Keywords in Text Categorization", Proceedings of the 21st International Conference on Computational Linguistics in 2006.
16. Savary, L. and Zeitouni, K. "Indexed bit map (ibm) for mining frequent sequences", PKDD Knowledge Discovery in Databases: PKDD, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases', Vol. 3721 of Lecture Notes in Computer Science, Springer, Porto, Portugal, in 2005.
17. Harb, A., Plantié, M., Dray, G., Roche, M., Trousset, F. and Poncelet, P. , "Web Opinion Mining: How to extract opinions from blogs?", International Conference on Soft Computing as Tran disciplinary Science in 2008.
18. ZhongchaoFei, Jian Liu, and Gengfeng Wu: "Sentiment Classification Using Phrase Patterns", Proceedings of the Fourth International Conference on Computer and Information Technology in 2004.
19. Lin,W.H., Wilson,T., Wiebe,J. and Hauptmann, A., "Which Side are You on? Identifying Perspectives at the Document and Sentence Levels", Proceedings of the 10th Conference on Computational Natural Language in 2006.
20. Camelin, N., Damnati, G., Béchet, F. and De Mori, R., "Opinion Mining in a Telephone Survey Corpus", International Conference on Spoken Language Processing in 2006.
21. Mehra, N., Khandelwal, S. and Patel, P., "Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews", Stanford University, USA in 2002.
22. Riloff, E. and Wiebe, J. "Learning Extraction Patterns for Subjective Expressions", Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language in 2003.
23. Yu,H. and Hatzivassiloglou ,V. 2003,"Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences", Proceedings of EMNLP-03, 8th Conference on Empirical Methodsin Natural Language Processing pages 129–136 in 2003.
24. Zaki, M. J. "Efficient enumeration of frequent sequences", in '7th International Conference on Information and Knowledge Management', ACM Press, Bethesda ,Maryland, United States, pp. 68–75 in 1996.
25. Wilson, T., Wiebe, J. and Hwa, R, "Just how mad are you? Finding strong and weak opinion clauses", Proceeding of National Conference on Artificial Intelligence in 2004.
26. Rillof,E., Wiebe, J., and Philips, W., "Exploiting Subjectivity Classification to Improve Information Extraction", Proceedings of the 20th National Conference on Artificial Intelligence in 2005.
27. Kim, Y. and Myaeng, S., "Opinion Analysis based on Lexical Clues and their Expansion", Proceedings of NII Test Collection for Information Retrieval in 2007.
28. Y. Kim, S.-H. Myaeng, "Opinion Analysis based on Lexical Clues and their Expansion Information and Communications" Proceedings of NTCIR-6 Workshop Meeting, Tokyo, Japan in May 15-18, 2007
29. Hiroshi, K., Tetsuya, N., and Hideo, W. "Deeper sentiment analysis using machine translation technology". In Proceedings of the 20th international Conference on Computational Linguistics (Geneva, Switzerland) in 2004.
30. Esuli, A. and Sebastiani, F. "SentiWordNet: A publicly Available Lexical Resource for Opinion Mining", Proceedings of the Fifth International Conference on Language Resource and Evaluation (LREC'06) on Artificial Intelligence in 2006.
31. P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424, 2002.
32. D Wang, S Zhu, and T Li. "SumView: A Web-based engine for summarizing product reviews and customer opinions" Expert Syst. Appl. 40(1):27-33 (2013).
33. B. Liu "Sentiment Analysis and Opinion Mining", April 22, 2012.
34. Ah-hwee Tan "Text Mining: The state of the art and the challenges", In Proceedings of the PAKDD Workshop on Knowledge Discovery from Advanced Databases in 1999.
35. Kim, Y. and Myaeng, S., "Opinion Analysis based on Lexical Clues and their Expansion", Proceedings of NII Test Collection for Information Retrieval in 2007.
36. D. K. Evans, L.-W. Ku, Y. Seki, H.-H. Chen, and N. Kando, "Opinion analysis across languages: An overview of and observations from the NTCIR6 opinion analysis ilot task," Applications of Fuzzy Sets TheoryLecture Notes in Computer Science Volume 4578, 2007, pp 456-463
37. L. Dini and G. Mazzini, "Opinion classification through information extraction," in Proceedings of the Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields Data Mining in 2002.
38. Ana-Maria Popescu , Oren Etzioni, "Extracting Product Features and Opinions from Reviews", Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, p.339-34 in 2005.
39. M. Baroni and S. Vegnaduzzo "Identifying subjective adjectives through web-based mutual information". In E. Buchberger, editor, In Proceedings of the Conference for the Processing of Natural Language and Speech KONVENS), pages 17--24, in 2004.
40. T. Yao and L. Li "A Kernel-based Sentiment Classification Approach for Chinese Sentences" World Congress on Computer Science and Information Engineering in March 31 2009-April 2 2009.
41. Q. Su, X. Xu1, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen1and Z. Su "Hidden Sentiment Association in Chinese Web" Proceedings of the 17th international conference on World Wide Web Pages 959-968 in 2008.

42. M.S. BinWahlan, N, Salim and L. Suanmali, "Swarm Based Features Selection for Text Summarization", International Journal of Computer Science and Network Security", 9(1): pp. 175-179 in 2009.
43. Jin, W. and Ho, H.H., "A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining", In Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada, 2009.
44. Wilson, T., Wiebe, J. and Hoffmann, P., "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", Proceeding of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing in 2005.
45. Whitelaw, C., Grag, N. and Argamon, S., "Using Appraisal Groups for Sentiment Analysis", Proceedings of the ACM SIGIR Conference on Information and Knowledge Management in 2005.
46. D. Wang, S. Zhu, and T Li. "SumView: A Web-based engine for summarizing product reviews and customer opinions" Expert Syst. Appl. 40(1):27-33 (2013).
47. E. BjørkelundTH and BurnettKNørvåg "A Study of Opinion Mining and Visualization of Hotel Reviews" Proceedings of the 14th International Conference on Information Integration and Web-based Applications &ServicesPages 229-238 in 2012.
48. Su-Ke Li, Zhi Guan, Li-Yong Tang and Zhong Chen "Exploiting Consumer Reviews for Product Feature Ranking" Journal of Computer Science and TechnologyJanuary Volume 27, Issue 3, pp 635-649 in 2012.
49. Jin, W. and Ho, H.H. 2009, "A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining", In Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada in 2009.
50. W. Wang, H. Xu and W. Wan "Implicit feature identification via hybrid association rule mining", ESWA 8310 No. of Pages 14, Model 5G 3 January 2013.
51. A. Aamodt, H. A. Sandtorv, O. M. Winnem "Combining Case Based Reasoning and Data Mining - A way of revealing and reusing RAMS experience" Safety and Reliability; Proceedings of ESREL, Trondheim, June 16-19, 1998.
52. R. F. Brena, J. L. Aguirre, C. Chesñevar, E. H. Ramírez and L. Garrido "Intelligent Agents: The Right Information at the Right Time" Knowledge and Information Systems, Volume 12, Issue 2, pp 203-227 in 2007.
53. A. BALAHUR and A. MONTOYO, "A Feature Dependent Method for Opinion Mining and Classification" Natural Language Processing and Knowledge Engineering, 2008. NLP-KE '08. International Conference on Date of Conference: 19-22 Oct. 2008.
54. X. WANG, G. HONG FU "Chinese subjectivity detection using a sentiment density-based naive bayesian classifier" conference on machine learning and cybernetics (icmlc) international in 2010.
55. V Balakrishnan, Mohammad R. Shakouri, H Hoodeh and Loo, Huck-Soo, "Predictions Using Data Mining and Case-based Reasoning: A Case Study for Retinopathy" international journal of computer and information engineering 6, 2012.
56. S.Poria, A. Gelbukh, E. Cambria, D.r Das, S. Bandyopadhyay, ""Enriching SenticNet Polarity Scores Through Semi-Supervised Fuzzy Clustering" IEEE 12th International Conference on Data Mining Workshops in 2012.
57. Z. ZHAI, H. XU, J. LI and P. JIA "Sentiment Classification for Chinese Reviews Based on Key Substring Features" Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference onDate of Conference: 24-27 Sept. 2009
58. Q. ThuyHa,T. ThanhVu,C. To Luu "An Upgrading Feature-Based Opinion Mining Model on Vietnamese Product Reviews"AMT'11 Proceedings of the 7th international conference on Active media technology Pages 173-185 in 2011.
59. H. Nguyen, T. Xuan, A. Cuong Le, L. M. Nguyen, "Linguistic Features for Subjectivity classification" International Conference on Asian Language Processing International Conference on Asian Language Processing in 2012.
60. W. Hu, Z. Gong, J. Guo, "Mining Product Features from Online Reviews" e-Business Engineering (ICEBE), IEEE 7th International Conference onDate of Conference: 10-12 Nov. 2010.
61. X. Su, G. Gao, Y. Tian, "A Framework to Answer Questions of Opinion Type" Seventh Web Information Systems and Applications Conference in 2010.
62. W. J. Jia, S. Zhang, Y.Ju Xia, J. Zhang and H. Yu "A Novel Product Features Categorize Method based on Twice-Clustering", International Conference on Web Information Systems and Mining in 2010.
63. M. Cheong, V. Lee "A Study on Detecting Patterns in Twitter Intra-topic User and Message Clustering" International Conference on Pattern Recognition 2010.
64. D. Culibrk, M. Mirkovic, P. Lugonja, V.Crnojevic, "Mining Web Videos for Video Quality Assessment" Soft Computing and Pattern Recognition (SoCPaR), International Conference of Date of Conference: 7-10 Dec. 2010.
65. S. Wang ; Taiyuan X. Yin ; J. Zhang and Ru Li "Sentiment clustering of product object based on feature reduction" Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on Date of Conference: 29-31 May 2012.
66. X. Ding, B. Liu and P. S. Yu "A Holistic Lexicon-Based Approach to Opinion Mining" Proceedings of the International Conference on Web Search and Data Mining Pages 231-240 in 2008.
67. Timothy L. Acorn, Sherry H. Walden "Smart: Support: Management Automated Reasoning Technology for Compaq Customer Service" In proceeding of: Proceedings of the Fourth Conference on Innovative Applications of Artificial Intelligence in 1992.
68. K. Tateishi, Y. Ishiguro, and T. Fukushima, "Opinion information retrieval from the internet," Information Processing Society of Japan (IPSJ) SIG Notes, 2001, vol. 69, no. 7, pp. 75–82, 2001.
69. Watson, I. and Marir, F., "Case-Based Reasoning, published in the knowledge engineering review vol.9 No. 4 in 1994.
70. Recio1,J A. Agudo1,B D. G´omez-Mart´in1,M A and Wiratunga2,N., "Extending jCOLIBRI for Textual CBR", Proceedings of the Sixth Spanish Conference on Programming and Languages, (SCPL'05) in 2005.
71. MC Wu, YF Lo and SH Hsu, "A fuzzy CBR technique for generating product ideas" Expert Systems with Applications, 34 (1), pg. 530-540 January 2008.
72. E. Rashid, S. Patnayak and V. Bhattacherjee, "A Survey in the Area of Machine Learning and Its Application for Software Quality Prediction" ACM SIGSOFT Software Engineering Notes Volume 37 Issue 5, September 2012.
73. M. K.. Farooq, M. Khan, S. Shamail and Mian M. Awais, "Intelligent Project Approval Cycle for Local Government Case-Based Reasoning Approach" In proceeding of: ICEGOV 2009, Proceedings of the 3rd International Conference on Theory and Practice of Electronic Governance, Bogota, Columbia, November 10-13, 2009.
74. T. jie Zhan and C.-hung Li, "Product Feature Mining with Nominal Semantic Structure" IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology in 2010.
75. A.A. Kardan, M. Ebrahimi "A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups" Information Sciences 219 93–110 international journal in 2013.
76. D. Martens "Predicting going concern opinion with data mining" Decision Support Systems 45 765–77 international journal in 2008.
77. K. Xu et al. "Mining comparative opinions from customer reviews for Competitive Intelligence" Decision Support Systems 50 743–754 international journal in 2011.
78. C.C. Chen, Y.-D. Tseng "Quality evaluation of product reviews using an information quality framework" Decision Support Systems 50 755–768 international journal in 2011.
79. O. Vechtomova "Facet-based opinion retrieval from blogs" Information Processing and Management 46 71–88 international journal in 2010.
80. L. Robaldo, L. Di Caro "Opinion Mining" Computer Standards & Interfaces international journal in 2012.
81. S.-T. Li, F.-C. Tsai "A fuzzy conceptualization model for text mining with application in opinion polarity classification" Knowledge-Based Systems 39 23–33 international journal in 2013.
82. Q. Ye et al. "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches" 2Expert Systems with Applications 36 6527–6535 international journal in 2009.

83. Y.-H. Tung et al. "A rule-based CBR approach for expert finding and problem diagnosis" Expert Systems with Applications 37 2427–2438 international journal in 2010.
84. M. RushdiSaleh "Experiments with SVM to classify opinions in different domains" Expert Systems with Applications 38 14799–14804 international journal in 2011.
85. H. Kang et al. "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews" Expert Systems with Applications 39 6000–6010 international journal in 2012.
86. L. Chen et al. "Comparison of feature-level learning methods for mining online consumer reviews" Expert Systems with Applications 39 9588–9601 international journal in 2012.
87. R. Moraes et al. "Document-level sentiment classification: An empirical comparison between SVM and ANN" Expert Systems with Applications 40 621–633 international journal in 2013.
88. Richard A. Huebner "A survey of educational data" research in higher education journal Volume 19 - April, 2013
89. R. Kaur and S. Aggarwal "A Survey of Text Mining Techniques and Applications" JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 1, NO. 1, AUGUST in 2009
90. F. Sebastiani "Text Mining" International Journal of Database Theory and Application Vol. 4, No. 3, September, 2011
91. Md. Z. Hasan, K. J. A. Chisty and N. Zaman Ayshik "Research Challenges in Web Data Mining", International Journal of Computer Science and Telecommunications Volume 3, Issue 7, July 2012.
92. Hong Yu, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences" Proceeding EMNLP '03 Proceedings of the conference on Empirical methods in natural language processing Pages 129-136 in 2003.
93. X. Su, G.i Gao, Y. Tian, "A Framework to Answer Questions of Opinion Type "Seventh Web Information Systems and Applications Conference in 2010.
94. W. J. Jia, S. Zhang, Y.Ju Xia, J. Zhang and H. Yu "A Novel Product Features Categorize Method based on Twice-Clustering",2010 International Conference on Web Information Systems and Mining in 2010.
95. M. Cheong, V. Lee "A Study on Detecting Patterns in Twitter Intra-topic User and Message Clustering" 2010 International Conference on Pattern Recognition 2010.
96. D. Culibrk, M. Mirkovic, P. Lugonja, V.Crnojevic, "Mining Web Videos for Video Quality Assessment" Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of Date of Conference: 7-10 Dec. 2010.
97. S. Wang ; Taiyuan X. Yin ; J. Zhang and Ru Li "Sentiment clustering of product object based on feature reduction" Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on Date of Conference: 29-31 May 2012.
98. T. jie Zhan and C.-hung Li, "Product Feature Mining with Nominal Semantic Structure" 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology in 2010.
99. Picovici, D and Mahdi, A.E, "New output-based perceptual measure for predicting subjective quality of speech" Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on Date of Conference: 17-21 May 2004.
100. X. Ding, B. Liu and P. S. Yu "A Holistic Lexicon-Based Approach to Opinion Mining" Proceedings of the 2008 International Conference on Web Search and Data Mining Pages 231-240 in 2008.