Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 5 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Sentiment Analysis and Opinion Mining from Social Media : A Review

By Savitha Mathapati, S H Manjula & Venugopal K R
University Visvesvaraya College of Engineering

Abstract - Ubiquitous presence of internet, advent of web 2.0 has made social media tools like blogs, Facebook, Twitter very popular and effective. People interact with each other, share their ideas, opinions, interests and personal information. These user comments are used for finding the sentiments and also add financial, commercial and social values. However, due to the enormous amount of user generated data, it is an expensive process to analyze the data manually. Increase in activity of opinion mining and sentiment analysis, challenges are getting added every day. There is a need for automated analysis techniques to extract sentiments and opinions conveyed in the user-comments. Sentiment analysis, also known as opinion mining is the computational study of sentiments and opinions conveyed in natural language for the purpose of decision making.

Keywords: domain adaptation; machine learning; opining mining; sentiment analysis; sentiment classification.

GJCST-C Classification : H.2.8,J.4

SENTIMENTANALYSISANDOPINIONMININGFROMSOCIALMEDIAA REVIEW

Strictly as per the compliance and regulations of:

Sentiment Analysis and Opinion Mining from Social Media: A Review

Savitha Mathapati [α], S H Manjula [σ] & Venugopal K R [ρ]

Abstract- Ubiquitous presence of internet, advent of web 2.0 has made social media tools like blogs, Facebook, Twitter very popular and effective. People interact with each other, share their ideas, opinions, interests and personal information. These user comments are used for finding the sentiments and also add financial, commercial and social values. However, due to the enormous amount of user generated data, it is an expensive process to analyze the data manually. Increase in activity of opinion mining and sentiment analysis, challenges are getting added every day. There is a need for automated analysis techniques to extract sentiments and opinions conveyed in the user-comments. Sentiment analysis, also known as opinion mining is the computational study of sentiments and opinions conveyed in natural language for the purpose of decision making. Preprocessing data play a vital role in getting accurate sentiment analysis results. Extracting opinion target words provide fine-grained analysis on the customer reviews. The labeled data required for training a classifier is expensive and hence to over come, Domain Adaptation technique is used. In this technique, Single classifier is designed to classify homogeneous and heterogeneous input from di_erent domain. Sentiment Dictionary used to find the opinion about a word need to be consistent and a number of techniques are used to check the consistency of the dictionaries. This paper focuses on the survey of the existing methods of Sentiment analysis and Opinion mining techniques from social media.

Keywords: domain adaptation; machine learning; opining mining; sentiment analysis; sentiment classification.

I. Introduction

Due to the huge growth of social media on the web, opinions extracted in these media are used by individuals and organizations for decision making. Each site contains a large amount of opiniated text which makes it challenging for the user to read and extract information [1]. This problem can be overcome by using sentiment analysis techniques. The main objective of sentiment analysis is to mine sentiments and opinions expressed in the user generated reviews and classifing it into different polarities. The output is the data annotated with sentiment labels. Machine learning techniques are widely used for sentiment classification [2]. For a specific domain *D*, sentiment data *Xi* and *Yi* denoting data *Xi* has polarity *Yi*. If the overall sentiment expressed in *Xi* is positive, then *Yi* is +1, else -1. Labelled sentiment data is a pair of sentiment text and its corresponding sentiment polarity *fXi,Yig*. If *Xi* is not assigned with any polarity data *Yi*, then it is a unlabelled sentiment data. In supervised sentiment classi_cation method, classifier is trained using labeled data from a particular domain. Semisupervised classification method, combines unlabeled data along with few labeled training data to construct the classifier [3].

Applications: There are variety of information in the form of news blogs, twitter *etc..* are available in the social media about different products. Sentiment Analysis can summarize and give a score that represents the opinion of that data. This is used by customers depending on their need. There are a number of applications of sentiment analysis and opinion mining. The area where Sentiment

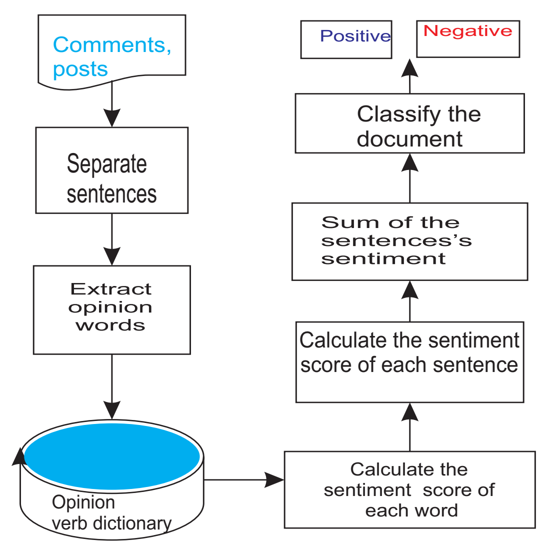

Figure 1: Architecture of Sentiment Analysis

Analysis is used is in Finance, Politics, Business and public actions. In business Domain, Sentiment analysis is used to detect the customer's interest in their product. Sentiment analysis in political do-main is used to get the clarity on the politician's position. Opinion Mining is also used to find the public interest on the newly applied rules by the goverenment.

Motivation: Current trend is to look for opinions and sentiments in the product reviews that are available in large scale in social media. Before making decision, we tend to look at the sentiment analysis results of the opinion given by different users. This helps any customer to decide his opinion on that product. As data available in large scale, it is a laborious process to look

Author α σ ρ: Department of Computer Science and Engineering University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India. e-mail: hiremathsavitha@gmail.com

into all the user opinion. Hence Sentiment analysis is require. The main Objective of sentiment analysis is to classify the sentiment into different categories. Figure 1, shows the overall architecture of the sentiment analysis. Document level, sentence level and aspect level are the different levels of sentiment classification. Classifying each document into positive or negative class is called document-level sentiment classification. While expressing the sentiment of a document by this type of classifier, it assumes that document contains opinion of the user about a single object. Aspect level sentiment analysis classify the opinion about a document assuming that the opinion is expressed about different aspects in a document.

Sentiment classifiers, designed using data from one domain may not work with high accuracy if the same is used to classify the data from a different domain. One of the main reasons is that the sentiment words of a domain can be different from another domain. Thus, Domain adaptations are required to bridge the gaps between domains. The Domain used to train the classifier is called source domain and the domain to which we apply the trained classifier is called the target domain. The advantage of this method is that we need some or no labeled data of the target domain, where labeled data is costly as well as in-feasible to manually label the reviews for each domain type. This type of classification is called Cross Domain Sentiment Classification. Heterogeneous domain adaptation is required when domains of different dimension are input to the topic adaptive sentiment classifier.

Sentiment classifiers can be broadly classified into machine learning and lexicon based. Machine learning algorithms are used in machine learning approach. These algorithms can work in supervised, semi-supervised or unsupervised learning methods. Supervised learning methods give more accurate results compared to semi-supervised and unsupervised learning methods, but it requires labeled data which is expensive and time consuming process. Semi-supervised approach uses Easy Adapt (++[EA++]) which is easier than the Easy Adapt [EA] which requires labeled data from source and target domain. This is because it uses both labelled and unlabeled data from the target domain which results in superior performance theoretically and empirically over EA and hence it can be efficiently used for preprocessing [4]. Lexicon based approach utilizes Sentiment lexicon to analyze the sentiments in a review. Lexicon based approach can use dictionary or corpus to classify the sentiment words. Due to the shortage of labeled data, a single classifier can be designed to classify reviews from different domains. But classifier designed to classify data from one domain may not work efficiently on other domain. This is due to domain specific words which are different for every domain.

Support vector machine and Naive baye's classifiers are the important classifiers that support machine learning approach. Support vector machine classify data by finding hyper-plane that separates into different classes. Naive Baye's classifier is a probabilistic classifier based on Bayes theorem and the strong independence between the features. As there is a shortage of labeled data, a single classifier can be designed to classify reviews from different domains. But classifier designed to classify data from one domain may not work efficiently on other domain. This is due to domain specific words which are different for different domain.

Organization: The paper is organized as follows. Section 2 deals with the difffeerent techniques of data Preprocessing. In Section 3, Domain Adaptation Methods are discussed along with importance and applications of Heterogeneous Domain Adaptation. Section 4 give a comparison of different Topic Adaptive Sentiment Classi_cation methods. Sections 5 and 6 gives an overview of the Extracting Opinion Targets and Words and Different levels of Sentiment Analysis respectively. Section 7 gives a brief overview on how to work on inconsistent dictionaries. Dificulties and Solutions of the Polarity Shifting Detection are discussed in Section 8. Intrinsic and Extrinsic Domain Relevance is discussed in section 9. Section 10 contained information regarding Content-Based and Policy- Based Filtering policies. Section 11 brief about the Evaluation methods and paper is concluded in Section 12.

II. Preprocessing Data

Data provided in the form of reviews by the users contain lot of noise which need to be removed before it is classified. Haddi *et al.*[5] have explored the role of preprocessing in improving the SVM classifier results by selecting appropriate features. Selection of relevent features increase the accuracy of the classification process. Different techniques used are Feature Frequency, Term Frequency Inverse Document Frequency, Feature Presence. Boa *et al.* [6] show the effect of *urls*, repeated letters, negation, lemmatization and stemming on the performance of the classifier. Bigrams and emotion features addition improves the accuracy of the classifier [7].

There are mainly three steps in data processing, to-kenization, normalization and part-of-speech(POS) tagging. Transfering inected form to base form, also known as lemma is called lemmatization. This reduces the sparseness of the data which make text classification efficient [8]. Stemming processes a word without knowledge of the context. Whereas lemmatization considers contextual part-of-speech information while finding the base form of a word.

Global Journal of Computer Science and Technology (C) Volume XVI Issue V Version I Year 2016

Unigrams and bigrams can be selected as training features. Pang *et al.* [9] show that unigrams turned out to be more effective compared to using bigrams. This leads to less features which give high performance. Stop words are excluded as they are not helpful for our classification.

III. Domain Adaptation Methods

Domain adaptation methods have been used for diffe- rent research fields. According to the data in the target domain, the domain adaptation methods are generally divided into three categories: Supervised, Semi-super- vised and Unsupervised domain adaptation methods. Supervised domain adaptation only use the labelled data in the target domain, Semisupervised domain adaptation use both the labelled and unlabelled data in the target domain and Unsupervised domain adaptation use only the unlabeled data in the target domain [4], [10]. Xavier *et al.* [11] proposed an efficient method for domain adaptation without the requirement of labeled data. This method classifies reviews from multiple domains by extracting the topic adaptive words from the unlabeled tweets using deep learning approach. SUIT model [12] considers the topic aspects and opinion hol- ders for domain adaptation using supervised learning.

Daume *et al.* [18] proposed a feature augmentation method for domain adaptation. This method augments the source domain feature space using feature from labeled data from the target domain. Cheng *et al.* [17] proposed semi supervised domain adaptation method that maps source to target feature space. Methods proposed in [19] donot consider labeled data while considering learning feature representation. Ando *et al.* [20] proposed multitasking algorithm to select pivot features between source and target domains which is used to build pseudo-tasks for building correspondence among the features.

Structural Correspondence Learning uses unlabeled data from both source and target domain to obtain common features referred to as Pivots which behave in the same way in both domains and to find the correspondence between them. Non-pivot features which co-occur with pivot features are also considered. This technique is tested on the part of speech tagging and show the gain in performance for varying amount of source and target training data [21].

Table 1: Summary of the Survey of Domain Adaptation Techniques

Author	Concept	Advantages	Disadvantages
Bollegala *et al.*, (2016), [13]	Project both source and target domain in same lower dimensional embedding and then learning classifier on this embedded feature	Only source domain labeled data is used	Single rule is applied at a time
Liu *et al.*, (2015), [14]	Updates topic adaptive features based on collaborative selection of unlabelled data	Single classifier classifies multiple topic tweets	Few topic adaptive sentiment words are not selected due to the threshold applied while selecting the words
Quynh *et al.*, (2015), [15]	linguistic resources are used to generate additional training examples	Percentage of new training examples is high	Errors of syntactic parsing may cause problems
Xiao *et al.*, (2015), [16]	Feature Space Independent semi-supervised domain adaptation	Both Homogeneous and Heterogeneous domain adaptation is implemented	-
Cheng and Pan (2014), [17]	Linear Transformation from source to target damain is used with Semi-Supervised Adaptation	Method can be used in general for all variety of loss functions	Practical Domain Adaptation problems are not Considered

J Blitzer et al [22] proposed a method where the SCL algorithm is extended which reduces the error between the domains by 30 to 46 percent over supervised baseline. Movies reviews are the most studied domain in the early days, but at present the number of domains are increasing widely. The sentiment classification system has to collect data for each new domain. The pivots are selected not only by considering the frequency of occurrence but also by using the manual information of the source labels by using very small number of labeled information. The distance between the domains is obtained which is the measure of loss due to domain adaptation from one to another. Spectral Feature Alignment require only small amount of

Global Journal of Computer Science and Technology (C) Volume XVI Issue V Version I Year 2016

source domain labeled data and no label data from target domain. To span the gap between source and target domain spectral feature alignment algorithm is used to align the domain specific words into a unified cluster with the help of domain independent words as a bridge. SFA provides a new representation of cross-domain data by using the relationship between domain specific and domain independent features(pivots) by clustering them into the same latent space. These clusters reduce the mismatch between domain specific words of both domains.The classifier is trained on the new representation.

Bipartite graph is constructed to study the relationship between domain specific and domain independent words [23], [24]. A sentiment sensitive the saurus is created by using labeled data from diverse source domains and unlabeled data from both source and target domains to find the association between the words in different domains. The created thesaurus is used to expand the feature vector to train the binary classifier. The feature vector expansion is done by appending the additional features that represent the source and target domain reviews to minimize the mismatch of features [25], [13].

Locality Preserving Projections is a linear projective map that emerges by resolving the different problem and by maintaining the locality of the constitution of data set. When two data overlap on the other, with the decreasing dimensionalities in the ambient space the Locality Preserving Projections are derived by determining the optimal linear estimations to the eigen functions of the Laplace Beltrami operator. Training and trial data when drawn from same distribution methods of Discriminative learning execute well. Infinite number of labeled data are available for source domain, but, focus is to find a classifier that performs effectively on target with little or no labeled data. First, we have to evaluate the conditions on which the classifier performs well on the target domain. Second, having compact labeled data for target domain and huge labeled data for source domain we need to combine them during training to attain minimum mistakes at test time [26].

a) Heterogeneous Domain Adaptation

Domain adaptation methods assume that the data from different domains are represented by the same type of features with same dimensions. These methods cannot classify if the dimensions of source and target data are different. Technique of classifying such data is called Heterogeneous domain adaptation. Shi *et al.* [27] propose a solution where classification of high accuracy can be obtained even with the different feature space and different data distribution. Spectral embedding is used to unify the feature space of both source and target domains. It proposes Heterogeneous spectral mapping to find the common feature subspace by understanding two feature mapping matrix. Gap between the two domains in Domain adaptation methods can be achieved by re-weighting source instances [28], [29], target instances are self-labeled [30], [31], introducing new feature representations [22], [32], [33]. These methods can be applied when both domains have same feature representations. In real world, feature representation in the source domain can be completely different from target do-main while doing cross domain sentiment classification. Example for this is cross language text classification, where reviews from different language domains are represented by words in different languages. Text-aided image classification can also be executed where source domain has word features and target domain has visual features.

Number of approaches are employed for hetero geneous domain adaptation, such as heterogeneous spectral mapping [27], feature mapping, feature projection and transformation [34], [35], manifold alignment [36] and auxiliary resources [37]. Xiao *et al.* [16] proposed a method which can do homogeneous and heterogeneous domain adaptation across domains. In this process, source domain is assumed to have large set of labeled data and unlabeled data compared to target domain data. Instead of focusing on the feature divergence, each domain instances are employed kernelized representation. Table 1 gives the summary of the survey of various Domain Adaptation techniques.

IV. Topic Adaptive Sentiment Classification

Sentiment classifier trained using data from one domain may not give a good accuacy if the same classifier is used to classify data from different do-main. For example, sentiment words of kitchen do-main are different from book domain. Blitzer *et al.* [22] proposed an approach called structural correspondance learning for domain adaptation where it used pivot features to bridge the gap between source and target domain. Pan *et al.* [23] proposed a method called spectral feature alignment where domain specific words from different domains are aligned into unified clusters. Bollegala *et al.* [25] proposed a method for classification when we do not have labeled data of target domain, but we have few labeled data of other domains. This method automatically creates a sentiment sesitive Thesaurus using labeled and unlabeled data from multiple source domains. Constructed Thesaurus is then used to enlarge the feature vectors to train the classifier. Choi *et al.* [38] proposed linear integer programming method that can adapt an existing lexicon into a new one and find the relations among words and opinion expressions to find the most likely polarity of each lexicon item for the given domain.

Subjectivity analysis is concerned with extracting infor- mation about opinions, sentiments and other private states expressed in texts. Stoyanov *et al.* [39] proposed a method which collectively considers the relation among words and opinion statements to get the polarity of the sentiment words of the given domain. He *et al.* [40] and Gao *et al.* [41] gave a probabilistic topic model which bridge each pair of domains in a semantic level. Compared to review data, Twitter data contain more variety topics from various domains. To train a topic specific classifier, labeled data is required. Aspect level sentiment analysis detect topic, relation of topic aspects, opinion words and sentiment holders in a document [42], [43]. Supervised learning is used in SUIT model [44] considering topic aspects and opinion holders for cross domain sentiment classification. Mejova *et al.* [45], [46] have shown that by considering news, blogs and twitter data set, cross media sentiment classification can be done. Shenghua Liu *et al.* [14] proposed that a classifier designed using multiple domain twitter inputs can be used as a specific classifier to classify tweets from a specific domain. Microblogs as a social media has become an interesting input for sentiment analysis [47], [48], [49]. Tumasjan *et al.* [50] concluded that twitter messages are more oriented towards the political opinion. Supervised learning of a sentiment classifier need labeled tweets which is expensive and rarely available.

Semi-supervised Support vector machine is one of the efficient model which classify data with less labeled data and utilizing more unlabeled data. When features can be easily split into different views, co-training framework [51] achieves good results.

V. Extracting Opinion Targets and Opinion Words

Extracting opinion target and opinion words one of the important task of opinion mining. More attention has been given to focus on these tasks [52], [53]. Extraction can be classified into sentence level and corpus level extraction. Identifing opinion target/word in each sentence refers to sentence level extraction [54], [55]. Extractors such as CRFs and HMM are built using sequence labelling models. Huang [56] shows that opinion extraction can be done using lexicalized HMM model. These methods need labeled training data to train the model. Overall performance of extraction reduces if less amount of labeled training data or labelled data from different domains other than the current domain is used. Based on the transfer learning method Li *et al.* [57] proposed that Cross do-main sentiment extraction of opinion words/ targets. Performance of extraction depend more on the relevance between source and target domain.

Most of the earlier methods applied a unsupervised extraction process. Important component of this method is to detect opinion relations and finding opinion associations among the words. Hu *et al.* [58] show that nearest neighbour rule can also exploit opinion relations among words. To obtain the good accuracy of detecting opinion relations among the words, only considering nearest neighbour rule and co-occurrence information is not suffficient. Specific patterns are designed by Zhang *et al.* and these are used in [59] which considerably increased recall. They also used HITS algorithm to calculate opinion target confidence to increase precision. Word Alignment Model is one of the important algorithm to extract opinion/target. Liu *et al.* [60] implemented WAM based opinion/target extraction. Thy used unsupervised WAM to capture opinion relations in sentences. From opinion relations, random walk framework is used to extract opinion targets. To detect implicit topics and opinon words, topic modeling is employed [61], [62]. The purpose of these method is not to extract opinion target/word, instead clustering all words with respect to the aspect in reviews.

VI. Sentiment Analysis at Different Levels

Sentiment analysis approaches extract sentiment words from the text and find the orientation of words to classify them as positive, negative or neutral words. Initially, sentiment analysis focused on the semantic orientation of adjectives. The techniques of analysing the sentiment words are largely used in filtering text, discovering the public opinions, customer relationship [63]. Sentiment analysis can be done at different levels of granularity from document level to sentence level. Pang *et al.* [64] proposed three machine learning algorithms: support vector machines, maximum entropy classification, and Naive Baye's give best results compared to human created baselines [64]. Rule based and Learning based approaches are the different categories of the sentiment analysis approaches. This approach uses the handbuild lexicon. Bloom *et al.* [65] propose a method that extracts the sentiment orientation from lexicon and classify the sentence or document by analysis the patterns that occure in text. Wiebe *et al.* [66] provided a lexicon containing subje-ctive words such as verbs, nouns, adjectives with their polarity and strength associated with them.

The polarity of word can change depending on the context in a sentence. Number of methods are proposed to find the sentiment orientation of words by considering the context of a sentence [67]. Yuen *et al.* gave an approach that calculate the sentiment orientation of words on the basis of morphemes and its statistical association with strong polarised words. To measure semantic polarity of adjectives, wordnet can also be used [68]. Hu and Liu [69] proposed a method where linguistic patterns called sequential rules are used to extract opinion features from reviews which can be mined from labeled data which is used for training sequences of words. Kim *et al.* [70] proposed a method

Global Journal of Computer Science and Technology (C) Volume XVI Issue V Version I Year 2016

for identifying an opinion with its holder and topic, given a sentence in on-line news media texts is extracted.

Millions of people daily post their comments on variety of topics with the help of social media. It is very difficult to analyse this information as it is huge and generally it is multidimentional and time varing. Wang *et al.* [71] proposed a visualization system that analysis the sentiments that are expressed in the public comments and give the short term trend of the sentiments. Relationship map is used to visualize the changes in the attributes and evolution model is used to compare the time varing parameters. In general, most of the machine learning algorithms learn single task at a time. Li*et al.* [72] investigated on Collaborative Multitasking Learning algorithm. The aim of the work is to focus on improving the performance of all tasks insted of single primary task. Online data is used for learning so that it can be processed as and when it arrives. This makes method more realistic. Collaborative Online Multitasking Learning algorithm results in improved classification performance. Relation between the tasks is assumed to be uniform and considering the relatedness degree among the tasks still improve the performance.

Social media is a great place for students to share their experiences, ideas, emotions, stress and to seek social support. To understand and reect the students experiences in the social media, human intervention is required. As the data available is huge, we need a automated classifier. To address this problem, Chen *et al.* [73] proposed a platform where Students Learning Experience is analysed by integrating large scale data mining techniques and Qualitative analysis. All students may not be active in social media, resulting in only few students who are ready to share their thoughts post their ideas. This work only focus on the text content, where as images and videos also add lot of information. Various research work is done on extracting sentiments from the comments the user has posted. Tan *et al.* [74] worked on finding the sentiment variations in the twitter that give insite about the reason behind the cause of sentiment variations. Latent Dirichlet Allocation model is used to analyse the possible reason for the sentiment variations.

VII. Polarity Inconsistent Dictionaries

Sentiment dictionaries are used to find the polarity of opinion words in the reviews. Orientation of opinion words in the reviews can be found using sentiment dictionaries. Final orientation of a sentence or a document is the addition of orientation of each word. There are different types of sentiment dictionaries. Domain independent sentiment dictionaries are created manually or semi automatically used by all domain reviews. Major problems with sentiment dictionaries is the inconsistency in intra and inter dictionary. Fragut *et al.* [75] show that inconsistency problem is NP complete. Inconsistency in dictionaries can be detected using fast SAT solver. There are corpora and Wordnet-based sentiment polarity lexicon used. To derive sentiment lexicon, Wordnet based approach uses lexicon relations defined in wordnet. Measuring the rela tive distance of a word from examples determine the sentiment of adjectives in Wordnet [76]. Synonyms and antonyms are used to increase the sets of words. One more method to increase the set of words by adding all synomyms of a polar word with polarity and antonyms with reverse polarity [77].When seed polar words are very few such as low resource language, method suffer from low recall [78]. In QW [79] synsets in word net are automatically anotated. If two synsets are assigned opposing polarites, then they are discarded. Machine learning algorithms as well as stochastic algorithms [80] can be used to classify words into different polarities.

VIII. Polarity Shifting Detection

Sentiment classifiers are intended to classify the document into different catagories. Bag of words model is used to represent the text which need to be classified. In BOW model, the text is represented in the form of vector of words. As BOW changes word order and remove some syntatic information, it is not an efficient model for sentiment classification. To remove this advantage, linguistic knowledge is introduced to enhance the efficiency of BOW. However, accuracy improvement is very less due to the basic aws in BOW. Polarity shift problem is the most important dificuly in the BOW model. Features are also used to determine whether the phrase is positive or negative contextual polarity and overall aim is to use the phrase-level sentiment analysis. Several approaches are proposed to address the polarity shift problem [81].

Polarity shift problem also has a problem of extra annotations and linguistic knowledge and some efforts are done on solving this problem [82], [83]. Nakagawa *et al.* [84] proposes a dependency tree-based method for Japanese and English sentiment classification using conditional random variables. The polarity of each dependency subtree of a subjective sentence is represented by a hidden variable. Sentiment classification is done by calculating the values of the hidden variables that are calculated in consideration of interaction between the variables. Liu *et al.* [85] proposes linguistic rules to deal with the problem together with a new option aggregation function and classifies the review or opinion whether it is a positive or negative. Ding *et al.* [86] came up with holistic lexicon based approach to determine the semantic orientation of the reviews obtained by opinion mining and uses a new function for aggregating multiple opinion words in the same sentence. Ding *et al.* [87] deals with the assigning of entities to the opinion extracted using a pattern based method. It also finds the entities of the comparative sentences whose entities are not explicitly mentioned by extracting large opinions using state of the art technique. Several techniques for opinion mining features based on data mining and Natural language processing (NLP) methods on product reviews. It gives a feature-based summary of a large number of product reviews by customer.

Turney *et al.* [88] proposed a concept based on simple unsupervised learning algorithm for rating a

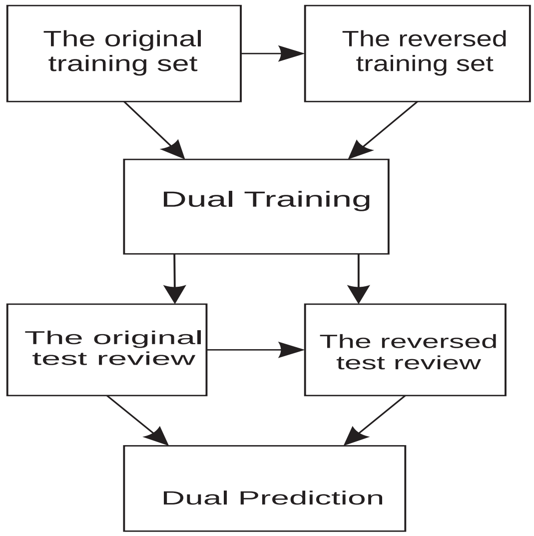

Figure 2: Dual sentiment analysis

review as recommended or un-recommended. The algorithm extract phrases which has adjectives or adverbs and estimates the semantic orientation of each phrase and classifies the reviews based on average semantic orientation. Turney *et al.* [89] provides general technique to measure semantic orientation to semantic association. It evaluates the semantic orientation using Pointwise mutual information and Latentsentiment Analysis(LSA) methods. Determining the polarity of sentiment-bearing expression, by considering the effect of interaction among words or constituents is important. It provides novel training-based approach which incorporates the structural inference to the learning procedure by the compositional semantics [90].

a) Data Expansion Technique

Expanding the data has been seen in the handwritten recognition, where the performance of this method is improved by adding few more training data. Figure 2 gives the process for dual sentiment analysis. In text mining, Agirre *et al.* [91] proposed a method to expand the amount of labeled data unique expressions in definitions from wordnet for a task of word sense disambiguation. Fujita *et al.* [92] proposed a method which provides training data using sentences from the external dictionary. Xia *et al.* [93] proposed a novel method of data expansion. The original and reversed reviews are constructed in one to one correspondence. The data expansion happens both in training stage and also during testing stage.

IX. Intrinsic and Extrinsic Domain Relevance

Opinion feature indicate attribute of an entity or an entity on which user express their opinions. Many approaches are proposed to extract to classify movie review opinion features. One of the efficient method is supervised learning method. This method works well in a given domain and if it needs to work for other domain, it has to be retrained [94].

By defining domain independent syntactic rules, Unsupervised approaches [95] identify opinion features. Wiebe *et al.* [96] proposed a supervised classification method to predict sentence subjectivity [97]. Pang *et al.* [98] proposed three machine learning algorithms to classify movie reviews into different sentiments. They are Naive Baye's, Support vector machines and maximum entropy [99].

A document can contain both Subjective and objective sentences. Due to this, Sentiment classifier may consider irrelevent text. Pang *et al.*, [100] proposed sentiment level subjectivity detector which identifies subjective or objective sentences. Then objective sentences are discarded which improves the classification results. Subjective sentences are further classified into positive and negative [101].

Wiebe *et al.* [102] proposed a method which uses naive Bayesian classifier to classify subjective sentences. One of the restriction for this method is the shortage of training set. Riloff *et al.* [103] proposed bootstrapping method which automatically label the training data so that lack of training data problem is solved.

X. Information Filtering

Online social networks has become a popular interactive medium to communicate between the users.Every day there is exchange of huge amountof information between the users. Information may be text, audio and video data. But the disadvantage is user wall is posted with so many different varieties of information in which the user may not be interested in some perticular type of data. This leads to the requirement of filtering the messages on the user wall before posting it [104], [105]. User is given the authority to decide which content type of messages need to be blocked. Information filtering of textual documents is of a great concern in recent years [106]. Vanetti *et al.* [107] proposed a Filtered Wall(FW), an automated system which is able to filter unwanted messages from online social network users. To mechanically assign with every text messages a set of categaries based on content, Machine learning text categorization techniques [107] are used.

a) Content-Based Filtering

Information filtering system are used to classify continuously generated messages sent by information produces and post messages on to the user wall that may satisfy the user requirements. Content based filtering system selects messages based on the interrelationship between the contents of the messages and the user preferences. Content-based filtering system mainly use the machine learning algorithms. Here classifier is trained by learning from the labeled examples. Text is mapped into a condensed representation of its content and then applied to training by feature extraction procedure. Hirsh *et al.*, [108] improved the short text messages using semisupervised learning strategy. It is based on the combination of labeled training data and secondary corpus of unlabeled data. Another approach proposed by Bobicev *et al.*, [109] is to adapt a statistical learning method that performs well.

b) Policy based Personalization

Classification mechanisms for personalizing access in OSNs is of recent interest. In [110], focus is on twitter and each tweet is associated with set of categories depending on its content. User selects tweets depending on the content that they are interested in. Contradicting to this, Golbeck *et al.*, [111] proposed *filmTrust*, that gives OSN trust relationship and this does not provide filtering policy layer by layer. Hence, user cannot exploit the classification results.

c) Text Representation

To increase the performance of classifier, taking out suitable set of features which present the text of a document is necessary. There are divergent sets of features for text classification. BOW, Document properties (Dp) and Contextual features(CF) [112], [113] are considered for short text messages. BOW and DP are used in [112] and they are completely derived from the information present with in the text of the message. Contextual features play an important role in finding the semantics of the messages.

XI. Evaluation

The performance of variety of methods that are used in sentiment analysis is compared by measuring few parameters like precision, recall and Fscore. Precision is a part of retrieved data that are more applicable. Whereas recall is the part of relevant data that are retrieved. F-measure is calculated using both recall and precision. As given in the Table 2, we have compared various works with respect to classifiers used, feature extraction methods and different measurement metrics. Matrics considered are Accuracy(A), Precision(P), Recall(R) and F-score(F).

XII. Conclusions

Variety of applications of sentiment analysis are widely used. They include classifing reviews, sum marizing review etc. In this paper, we have discussed different approaches of sentiment classification and its performance. Domain adaptation is required as it reduces number of classifier user for the sentiment analysis. Different approaches of domain adaptation are compared using supervised, semisupervised and unsupervised learning methods. Heterogeneous domain adaptation is able to classify data with different dimensions. Extracting opinion words and target words is crucial for the performance of the classifier. Efficient algorithms for extracting opinions words and opinion target are discussed. Data expansion techniques are discussed which is used in dual sentiment analysis that reduces the number of training labeled data used for the classification. Information filtering is a online social network user friendly concept which gives user exible choices.

References Références Referencias

1. G. U. Vasanthakumar, P. D. Shenoy, and K. R. Venugopal, PTIB: Profiling Top Inuential Blogger in Online Social Networks," vol. 10, no. 1, pp. 77-91, 2016.
2. N. Godbole, M. Srinivasaiah, and S. Skiena, Large-Scale Sentiment Analysis for News and Blogs." *ICWSM*, vol. 7, no. 21, pp. 219 -222, 2007.
3. S. Li, C.-R. Huang, G. Zhou, and S. Y. M. Lee, Employing Personal / impersonal Views in Supervised and Semi-supervised Sentiment Classification," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 414-423, 2010.
4. J. Jiang and C. Zhai, Instance Weighting for Domain Adaptation in NLP," in *Proceedings of the 45th*

Annual meeting of the Association for Computational Linguistics, vol. 7, pp. 264-271, 2007.

5. E. Haddi, X. Liu, and Y. Shi, The Role of Text Pre-processing in Sentiment Analysis," *Proceedings of the First International Conference on Information Technology and Quantitative Management, Procedia Computer Science*, vol. 17, pp. 26 -32, 2013.
6. Y. Bao, C. Quan, L. Wang, and F. Ren, The Role of Pre-processing in Twitter Sentiment Analysis," in *Proceedings of the International Conference on Intelligent Computing Methodologies*. Springer, pp. 615 - 624, 2014.
7. K. R. Venugopal, K. G. Srinivasa, and L. M. Patnaik, *Soft Computing for Data Mining Applications*. Springer, 2009.
8. D. Sejal, K. G. Shailesh, V. Tejaswi, D. Anvekar, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, “Query Click and Text Similarity Graph for Query Suggestions," in *Proceedings of the International Workshop on Machine Learning and Data Mining in Pattern Recognition*. Springer, pp. 328{341, 2015.
9. B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up?: Sentiment Classification using Machine Learning Techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Languag Processing, ACL*, vol. 10, pp. 79-86, 2002.
10. D. Davidov, O. Tsur, and A. Report, Semi-supervised Recognition of Sarcastic Sentences in Twitter and Amazon," in *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics, pp. 107-116, 2010.
11. X. Glorot, A. Bordes, and Y. Bengio, Domain Adaptation for Large-scale Sentiment Classification: A Deep Learning Approach," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 513-520, 2011.
12. F. Li, S. Wang, S. Liu, and M. Zhang, Suit: A Supervised User-item Based Topic Model for Sentiment Analysis," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1636-1642, 2014.
13. D. Bollegala, T. Mu, and J. Goulermas, Cross-domain Sentiment Classification using Sentiment Sensitive Embeddings," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 398 - 410, 2016.
14. S. Liu, X. Cheng, F. Li, and F. Li, TASC: Topic-Adaptive Sentiment Classification on Dynamic Tweets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1696 -1709, 2015.

Table 2: Comparison of the Various Sentiment Analysis Techniques along with Performance

Author	Concept	Feature selection method	Data source	A	P	R	F
Shenghua Liu *et al.*, (2015), [14]	Multiclass SVM Classifier	PMI-IR	Sander-Twitter Corpus	0.54	0.55	0.52	0.53
Xiao *et al.*, (2015), [16]	Hilbert Schmidt Independence Criteria	Prediction function	Amazon product reviews	0.79	-	-	-
Rui Xia *et al.*, (2015), [93]	Naive bayes classifier	Data Expansion Technique	Amazon product reviews	0.81	-	-	-
Wen li *et al.*, (2014), [114]	SVM Classifier	Heterogeneous feature augmentation	Object dataset	0.54	-	-	-
Jianping cao *et al.*, (2014), [115]	Rule based classifier	Sentiment polarity score	tianya.cn	0.82	0.84	0.30	-
Li Cheng *et al.*, (2014), [17]	SVM classifer	Semi-supervised transfer component analysis	Amazon product reviews	0.63	-	-	-
Zhen Hai *et al.*, (2014), [116]	Differentiating opinion feature statistics across domains	Intrinsic and Extrinsic domain relevance	Cell phone review corpus	-	0.65	0.61	0.63

15. M.-F. M. Quynh Thi Ngoc Do, Steven Bethard, Domain Adaptation in Semantic Role Labeling Using a Neural Language Model and Linguistic Resources," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1812-1823, 2015.
16. M. Xiao and Y. Guo, Feature Space Independent Semi-supervised Domain Adaptation *via* Kernel Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 54 -66, 2015.
17. L. Cheng and S. J. Pan, Semi-Supervised Domain Adaptation on Manifolds," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 12, pp. 2240 - 2249, 2014.

18. H. Daum'e III, Frustratingly Easy Domain Adaptation," *arXiv preprint arXiv: 0907.1815*, 2009.
19. S. J. Pan, J. T. Kwok, and Q. Yang, Transfer Learning *via* Dimensionality Reduction." in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, vol. 8, pp. 677 -682, 2008.
20. R. K. Ando and T. Zhang, A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data," *The Journal of Machine Learning Research*, vol. 6, no. 2, pp. 1817-1853, 2005.
21. J. Blitzer, R. McDonald, and F. Pereira, Domain Adaptation with Structural Correspondence Learning," in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp.120-128, 2006.
22. J. Blitzer, M. Dredze, F. Pereira *et al.*, Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification," in *Proceedings of the 45th Annual Meeting of the Association of the Computational Linguistics, ACL*, vol. 7, pp. 440 - 447, 2007.
23. S. J. Pan, X. Ni, J.-T. Sun, Q. Yang, and Z. Chen, Cross-domain Sentiment Classification *via* Spectral Feature Alignment," in *Proceedings of the 19th International Conference on World wide web*. ACM, pp. 751-760, 2010.
24. G. Vasanthakumar, P. D. Shenoy, and K. R. Venugopal, PFU: Profiling Forum Users in Oline Social Networks, a Knowledge Driven Data Mining Approach," in *Proceedings of the IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, pp. 57- 60, 2015.
25. D. Bollegala, D. Weir, and J. Carroll, Using Multiple Sources to Construct a Sentiment Sensitive The saurus for Cross-domain Sentiment Classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 132-141, 2011.
26. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, A Theory of Learning from Different Domains," *Machine Learning*, vol. 79, no. 1, pp. 151-175, 2010.
27. X. Shi, Q. Liu, W. Fan, P. S. Yu, and R. Zhu, Transfer Learning on Feature Spaces *via* Spectral Transformation," in *Proceedings of the IEEE 10th International Conference on Data Mining (ICDM)*. IEEE, pp. 1049 -1054, 2010.
28. M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, Direct Importance Estimation with Model Selection and its Application to Covariate Shift Adaptation," in *Proceedings of the Advances in Neural Information Processing Systems*, pp1433-1440, 2008.
29. H. Shimodaira, Improving Predictive Inference unde Covariate Shift by Weighting the Log-likelihood Function," *Journal of Statistical Planning and Inferencevol.* 90, no. 2, pp. 227-244, 2000.
30. M. Chen, K. Q. Weinberger, and J. Blitzer, Cotraining for Domain Adaptation," in *Proceedings othe Advances in Neural Information Processing Sytems*, pp. 2456-2464, 2011
31. G. Tur, Co-adaptation: Adaptive Co-training for Semi-supervised Learning," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009.* IEEE, pp. 3721-3724, 2009.
32. Kumar, A. Saha, and H. Daume, Coregularization based Semi-supervised Domain Adaptation," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 478-486, 2010.
33. J. Blitzer, S. Kakade, and D. P. Foster, Domai Adaptation with Coupled Subspaces," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 173 -181, 2011.
34. K. Saenko, B. Kulis, M. Fritz, and T. Darrell, Adapting Visual Category Models to New Domains," in *Proceedings of the European Conference on Compute Vision -ECCV*. Springer, pp. 213{226, 2010.
35. B. Kulis, K. Saenko, and T. Darrell, What You Saw is not What You Get: Domain Adaptation using Asymmetric Kernel Transforms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1785-1792, 2011.
36. C. Wang and S. Mahadevan, Heterogeneous Domain Adaptation using Manifold Alignment," in *Procee dings of the International Joint Conference on Artifi cial Intelligence, IJCAI*, vol. 22, no. 1, pp. 1541, 2011.
37. W. Dai, Y. Chen, G.-R. Xue, Q. Yang, and Y. Yu, Translated Learning: Transfer Learning across Different Feature Spaces," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 353 - 360, 2008.
38. Y. Choi and C. Cardie, Adapting a Polarity Lexicon using Integer Linear Programming for Domain Specific Sentiment Classification," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing, ACL*, vol. 2, pp. 590 - 598, 2009.
39. V. Stoyanov and C. Cardie, Topic Identification for Fine-grained Opinion Analysis," in *Proceedings of the 22nd International Conference on Computational Linguistics, Association for Computational Linguistics*, vol. 1, pp. 817- 824, 2008.
40. Y. He, C. Lin, and H. Alani, Automatically Extracting Polarity-bearing Topics for Cross-domain Sentiment Classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 123-131, 2011.
41. S. Gao and H. Li, A Cross-domain Aadaptation Method for Sentiment Classification using Probabilistic Latent Analysis," in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*. ACM, pp. 1047{1052, 2011.

42. V. Stoyanov and C. Cardie, Topic Identification for Fine-grained Opinion Analysis," in *Proceedings of the 22nd International Conference on Computational Linguistics, Association for Computational Linguistics*, vol. 1, pp. 817-824, 2008.
43. D. Das and S. Bandyopadhyay, Emotion Coreferencing-Emotional Expression, Holder, and Topic," *Computational Linguistics and Chinese Language Processing*, vol. 18, no. 1, pp. 79-98, 2013.
44. F. Li, S. Wang, S. Liu, and M. Zhang, Suit: A Supervised User-item based Topic Model for Sentiment Analysis," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 1636-1642, 2014.
45. Y. Mejova and P. Srinivasan, Crossing Media Streams with Sentiment: Domain Adaptation in Blogs, Reviews and Twitter." in *Sixth International AAAI Conference on Weblog and Social Media*, pp. 234 - 241, 2012.
46. D. Das and S. Bandyopadhyay, Extracting Emotion Topics from Blog Sentences: Use of Voting from Multiengine Supervised Classifiers," in *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*. ACM, pp. 119-126, 2010.
47. S. Liu, F. Li, F. Li, X. Cheng, and H. Shen, Adaptive Cotraining SVM for Sentiment Classification on Tweets," in *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. ACM, pp. 2079 - 2088, 2013.
48. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, Sentiment Analysis of Twitter Data," in *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, pp. 30 - 38, 2011.
49. S. Liu, W. Zhu, N. Xu, F. Li, X.-q. Cheng, Y. Liu, and Y. Wang, Co-training and Visualizing Sentiment Evolvement for Tweet Events," in *Proceedings of the 22nd International Conference on World Wide Web companion*. International World Wide Web Conferences Steering Committee, pp. 105-106, 2013.
50. A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, vol. 10, pp. 178 -185, 2010.
51. D. Das and S. Bandyopadhyay, Identifying Emotion Topic An Unsupervised Hybrid Approach with Rhetorical Structure and Heuristic Classifier," in *Proceedings of the International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*. IEEE, pp. 1-8, 2010.
52. Y. Kim and S. R. Jeong, Opinion-Mining Methodology for Social Media Analytics." *TIIS*, vol. 9, no. 1, pp. 391-406, 2015.
53. H. Yu, Structure-aware Review Mining and Summarization," in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 653 - 661, 2010.
54. T. Ma and X. Wan, Opinion Target Extraction in Chinese News Comments," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, pp. 782-790, 2010.
55. Q. Zhang, Y. Wu, T. Li, M. Ogihara, J. Johnson, and X. Huang, Mining Product Reviews Based on Shallow Dependency Parsing," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 726-727, 2009.
56. W. Jin, H. H. Ho, and R. K. Srihari, A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining," in *Proceedings of the 26th Annual International Conference on Machine Learning*. Citeseer, pp. 465 -472, 2009.
57. F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, Cross-domain Co-extraction of Sentiment and Topic Lexicons," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers*, vol. 1, pp. 410-419, 2012.
58. M. Hu and B. Liu, Mining Opinion Features in Customer Reviews," in *AAAI*, vol. 4, no. 4, pp. 755-760, 2004.
59. G. Qiu, B. Liu, J. Bu, and C. Chen, Opinion Word Expansion and Target Extraction Through Double Propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9-27, 2011.
60. K. Liu, L. Xu, and J. Zhao, Opinion Target Extraction using Word-based Translation Model," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1346 -1356, 2012.
61. W. X. Zhao, J. Jiang, H. Yan, and X. Li, Jointly Modeling Aspects and Opinions with a Maxent-LDA hybrid," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 56-65, 2010.
62. Mukherjee and B. Liu, Modeling Review Comments," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers*, vol. 1, pp. 320-329, 2012.
63. V. Jha, N. Manjunath, P. D. Shenoy, K. Venugopal, and L. Patnaik, HOMS: Hindi Opinion Mining System," in *IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)*. IEEE, pp. 366 - 371, 2015.
64. B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up?: Sentiment Classification using Machine Learning Techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, ACL*, vol. 10, pp. 79-86, 2002.
65. K. Bloom, N. Garg, and S. Argamon, Extracting Appraisal Expressions," in *HLT-NAACL*, pp. 308-315, 2007.

66. J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, Learning Subjective Language," *Computational linguistics*, vol. 30, no. 3, pp. 277-308, 2004.
67. L. Vibha, G. Harshavardhan, K. Pranaw, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, Classification of Mammograms using Decision Trees," in *Proceedings of the 10th International Database Engineering and Applications Symposium (IDEAS'06)*. IEEE, pp. 263 -266, 2006.
68. G. Vasanthakumar, A. K. Upadhyay, P. F. Kalmath, S. Dinakar, P. D. Shenoy, and K. Venugopal, UP3: User Profiling from Profile Picture in Multi-Social Networking," in *Proceedings of the Annual IEEE India Conference (INDICON), IEEE*, pp. 1-6, 2015.
69. M. Hu and B. Liu, Opinion Feature Extraction Using Class Sequential Rules." in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, pp. 61- 66, 2006.
70. S.-M. Kim and E. Hovy, Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text," in *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. ACL, pp. 1-8, 2006.
71. C. Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang, SentiView: Sentiment Analysis and Visualization for Internet Popular Topics," *IEEE transactions on human-machine systems*, vol. 43, no. 6, pp. 620-630, 2013.
72. G. Li, S. C. Hoi, K. Chang, W. Liu, and R. Jain, Collaborative Online Multitask Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1866-1876, 2014.
73. X. Chen, M. Vorvoreanu, and K. Madhavan, Mining Social Media Data for Understanding Students' Learning Experiences," *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 246-259, 2014.
74. S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. Bu, C. Chen, and X. He, Interpreting the Public Sentiment Variations on Twitter," *IEEE transactions on knowledge and data engineering*, vol. 26, no. 5, pp. 1158 - 1170, 2014.
75. E. Dragut, H. Wang, C. Yu, P. Sistla, and W. Meng, Polarity Consistency Checking for Sentiment Dictionaries," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 997-1005, 2012.
76. G. Vasanthakumar, B. Prajakta, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, PIB: Profiling Inuential Blogger in Online Social Networks, A Knowledge Driven Data Mining Approach," *Procedia Computer Science*, vol. 54, pp. 362-370, 2015.
77. S.-M. Kim and E. Hovy, Identifying and analyzing judgment opinions," in *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. ACL, pp. 200-207, 2006.
78. D. Rao and D. Ravichandran, Semi-supervised Polarity Lexicon Induction," in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 675-682, 2009.
79. V. Jha, R. Savitha, S. Hebbar, P. D. Shenoy, and K. R. Venugopal, HMDSAD: Hindi Multi-domain Sentiment Aware Dictionary," in *Proceedings of the International Conference on Computing and Network Communications (CoCoNet), IEEE*, pp. 241-247, 2015.
80. A. Hassan and D. Radev, Identifying Text Polarity using Random Walks," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. ACL, pp. 395-403, 2010.
81. Kennedy and D. Inkpen, Sentiment Classification of Movie Reviews using Contextual Valence Shifters," *Computational Intelligence*, vol. 22, no. 2, pp. 110-125, 2006.
82. S. Li, S. Y. M. Lee, Y. Chen, C.-R. Huang, and G. Zhou, Sentiment Classification and Polarity Shifting," in *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, pp. 635 - 643, 2010.
83. T. Wilson, J. Wiebe, and P. Hoffmann, Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis," *Computational Linguistics*, vol. 35, no. 3, pp. 399-433, 2009.
84. T. Nakagawa, K. Inui, and S. Kurohashi, Dependency Tree-based Sentiment Classification using CRF s with Hidden Variables," in *Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786 - 794, 2010.
85. X. Ding and B. Liu, The Utility of Linguistic Rules in Opinion Mining," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 811- 812, 2007.
86. X. Ding, B. Liu, and P. S. Yu, A Holistic Lexicon-based Approach to Opinion Mining," in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, pp. 231-240, 2008.
87. X. Ding, B. Liu, and L. Zhang, Entity Discovery and Assignment for Opinion Mining Applications," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 1125-1134, 2009.
88. P. D. Turney, Thumbs Up or Thumbs Down? : Semantic Orientation Applied to Unsupervised Classification of Reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 417-424, 2002.
89. P. D. Turney and M. L. Littman, Measuring Praise and Criticism: Inference of Semantic Orientation

Global Journal of Computer Science and Technology (C) Volume XVI Issue V Version I Year 2016

from Association," *ACM Transactions on Information Systems (TOIS)*, vol. 21, no. 4, pp. 315-346, 2003.

90. Y. Choi and C. Cardie, Learning with Compositional Semantics as Structural Inference for Sub sentential Sentiment Analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 793-801, 2008.
91. E. Agirre and D. Martinez, Exploring Automatic Word Sense Disambiguation with Decision Lists and the Web," in *Proceedings of the COLING-2000 Work shop on Semantic Annotation and Intelligent Content*. Association for Computational Linguistics, pp. 11 -19, 2000.
92. S. Fujita and A. Fujino, Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-supervised Learning Method," *ACM Transactions on Asian Language Information Processing (TALIP)*, vol. 12, no. 2, p. 7, 2013.
93. R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi, and T. Li, Dual Sentiment Analysis: Considering Two Sides of One Review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2120-2133, 2015.
94. N. Jakob and I. Gurevych, Extracting Opinion Targets in a Single-and Cross-domain Setting with Conditional Random Fields," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 1035-1045, 2010.
95. G. Qiu, B. Liu, J. Bu, and C. Chen, Opinion Word Expansion and Target Extraction through Double Propagation," *Computational linguistics*, vol. 37, no. 1, pp. 9-27, 2011.
96. V. Hatzivassiloglou and J. M. Wiebe, Effects of Adjective Orientation and Gradability on Sentence Subjectivity," in *Proceedings of the 18th Conference on Computational linguistics, ACL*, vol. 1, pp. 299-305, 2000.
97. V. Jha, N. Manjunath, P. D. Shenoy, and K. R. Venugopal, HSAS: Hindi Subjectivity Analysis System," in *Proceedings of the 2015 Annual IEEE India Conference (INDICON)*. IEEE, pp. 1-6, 2015.
98. B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up?: Sentiment Classification using Machine Learn- ing Techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, ACL*, vol. 10, pp. 79-86, 2002.
99. L. Vibha, G. Harshavardhan, K. Pranaw, P. D. Shenoy, K. R. Venugopal, and L. M. Patnaik, Lesion Detection using Segmentation and Classification of Mammograms," in *Proceedings of the 25th ASTED International Multi-Conference: Artificial Intelligence and Applications*. ACTA Press, pp. 311-316, 2007.
100. B. Pang and L. Lee, A Sentimental Education: Sentiment Analysis using Subjectivity Summarization based on Minimum Cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pp. 271, 2004.
101. B. Liu, Sentiment Analysis and Subjectivity." *and-book of Natural Language Processing*, vol. 2, pp. 627- 666, 2010.
102. J. M. Wiebe, R. F. Bruce, and T. P. O'Hara, Development and Use of a Gold-standard Data Set for Subjectivity Classifications," in *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 246- 253, 1999.
103. E. Riloff and J. Wiebe, Learning Extraction Pat- terns for Subjective Expressions," in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pp. 105 -112, 2003.
104. K. Srinivasa, A. Singh, A. Thomas, K. R. Venugopal, and L. Patnaik, Generic Feature Extraction for Classification using Fuzzy C-means Clustering," in *Proceedings of the 3rd International Conference on Intelligent Sensing and Information Processing*. IEEE, pp. 33-38, 2005.
105. D. Sejal, V. Rashmi, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, Image Recommendation based on Keyword Relevance using Absorbing Markov Chain and Image Features," *International Journal of Multimedia Information Retrieval*, vol. 5, no. 3, pp. 185 - 199, 2016.
106. G. Adomavicius and A. Tuzhilin, Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engi- neering*, vol. 17, no. 6, pp. 734 -749, 2005.
107. M. Vanetti, E. Binaghi, E. Ferrari, B. Carminati, and M. Carullo, A System to Filter Unwanted Messages from OSN User Walls," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 2, pp. 285 - 297, 2013.
108. S. Zelikovitz and H. Hirsh, Improving Short Text Classification using Unlabeled Background Knowledge to Assess Document Similarity," in *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 1183-1190, 2000.
109. V. Bobicev and M. Sokolova, An Effective and Robust Method for Short Text Classification." in *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*, pp. 1444 -1445, 2008.
110. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, Short Text Classification in Twitter to Improve Information Filtering," in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 841-842, 2010.
111. J. Golbeck, Combining Provenance with Trust in Social Networks for Semantic web Content Filtering," in *Provenance and Annotation of Data*. Springer, pp. 101-108, 2006.

112. M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, Content-based Filtering in On-line Social Networks," in *Privacy and Security Issues in Data Mining and Machine Learning*. Springer, pp. 127-140, 2010.
113. M. Carullo, E. Binaghi, and I. Gallo, An Online Document Clustering Technique for Short Web Contents," *Pattern Recognition Letters*, vol. 30, no. 10, pp. 870 - 876, 2009.
114. L. Duan, D. Xu, and I. Tsang, Learning with Augmented Features for Heterogeneous Domain Adaptation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1134-1148, 2014.
115. J. Cao, K. Zeng, H. Wang, J. Cheng, F. Qiao, D. Wen, and Y. Gao, Web-Based Traffic Sentiment Analysis: Methods and Applications," *IEEE transactions on Intelligent Transportation systems*, vol. 15, no. 2, pp. 844-853, 2014.
116. Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, Identifying Features in Opinion Mining *via* Intrinsic and Extrinsic Domain Relevance," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 623-634, 2014.