A roadmap of sentiment analysis and its research directions

Article *in* International Journal of Knowledge and Learning · January 2015

DOI: 10.1504/IJKL.2015.073485

Rajni Mohana
Jaypee University of Information Technology

A roadmap of sentiment analysis and its research directions

Sukhnandan Kaur and Rajni Mohana*

Jaypee University of Information Technology,
Waknaghat, P.O. Waknaghat,
Teh Kandaghat, Distt. Solan, 173 234, (H.P.), India
Email: sukhnandan.kaur@mail.juit.ac.in
Email: rajni.mohana@juit.ac.in
*Corresponding author

Abstract: The exponential growth of data on the websphere accelerated the need of extracting meaningful information from it. This information can be used for better decision making. The automatic generation of sentiments from the text is called as sentiment analysis (SA). It is a collaborative process of natural language processing and data mining. This paper tries to deeply analyse the existing research work in the area of SA. It presents work done till date and segregates it in terms of level of granularity. An ideal sentiment analyser should have the intellectual capability similar to a human being. This paper mentions a roadmap of the research directions to achieve the goal of ideal sentiment analyser. These research directions include SA based on temporal summarisation, multi-linguality, etc. This paper also mentions various research aspects to work on these research directions. Future research in this direction will also refine the performance of decision making in decision support system.

Keywords: sentiment analysis; opinion mining; knowledge base; multi-linguality; fake review detection; temporal tagging.

Reference to this paper should be made as follows: Kaur, S. and Mohana, R. (2015) 'A roadmap of sentiment analysis and its research directions', *Int. J. Knowledge and Learning*, Vol. 10, No. 3, pp.296–323.

Biographical notes: Sukhnandan Kaur is currently a Research Scholar in the Department of CSE, at the Jaypee University of Information Technology (JUIT), Waknaghat, India. She has done her MTech from SVIET, Banur. Her area of interest is natural language processing which includes transliteration of Gurmukhi language to English and sentiment analysis of textual data.

Rajni Mohana is currently working as an Assistant Professor (Senior Grade) in the Department of CSE, at the Jaypee University of Information Technology (JUIT), Waknaghat, India. She has over 10 years of experience in Academia. She has various national and international publications to her credit. She is also a reviewer for various renowned international journals and international conferences.

1 Introduction

The notion of social network and methods presents us with a more scientific grasp of social data (the data gathered through social sharing). The data are in the form of blogs, comments, etc., where anyone can share their views about any named entity (product, person or place).

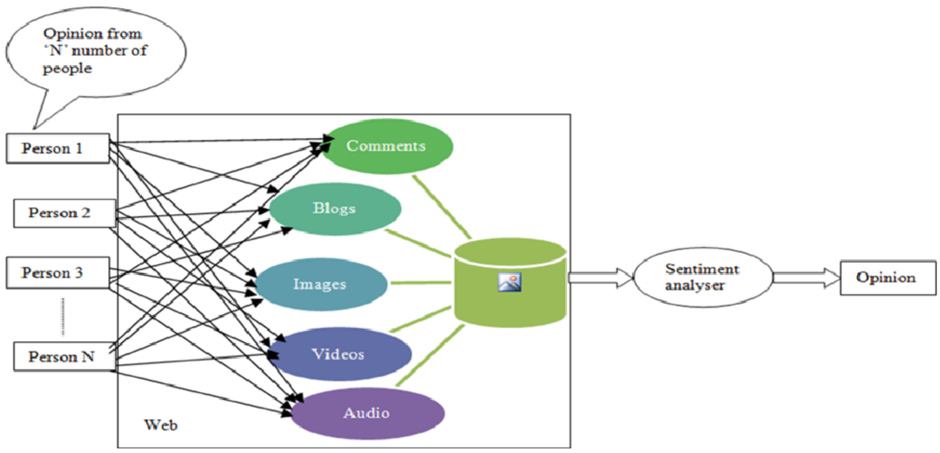
It is difficult to access a greater volume of social data, including data present in the literature. This huge amount of social data opens broader opportunities to streamline marketing efforts and measure campaign's impact on the efficient decision support system.

Ten years back, social sites-like Skype, Facebook, YouTube, Reddit, Twitter, Tumblr, Dropbox and Instagram did not exist. Nowadays social sites are flooded for sharing views. A summary of huge data for representing the usage of various social sites is shown in Table 1. Rank column tells the order of usage of various sites. It can be seen from Table 1 (http://onesecond.designly.com/) that Facebook and Twitter are the major information sharing media. Therefore, extracting the information manually from such a huge data available on the websphere as shown in Figure 1 is cost ineffective process. It also hinders the process of taking the right decision.

Table 1 Average information shared per second

S. No.	*Website*	*No. of posts per second*	*Ranking*
1	Skype	387,096	3
2	Facebook	39,731,495	1
3	Twitter	16,92,050	2
4	Instagram	162,050	4
5	Reddit	62,055	5

Figure 1 View of sentiment analysis (see online version for colours)



The exponential growth of data on the websphere accelerated the need of extracting meaningful information from such data. The discretionary evaluation of people's experience is done on resource scarce languages as well on the resource rich languages.

Its domain varies from a tiny personal thing to huge worldwide. The need of automatic generation of summaries of people's view about any product from available data upswing the development process in the field of sentiment generation.

The intersection between social media and user generated content arose a great deal of research in the area of sentiment analysis (SA). SA is present in many spheres of our daily lives, whether we realise or not. It affects how we shop, work, etc.

1.1 Motivation

Numerous researchers are working in the area of SA. There was a need to deeply analyse existing research work. This paper contributes in presenting a roadmap of the ongoing research directions and its overview is also presented in tabular format. It also tries to locate various research aspects which are still untouched or needs attention. There are many available survey in SA but we tried to compile the research done till date. We also identified various research directions-like unstructured sentences, fake review detection, temporal tagging and multi-lingualism.

The remainder of the paper is structured as follows: Section 2 presents the background knowledge of SA. Section 3 describes primary steps of constructing a sentiment analyser. Various performance metrics and their descriptions are presented in Section 4. Research gap in states-of-the-art SA system is shown in Section 5. Section 6 concludes with closing remarks.

2 Background

SA is a collaborative process of natural language processing and data mining. The work in SA is a subset of text engineering or deep learning. Deep learning is a way for the machine to interpret what they perceive. This can also be defined as to process various sentiment signals to support automatic decision generation.

Earlier diffusion of sentiment signals into binary form was the prime task of SA, i.e., positive and negative. Refinement in the level of granularity was proliferated along with the technical hikes in the machine learning. During the time span of early 2000s, researchers were working on the polarity check on the document, i.e., positive, negative or neutral signals. Prior, the evaluation was done at the document level, but gradually drifted to sentence level (i.e., considering only subjective sentences) and nowadays entity\feature level is increasing.

Due to this evolution, definition of SA has also changed it as shown in Figure 2.

Definition 1 (Liu, 2012)**:** A SA is a process having binary tuples, '*e*' is the entity for which the document is about and '*s*' is the sentiment about the document, i.e., $SA = \{e, g\}$.

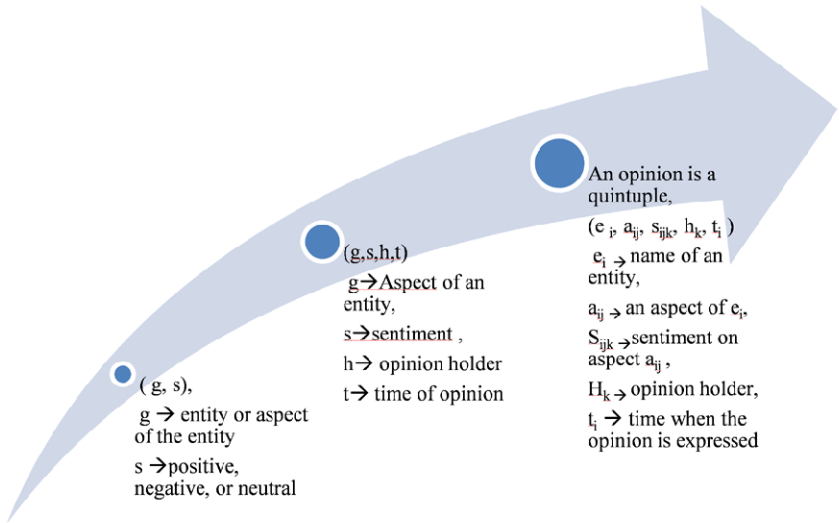
It gives the output as the only entity with its corresponding opinion. This definition does not focus on the issue that "*Who is the opinion holder*". The opinion holder may also change his opinion with time, due to which the dissimilarity can arise. Thus, researchers came up with another definition:

Definition 2 (Liu, 2012)**:** SA composed of four tuples, '*g*' is an aspect of the entity, '*s*' is sentiment, '*h*' is opinion holder and '*t*' is time of opinion, i.e., SA = {*g*, *s*, *h*, *t*}.

It was realised that a document can contain views about more than one aspect of an entity.

To cope up with this problem definition was again revised.

Figure 2 State-of-the-art sentiment analysis (see online version for colours)

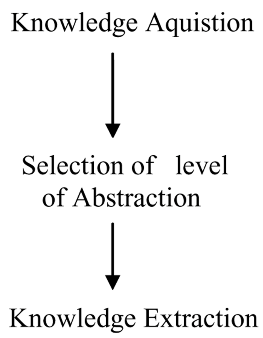


Definition 3 (Liu, 2012)**:** SA now is a quintuplet consisting '*e*' is an entity, '*a*' is an aspect of the entity, '*s*' is the sentiment on aspect, '*h*' is opinion holder and '*t*' is time of opinion, i.e., SA = {*e*, *a*, *s*, *h*, *t*}.

3 Components of building sentiment analyser

The steps to build a sentiment analyser is shown in Figure 3. It embodies the construction of a knowledge base, then identifying the level of abstraction and then choosing a method through which extraction and updation of the knowledge base can be taken up.

Figure 3 Components of building sentiment analyser



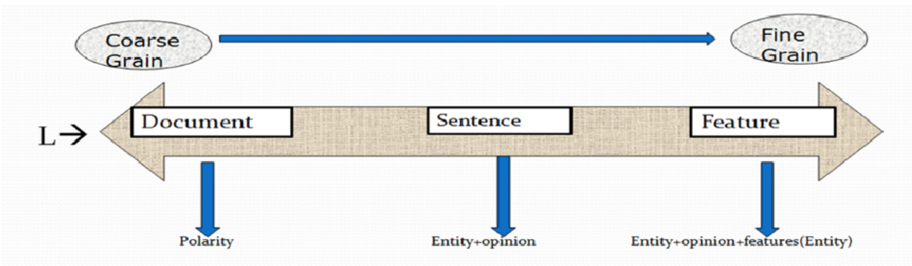
3.1 Knowledge acquisition

To extract real world sentiments through existing algorithms, it is required to train the analyser for learning about real world entities and sentiments (both explicit and implicit). Through various sources-like experts, lexicons, etc. training of the system can be done. The repository of such a huge knowledge is known as a knowledge base. It is the basic building block of sentiment analyser. The knowledge base can be built using linguistic experts, lexicons, textbooks, newspapers, magazines, etc. Linguistic experts are the key components to make machines programmed for accurate analysis by transferring their expertise to the knowledge base. Lexicons are also used to feed knowledge as it contains synonyms, semantic information about words in various languages. The above-mentioned two sources are time consuming and costly. To cope up with this, various autonomous techniques are used to feed knowledge automatically in the knowledge base. It is called as an autonomous expansion of knowledge base. The accuracy of automatic expansion of the knowledge base is less than the human annotators. On the contrary, knowledge engineers, although do not need to have expert domain knowledge. They only work on what is given to them in the knowledge base. They only need is to have a good sense in drawing efficient conclusions.

3.2 Selection of level of abstraction

SA facilitates the abstraction at three different levels, i.e., document level, sentence level and feature level as shown in Figure 4. Document level basically deal with the coarse grain granularity and feature level has the highest level of granularity, i.e., fine grain.

Figure 4 Level of granularity (see online version for colours)



Different researchers work on different levels as discussed below:

3.2.1 Document level

Document level is the upper level of abstraction in SA. Initially binary classification of the document is done at this level. It extracts the granularity at coarse grain level. Document level SA deduces the overall positive or negative opinion of the whole document. At the introductory stage of SA, work done at document level was in high demand.

The researchers mainly focused on the assumption found by Liu (2012) that the whole document was talking about only one entity. Thus extracted the opinion about that, but this assumption was failed when there are number of opinion holders submitting different opinions about an entity.

Hatzivassiloglou and McKeown (1998), and Moghaddam and Popowich (2010) proposed the SA task explicitly based on the adjectives, present in the English linguistic resources. They found that the exceptions to the rules of linguistics, i.e., conjunction rules-like and, nor, but, etc. They used supervised learning based on 21 million words of English. Their algorithm was not much effective as it required all the training data fully labelled. Unlabelled adjectives were not considered by the algorithm which in turn lowered down the accuracy.

Turney (2002) proposed unsupervised learning for SA. They used the adjectives and adverbs from which the orientation was calculated by an algorithm incorporating three steps as described below:

Step 1: Extracts phrases containing *adjectives* or *adverbs* and a context word to determine orientation by using POS tags (Brill, 1994).

Step 2: Estimates the orientation of the extracted phrases using the *pointwise mutual information (PMI)* (Turney, 2001). The opinion orientation (oo) of a phrase is computed based on its association with the positive reference word and its associations with the negative reference word.

Step 3: Computes the average oo of all phrases the review is classified as recommended if the average oo is positive.

This did not give good results when applied to different languages. As every language has its own linguistic rules.

Pang et al. (2002) gave a different approach of finding SA. They used supervised machine-learning methods such as Naive Bayes, Max. Entropy and SVM previously used for effective topic classification. In their work, they found that these machine-learning methods are better than human baselines for unigrams and bigrams. The results revealed that these machine-learning methods performed well in topic classification (McCallum and Nigam, 1998) than in SA. They reported that SVM had better performance Naive Bayes. Their work was dependent on binary classification of polarity, i.e., positive and negative.

Later O'Connor et al. (2010) included time series along with the existing methods. He calculated the sentiment by the inclusion of time series to find the political status of a person by using an unsupervised method. The problem faced by them was the different time zones present in the world. It hinders the evaluation of overall opinion of the document by considering the blogs, tweets, etc. posted by various people from distant places of the world.

To keep the temporal aspect of the SA and the time zone problem faced by O'Connor et al. (2010), 3 h burst time was considered by Thelwall et al. (2011).

Larsen and Aone (1999) proposed a method to extract the explicit features based on term frequency count. Terms having high frequency are denoted as candidate feature. After clustering these candidate features, they found that continuous centre adjustments in clusters gave better results. In their work, an efficient vector average damping algorithm was used for centre adjustment in clusters without the cost of additional time. The trade-off of the term frequency count approach used in this was that the non-frequent items were always ignored though may be valid entities.

Hung et al. (2012) worked on modifying the document quality by dividing it into five categories:

- high
- medium
- low
- duplicate (including redundant sentiments)
- spam.

This was done by using cosine similarity (sim $D_{x,y}$) between this WOM (word of mouth) (Hung et al., 2012) document X and its associated product description Y.

For generalisation of the task of opinion mining, many lexicons are built from the base lexicon, i.e., WordNet (Miller et al., 1990) in other languages by using various translation or transliteration schemes.

Stefano Baccianella et al. (2010) assigned a polarity score (numeric value) to each word from the WordNet called word score. Word score (w) is computed by averaging the score (both positive and negative) of the individual words present in the given text span related to the feature M by using equation (1).

$$\text{Word score } (W) = \frac{1}{n}\sum_{i=0}^{n} \text{posScore}(i) + \frac{1}{n}\left(-\sum_{i=0}^{n} \text{negScore}(i)\right), \tag{1}$$

where

posScore(i): positive score

negScore(i): negative score, respectively, found as of ith synset of word in text span S.

n: total number of synsets of word.

This helps to make a standard numeric value for each word and named as SentiWordNet (Esuli and Sebastiani, 2006).

It was observed that few work done in the context of cross domain analysis. Bollegala et al. (2013) proposed SA in cross domain at document level. For this work both supervised and unsupervised techniques of machine learning were used. The accuracy of this did not compete with the existing baseline methods. Later on, the work of SA focused on the meaning instead of the keyword matching approach.

Maas et al. (2011) proposed a technique for calculating the semantic likelihood. It combined unsupervised- and supervised-learning techniques. Their proposed vector-based model used to deal with semantic and sentiment similarities between different words. It did not work well in the cross domain SA.

In every linguistic world, apart from simple words, there are a number of complex words (like idioms, phrases, proverbs, etc.). These words also contribute to semantic analysis. Xie and Wang (2014) proposed a unsupervised technique for extracting sentiments from textual data using resources containing Chinese idioms to make a general classifier and found that it performs well for unannotated dataset. The research work in SA at the document is summarised in Table 2.

Table 2 Summarised work at document level

Author	*Work done*	*Trade off*
1. Turney (2002)	Used unsupervised techniques	Term having low frequency count was discarded from the analysis of sentiments
2. Pang et al. (2002)	Used various machine learning techniques like Naive Bayes, maximum entropy classification, and support vector machines Found that Naive Bayes gave worst results and SVM gave more accurate results	Did not perform well on a traditional topic based sentiment analysis
3. Liu (2012)	Found that the assumption was lacking when there is more than one opinion holder	The overall opinion was relatively difficult to calculate in case of opinion given by the number of users for the same named entity
4. O'Connor et al. (2010)	Extracted opinion using topic based on the occurrence of words, i.e., in the micro blogs	The relevant message selection was the loophole in this as all the messages are considered for analysis
5. Thelwall et al. (2011)	Found that the popular events were more concerned to negative posts Word frequency count was used for this	The time scale used for this is based on hours, due to which the change in time frame gave results as a different list of popular events
6. Hung et al. (2012)	Differentiated the document by its quality by using the cosine similarity measure along with SentiWordNet	A document containing fake review was not considered
7. Baccianella et al. (2010)	Results of evaluating SENTIWORDNET 3.0 were compared with a fragment of WORDNET 3.0 (manually annotated for positivity, negativity, and neutrality). The results indicated accuracy improvements of about 20% with respect to SENTIWORDNET 1.0 by adding semi supervised techniques	Did not work well for some words. E.g. term 'bad' SentiWordNet gave the results as pos = 0.625, neg = 0.125, obj = 0.25 On the other hand, the actual value by human annotators was pos = 0, neg = 1, obj = 0, which are completely conflicting (Brody and Diakopoulos, 2011)
8. Larsen and Aone (1999)	Extraction of features by using clustering techniques along with tfidf (term frequency inverse document frequency) and found that continuous adjustment in the centres of the clusters gave more refined results	Infrequent features were discarded which may be important and the calculation of the cluster centre was difficult as the assumption at document level is that single document talk about only one entity. When the document contained multiple topics, then the clustering was failed to deal with numbers of topics presented in a document

Table 2 Summarised work at document level (continued)

Author	*Work done*	*Trade off*
9. Hatzivassiloglou and McKeown (1998)	Used in various linguistic rules to extract the polarity of adjectives. Conjunction rules found in English language was used	Did not work well for all the sentences i.e., This store is good and costly Good and costly are positive and negative words respectively in practice, but according to conjunction rules of 'and' both of these either have positive or negative polarity
10. Bollegala et al. (2013)	Gave a cross domain sentiment analysis based on supervised and unsupervised learning techniques	The main problem was word sense disambiguation
11. Maas et al. (2011)	Worked based on the meaning of words. A vector based model was used for calculating the semantic orientation of various words	Did not work well with words having different meaning even in a single domain E.g., word 'hot' has different meaning for domain named food Hot coffee → +ve Hot beer → –ve
12. Xie and Wang (2014)	Worked in complex terms like idioms, etc. in Chinese language and found that it worked well for unannotated data	Accuracy of bootstrapping self training classifiers was low

It can be seen from the above table that fake review detection is the major area of concern for getting the quality review. It was also observed that mostly term frequency method is used for extraction of named entity as it is the simplest method. The term frequency method also has a shortcoming that the neglected infrequent items may contain valuable information. It can also be seen that the SVM is the most efficient method for complex sentences and Naive Bayes is good for short and simple sentences.

3.2.2 Sentence level

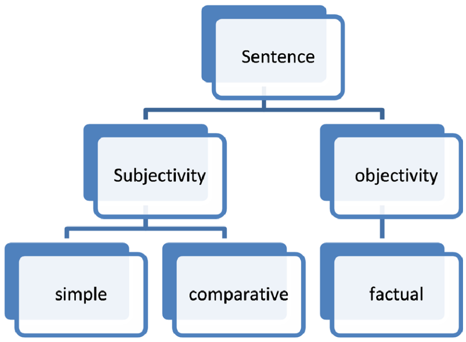
Document level and sentence level sentiment generation processes are almost the same except at sentence level the main task is the subjectivity detection. Sentence level SA involves subjectivity detection, thus increases the granularity than document level. The document contains sentences which are further categorised as subjective and objective in nature as shown in Figure 5. Analysis of an opinion depends on the individual's experience about any named entity. The main task at this level is to exclude all the objective sentences (containing factual information) and focus on subjective sentences.

E.g., India has 29 states. Tourists mostly like to visit religious places.

It can be seen that the sentences above have objective as well as subjective background as shown in Figure 5.

Hatzivassiloglou and Wiebe (2000) examined how subjectivity at sentence level gets affected by the orientation of adjectives in the SA and also figured out the gradability at sentence level focusing primarily on subjectivity. The goal of their research was to tell whether a given sentence is subjective or objective by considering the adjectives appearing in that sentence. This method did not work well in a cross domain. As word can be positive in one domain may be negative in another domain. It needed a long list of adjectives for the training data which increased the cost of SA task.

Figure 5 Sentence level (see online version for colours)



Pang and Lee (2004) also concluded that by compression the overall polarity is not affected. Here, compression means exclude the irrelevant part from the document which is not contributing to the generation of the sentiment. Minimum cut detection in graphs for SA along with Naive Bayes and support vector machines (SVMs) is used for subjectivity detection.

Furthermore, the low accuracy is another problem of adaptive methods because of disambiguation of sentiment words in different domains or languages. Disambiguation of polarity words present in the sentences actually does not provide the polarity.

E.g., This work is done by national trust for poor people.

In the above sentence, the opinion words *trust* and *poor* does not actually generate opinion. The accuracy depends if these are taken as neutral, which in the general framework of SA is treated as opinion words.

Wilson et al. (2005) determined the polarity of the context by the subjectivity clues. The evaluation consists of two steps, described below:

Step 1: Classifying clues in context as neutral or polar. They extracted 28 features summarised as: *Word features* (context, prior polarity, etc.).

> *Modification features*-linguistic (is intensifier, proceeded by adjective, dependency phrase info, etc.).
>
> *Sentence features* (pronoun in sentence, etc.)
>
> *Document Features* (document topic)

Step 2: The clue instances marked as polar are classified into their polarity (positive, negative, neutral or both). This time they used 10 features:

> *Word features* (token, word polarity)
>
> *Polarity features* (negation, modified polarity, etc.)

From their work they evaluated that the combination of all features yields best performance.

Liu (2012) addressed two problems in his paper with the existing methods:

- opinion words whose semantic orientations are context dependent
- sentences having more than one opinion word.

He gave the solution by aggregating multiple opinion words in the same sentence. Prior the work was more focusing on English language because of the less advancement in the translation and transliteration techniques. As the translation techniques improved, the work of SA also advanced. Nowadays, researchers are working to extract opinion in every possible language. Boiy and Moens (2009) proposed supervised machine learning for extracting the sentiments from blogs, review and forum in English, Dutch and French.

An and Hagiwara (2014) focused on deducing the emotion of a person by considering top five adjectives. If the top five adjectives are negative, then the person is deduced in a sad mood or else is happy. The summarised work of sentence level polarity is shown in Table 3.

Table 3 Summarised work at sentence level

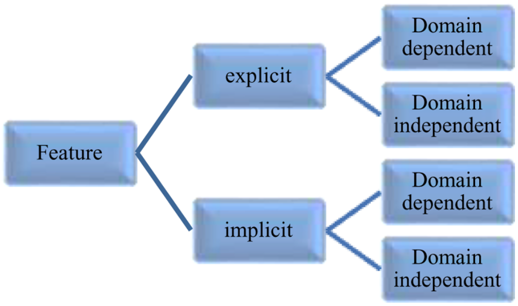
Author	*Work done*	*Trade off*
1. Pang and lee (2004)	Labelled the sentences in the document as either subjective or objective and discarding the latter	Word sense Disambiguation was the problem (Wilson et al., 2005)
	Applied a standard machine learning classifier to the resulting extract Naïve Bayes' and SVM were used for subjectivity and for semantic analysis was done by using minimum cut graphs	Separating irrelevant sentences from the relevant was a costly process in terms of time.
	Excluded the irrelevant or potentially misleading text	
	Found that Naive Bayes was efficient	
2 Hatzivassiloglou and Wiebe (2000)	Worked on identifying subjectivity by using negative and positive adjectives	Did not perform well for the use of lexicons for resource scarce language
3. Boiy and Moens (2009)	Dealt with informal text present on the web	The use of language corpus was varied which contain a huge number of misspelled training data
	Dealt with multilingual text in Dutch, English and French	
	Used machine models based on supervised learning	
4. Wilson et al. (2005)	Calculated the contextual polarity instead of assigning the prior polarity for context independent approach of sentiment analysis	Did not handle neutral polarity sentences efficiently, which degrades the system performance
5. An and Hagiwara (2014)	Found the emotion from the textual data by taking first five adjectives with the highest score	Did not work well for unstructured sentences or slangs
6. Liu (2012)	Dealt with context dependent opinion words	Did not work well for unstructured sentences
	Proposed a method of calculating the polarity of sentences having more than one word	

From the above table, it is figured out that handling unstructured data is needed. Multi-linguality has been taken care of but there is scope of work in multi-linguality in a single sentence. Research should be carried out for resource scarce language. Disambiguous polar words are also required to be handled properly.

3.2.3 Feature level

It is the highest level of granularity. Here features can be classified as implicit or explicit which further divided into domain independent and domain dependent as shown in Figure 6. Many researchers have worked at this level. Florian et al. (2003) described four named entity classifiers rule based, hidden Markov model (HMM), robust risk minimisation (RRM) classifier based on regularised winnow method (Zhang et al., 2002) and max entropy classifier for extracting named entities in English as well as in German. They found that RRM is better than others. It was also observed that these methods gave unsatisfactory results in a cross lingual environment.

Figure 6 Feature level abstraction (see online version for colours)



Zhu et al. (2009) used association rule mining (ARM) task to find the frequent features in Chinese language. The association was done by using the Apriori algorithm (Agarwal and Srikant, 1994) based on a breadth first search for calculating the frequent items. The reason of using the Apriori algorithm was to make the process faster than previously used approaches. To extract the candidate features from all the calculated frequent items, topic correlation filtration was used. The drawback of the work was that non-frequent items were always ignored.

Yi and Niblack (2005) followed the keyword approach, i.e., the spotter to find the feature by explicitly mention the predefined arbitrary itemset. After having a bunch of appearance of these itemsets, all get collected under one topic name by considering the synonym of the itemset. Disambiguator was also used to correctly find the items related to the topic, i.e., predefined set of items contains *Sun Microsoft* and during tokenisation we may get *sun* and *microsoft* as two words. He suggested that a relationship is needed to be found between both for extracting the named entity (Sun Microsoft). This is done by computing the term frequency-inverse document frequency score based on off-topic and on-topic items (Yi and Niblack, 2005).

Kucuktunc et al. (2012) used machine-learning techniques and found various characteristics of SA, such as the temporal nature of the sentiments, demographic effect on sentiments and contextual dependency was also figured out. The main findings (Kucuktunc et al., 2012) were as follows:

- there was a strong dependency on the topic
- demographic factors suggested a strong influence in our data
- sentiments had shown temporal variation on monthly, daily, hourly basis.

Zhao and Zhou (2009) has given a method to find the aspect by using the relationship between the opinion word and the feature word. The research is encouraged by developing multi-aspect (Zhu et al., 2012) SA. The main aim of this was to make the task of opinion mining domain independent. Part-of-speech taggers are used to find the relationship.

Later the researchers put their efforts in implicit feature extraction instead of explicit. Srivastava et al. (2010) worked for implicit feature extraction by using binary grammatical dependencies between different opinion words and feature words by using POS tagging in a specific domain of product reviews. Ding et al. (2008) found the contextual dependencies in the analysis of sentiment and explained well how these context-dependent opinion words can be handled efficiently. Basic intra-conjunction rules were developed for the purpose which says opinion on both sides of 'and' should have the same polarity.

Xie and Wang (2014) used a general classifier to reduce the impact of domain dependency by taking the idioms and proverbs having universally same meaning in all the contexts in which they appear by using the bootstrapping method.

Ding et al. (2009) worked on finding the implicit and explicit entity. Two approaches were considered: one was for explicit entities, i.e., entity discovery and another was entity assignment used for implicit entity extraction.

Pontiki et al. (2014) focused in the development of aspect-based SA. Their work is based on the extraction of entity, its aspects and polarity corresponding to each aspect (explicit as well as implicit).

Che et al. (2015) developed a process known as Sent_Comp. It takes only the valueable information for SA by compressing the long sentences into short at the sake of having the same polarity.

The work at feature level is summarised in Table 4.

Table 4 Summarised work at feature level

Author	*Work done*	*Trade off*
1. Florian et al. (2003)	Various machine learning methods are used for named entity recognition (NER) like RRM, TBL, Max. Entropy and HMM	Does not present the most effective results for German as it produced in English
	Found that the RRM algorithm suited well to deal with additional feature selection	Does not find context dependent opinion words, e.g., small, long, fast. The solution is given by Riloff and Wiebe (2003)
2. Kucuktunc et al. (2012)	Investigation of the connection between sentiments and the influence of factors such as gender, age, education level, the topic at hand, or even the time of the day on sentiments in the context of a large online question answering site by using the topic correlation method	Works well on explicitly mentioned topics, but the implicit topics are not considered in their work

Table 4 Summarised work at feature level (continued)

Author	*Work done*	*Trade off*
3. Zhao and Zhou (2009)	Work efficiently in a domain independent environment by using the bootstrapping method	Performance depends on seeding of training data. As much as seeding is done efficient analysis comes out
4. Srivastava et al. (2010)	Extracted sentiments or opinion by using the grammatical structure of the collection of words. i.e., BGD (binary grammatical dependencies)	Better results for domain specific feature extraction and does not suite well in domain independent feature extraction (Xieand Wang, 2014)
5. Ding et al. (2008)	Work on contextual dependency for extracting the features by introducing intra conjunction rules in a sentence	Cannot be applicable to all the languages due to the lack of a general structure as every language has a different linguistic structure
6. Xie and Wang (2014)	Used Chinese idiom resource by using unsupervised learning in a domain independent for building the sentiment classifier in a given domain and found that it is more stable	Trigrams or N-grams are not included Accuracy is low as compared to standard Naive Bayes
7. Zhu et al. (2009)	Worked for faster analysis of sentiments by using an apriori algorithm for S.A. from Chinese textual data	Did not give the general rules which can be implemented onto other languages
8. Yi and Niblack (2005)	Used keyword approach to find explicit feature extraction Word sense disambiguation is also solved to some extent	Unable to find the implicit features
9. Ding et al. (2009)	Efficient in finding explicit and implicit features in Chinese language	Domain dependent nature makes this task inefficient in some areas
10. Pontiki et al. (2014)	Foster research in entity, its aspects and polarity corresponding to each aspect	Domain dependent
11.Che et al. (2015)	Compressed the sentences based on the redundancy and retain the same polarity as the original sentence	Inclusion of the semantic network may result in different polarity. Because the meaning of a word is different in other domains

By recapitulating feature level SA, it is concluded that less work is done in implicit feature extraction in multi-lingual data. It also needs attention for cross-domain feature extraction.

3.3 Knowledge extraction

At this part of analyser, the results of queries are to deduce on the basis of pre-existing knowledge base through using various existing models and learning techniques. There is no hard rule for using any model.

3.3.1 Models

To access the huge repository of data which sometimes is very useful and sometimes meaningless, the need of machine-learning arose. There are various machine-learning models used by SA to extract meaningful data. The choice of appropriate model is very necessary and complex too. There is no hard rule on the choice of models. When a wrong model gives a correct prediction then that model is treated as a useful model. Grorge Box stated "Essentially all models are wrong, but some are useful".

Models are categorised in the following types:

i *Predictive models* interpret the future value of the question in the subject. These models may termed as today for future (TF) models. The effectiveness of the models used for the prediction depends on its outcome, i.e., the accuracy of predictive values. Naive Bayes, (Pang et al., 2002) ARM, (Pang et al., 2002) etc. methods are used to predict the upcoming values for the output.

ii *Descriptive models* (Liu, 2012) are actually used to summarise the analysis. Classification, clustering and ARM are the various techniques used for the descriptive models in SA.

iii *User-sensitive models* (Liu, 2012) *of opinions* are used to reflect the fact that the same opinion could be positive for one user group and negative for another. E.g. A camera can be rated as good for a photographer, but too complex for the casual user.

iv *Author authority models* account for the fact that personal authority often influences whether the reader will be influenced (Liu, 2012) significantly by a given review or not (e.g., an endorsement from an unknown person does not carry the same weight as that of a respected expert).

3.3.2 Learning techniques

There are various learning techniques in literature which are used to train the system as shown below:

i *Supervised learning* is used for the labelled data to train the system. Many machine-learning algorithms are used for supervised learning-like SVM, HMM, etc. Pang et al. (2002) proposed Naive Bayes, maximum entropy and SVM approach for SA to classify movie reviews into positive and negative opinions. Feature extraction in these is based on the combination of all active named entities. They reported that those algorithms did not perform as well on sentiment classification as in text classification. Florian et al. (2003) used HMM, a RRM classifier based on regularised winnow methods for named entity extraction in addition to the models used by Pang et al. (2002). It is found that RRM classifier was the best amongst all for feature extraction. Correlation method for feature extraction was used by considering the relationship of bigrams, trigrams and Ngrams with the topic by using a distance measure proposed by Liu et al. (2005). Supervised techniques gave the best results in a specific domain-oriented environment.

ii *Semi-supervised learning* algorithms are used when the input is the combination of labelled and unlabelled data. Graph-based semi-supervised-learning methods based on minimum cuts proposed by Pang and Lee (2004). Etzioni et al. (2004) proposed bootstrapping method for the minimisation of manual labelling of input data. The performance of bootstrapping depends on the seeding done for the extraction at the time of training the system. Riloff and Wiebe (2003) used semi-supervised technique, i.e., bootstrapping for annotated data using linguistic clues to extract patterns for subjectivity.

The most broadcast methods for annotating words or expressions with either subjectivity or polarity is the bootstrapping method.

iii *Unsupervised learning* algorithms are used to deal with completely unlabelled or annotated data analysis. Rules are built to train the system by using unlabelled data. Florian et al. (2003) proposed agglomerative classifier which was used to do classification based on the active features and their combination. Transformation-based learning classifier based on unsupervised techniques, i.e., rule-based. Chamlertwat et al. (2012) proposed unsupervised machine-learning technique based on available lexicon. Unsupervised learning is widely useful, although less accurate as most of the data on the web is annotated.

The output of this part of sentiment analyser is basically the results or actual opinion about any entity. It sometimes can also be used to extend the existing knowledge base by adding new entities and opinion words.

4 Performance measurement

There are the following metrics broadly used to evaluate the system performance. Precision, recall, F-measure, human annotator agreement and accuracy.

To calculate precision, recall and accuracy the following metrics need to be defined:

- *True positives (TP)*: Number of positive examples labelled as positive.
- *False positives (FP)*: Number of negative examples labelled as positive.
- *True negatives (TN)*: Number of negative examples labelled as negative.
- *False negatives (FN)*: Number of positive examples labelled as negative.
- *Correct output (CO)*: Number of outputs of the system which are considered as correct by the human annotators.

1 *Recall* (https://class.coursera.org/nlp/lecture/142): It is the percentage of named entities present in the corpus that are found by the learning system. It is poor in case of less training data due to which the system is unable to cover all the terms. Recall can be calculated by equation (2).

$$\text{Recall}(R) = TP / (TP + FN). \tag{2}$$

2 *Precision* (https://class.coursera.org/nlp/lecture/142)*:* It is the accurate number of named entities found by the learning system. It is found high if it gives correct results. Precision can be calculated by equation (3).

$$\text{Precision}(P) = TP / (TP + FP). \tag{3}$$

3 Ac*curacy* (https://class.coursera.org/nlp/lecture/142): It is defined as the ratio of addition of true positive, true negative and true positive, true negative, false positive, false negative. Accuracy can be calculated by equation (4).

$$\text{Accuracy}(A) = (TP + TN) / (TP + TN + FP + FN). \quad (4)$$

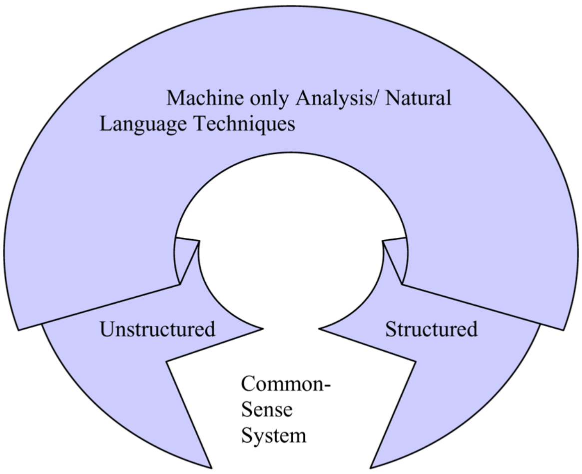
4 *Human agreement* (https://class.coursera.org/nlp/lecture/142) *(HA)*: It is the percentage of agreement between two or more humans for the correct output of the system. It is measured in terms of accuracy of the system by human annotators. It is also called as human annotator agreement. Human agreement can be calculated by equation (5).

$$\text{Human annotator agreement} = (CO) / (TP + TN + FP + FN). \quad (5)$$

5 Research gap

The future of SA is to design a system that needs to have a common sense similarity as humans. This is what a current state of an art system lacking. The research gap is shown in Figure 7. Today, due to the advancements in the processing task, many machine-learning techniques are improved. These are used for converting unstructured data to the structured data by using various available learning algorithms as shown in Figure 7. The research gap is to develop a common-sense system for SA of any natural text.

Figure 7 Way to common sense system (see online version for colours)



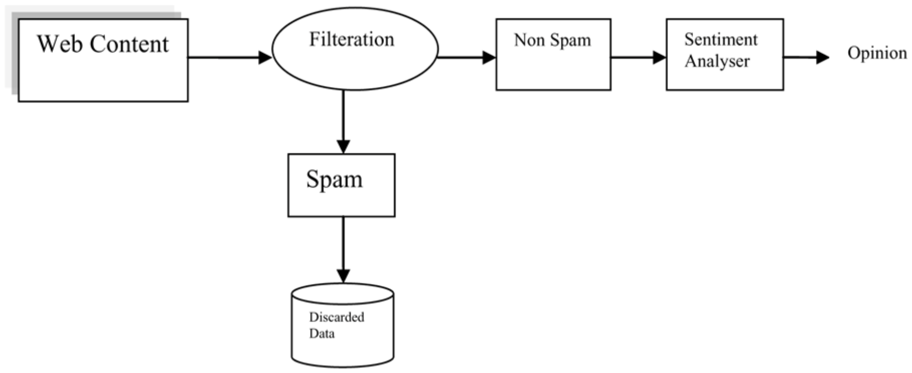
5.1 Current research directions

An ideal sentiment analyser should have the intellectual capability same as humans. The design of such system can be possible if research gaps are filled. To fill the research gap as shown in Figure 7 researchers need to work in the various directions. This section describes current research directions and various aspects which needs attention by the researchers to narrow down the path to ideal sentiment analyser.

5.1.1 Fake review detection

As the growth of electronic data is rising exponentially. The availability of data is no more the problem these days. Along with it, spam content is also increasing. The quality of SA is based on the truthfulness of the opinion feeding sentences to the system for extracting sentiment. The process of identification of the people hired by an organisation to post-superior reviews for their firm or to lower down the reputation of competitor is known as fake review detection. Fake review detection is the major concern these days as finding the true opinion of the product. Nowadays, researchers are working on refining the input by filtering the bogus reviews or comment and providing a fine comb for fraud detection (Nitin and Liu, 2008; Wang et al., 2012; McCord and Chuah, 2011) and better customer profiling as shown in Figure 8. Unlike other researchers, Wang et al. (2015) considered the spam reviews instead of fully neglecting them.

Figure 8 Fake review filtration



Fake review detection is categorised into link spam and content spam (Nitin and Liu, 2008).

The research issues in fake review detection comprise:

1 *Detection of a bunch of spammers or individual spammers*: The output of any decision support system depends on the quality of the reviews taken for the analysis. Therefore, the need of excluding the fake reviews from the analysis is accelerated. Many researchers have worked on identifying the group of spammers (Lim et al., 2010; Mukherjee et al., 2012) from the reviews and able to achieve satisfactory results. On the other hand, finding individual review is also equally important as finding the group spammers. It is one of the toughest tasks in fake review detection. Various actionable measures can be taken to have a quality reviews like block ids of bogus reviewers, fine them, etc.

2 *Domain independence*: There are many domains-like movie, restaurant, products, etc. where spam detection is required. The prime motive of researchers is to design techniques to achieve fake review detection irrespective of any domain. This is called as domain independence.

3 *Content-based/metadata-based/factual data*: The fake review detection is based on three approaches:

- Content-based approach contains the data actually posted by the user.
- Metadata which is attached to every post by the server-like time of post, date of post, etc.
- Factual data are the data contains any facts or figures. Sentences which contain factual data are also called as objective sentences.

It is figured out from Table 5 that researchers till now did not use the factual information in a SA task or fake review detection. The inclusion of factual information may intensifying the research in fake review detection.

Table 5 Brief overview of fake review detection

S. No.	*Author*	*Group spammers*	*Individual spammers*	*Domain independency*	*Content based*	*Meta-data*	*Factual information*	*Results*
1	Jindal and Liu (2007)	×	✓	×	✓	✓	×	Accuracy = 0.78
2	Lim et al. (2010)	✓	×	×	✓		×	H.A. = 0.64
3	McCord and Chuah (2011)	×	✓	×	×	✓	×	A = 0.957
4	Mukherjee et al. (2012)	✓	×	×	×	✓	×	H.A. = 0.79
5	Wang et al. (2012)	×	✓	×	×	✓	×	H.A. = 0.60
6	Mukherjee et al.(2013)	×	✓	×	✓	✓	×	H.A. = 0.76

5.1.2 Temporal nature of sentiment generation

Another important dimension of SA, which is often overlooked is the temporal aspect (time factor in the SA definition). Researchers not only identify opinions, but also identify when they were stated, which makes it possible to detect shifts in attitudes over time. The temporal dimension is particularly important in applications such as monitoring the impact of marketing campaigns or containing damage to brands and companies through quick response to problems as soon as a significant number of users start reporting them online. Hence, there is a need of automatic assignment of high weightage to the recent reviews and low to the previously posted blog, review, etc.

Following are the various spheres where SA includes the temporal nature of reviews:

1 *Explicit/implicit topic word*: Most researchers give topic word (named entity) explicitly in the query or deduce it implicit from the text. Extracting topic word implicitly from the text is the best option but it is a complex task.

2 *Geographical dispersion of time*: The task of SA based on the temporal aspect becomes more complex because of variation in time globally. Hence, it is very tedious to get a collaborative real-time sentiment of the whole world.

3 *Forecast analysis*: Researchers are working to enhance forecast analysis by including the time in the process of SA.

4 *Hinged weightage to reviews with respect to time:* With the passage of time, the value of previous reviews is degrading. The present day review is much more important. The process of SA should include hinged weightage with respect to time to have time-oriented review.

It can be seen from Table 6 that researchers did not give a figurative value, i.e., weightage to the reviews based on the time. Therefore, the need of deducing the sentiments based on the standard time scale by using metadata gets less attention.

Table 6 Overview of temporal nature of sentiment analysis

S. No.	*Author*	*Explicitly mention of topic keyword*	*Implicitly deduce topic keyword*	*Handling the geographical dispersion of time*	*Forecast analysis*	*Weightage hinged to reviews w.r.t. time*	*Results*
1	O'Connor et al. (2010)	✓	×	×	✓	×	R = 63.5
2	Thelwall et al. (2011)	✓	×	✓	×	×	P = 0.013
3	Razavi et al. (2013)	✓	×	×	×	×	H.A. = 0.69
4	Dias et al. (2014)	✓	✓	×	×	×	P = 0.078 R = 0.78
5	Fukuhara et al. (2007)	✓	×	×	×	×	

5.1.3 Multi-linguality

Most of the people these days use the internet for communication. To maximise the number of interested people that can understand the text, writer often selects English even if the great majority of his readers has the same native language. Alternatively, he might write his texts, both in his native language and in English. This not only doubles the work needed for writing a document but also for maintaining it. Only 29.4% people on the internet know English found by Banea et al. (2011) in 2011. Rest of more than 70% people feel comfortable using their native language. For communication over the web, they use either their native language or the mixed language, i.e., native and any other formal language supported by the present day systems. This makes the researchers focused on numerous languages used by different parts of the world for SA. Following are the various aspects of handling multi-lingualism in textual data:

1 *Use of translation/transliteration*: The earlier work done in multi-lingual data was very less. The lack of availability of efficient translation techniques makes most of SA tasks in multi-lingual based on manually built bilingual lexicons. The research in

the field of SA is promoted with the advent of machine translation and transliteration, multi-lingual (Zhu et al., 2012) aspect of SA.

2 *Automatic expansion of lexicons in different languages*: To work with different languages the need of parallel lexicons arose. To get all the words in the dictionary is near to impossible as natural language contains some new words most of the time. So, the extension of the current lexicon in any language is the need of an hour to work with multi-lingual data.

3 *Collocation*: A word in one language has multiple translations in various other languages. It is hard to find bilingual collocation correspondence. Collocation includes the translation of a word based on its sense in the particular context.

4 *Implicit/explicit SA*: Using translation techniques, it became easy to find explicit entities along with their opinion. Finding implicit entities is still a problematic task. To perform SA implicitly a parallel corpus is required.

5 *Multi-linguality in a single sentence*: Sentences containing only the base language are easy to deal with. On the other hand, the presence of words other than base language are rejected by treating them as stop words which sometimes are valuable for extracting opinion.

The existing translation and transliteration techniques are failing to detect the words without explicitly mention its language. From Table 7, it is concluded that the researchers need to concentrate for handling these multi-lingual sentences for the qualitative SA.

5.1.4 Unstructured sentences

Currently, most automated machine-learning technology can access only structured content. For the most part, structured data refers to information with a high degree of organisation such that inclusion in a relational database is seamless and readily searchable by simple, straightforward search engine algorithms or other search operations whereas unstructured data is essentially the opposite. The lack of structure makes compilation a time- and energy-consuming task. Unstructured data are (loosely speaking) usually for humans, who do not easily interact with information in the strict linguistic format. Unstructured data contain:

1 *Informal text*: Sentences does not follow any grammatical rules as in English the structured sentences follows: Subject Verb Object (SVO) or Object–Verb–Subject (OVS) format.

Structured sentences: I bought the phone. It is very expensive.

Unstructured sentences: bought phone. Very expensive.

2 *Use of slangs (short words/abbreviation/spelling correction)*: Internet gives the privilege to its users to freely write the contents in any desired format. The people try to use alphanumeric words to minimise the number of words.

e.g., use of gr8 instead of great.

Table 7 Brief overview of multilinguality in sentiment analysis

S. no.	*Author*	*Implicit feature extraction*	*Explicit feature extraction*	*Translation/ transliteration*	*Automatic expansion of bilingual lexicon*	*Collocation*	*Multilinguality in a single sentence*	*Results*
1	Denecke (2008)	×	✓	✓	×	×	×	*A*-0.58-0.66
2	Banea et al. (2011)	×	✓	✓	×	×	×	*R* = 83.15 *P* = 67.76 *A* = 69.44
3	Lin et al. (2012)	×	✓	✓	✓	✓	×	*A* = 0.706
4	Hogenboom et al. (2014)	✓	✓	✓	✓	×	×	*A* = 0.62
5	Boiy and Moens (2009)	×	✓	✓	×	×	×	*A* = 0.87

Table 8 Brief overview of handling unstructured data in sentiment analysis

S. no.	*Author*	*Short words*	*Normalisation (Elongation of words/ alphanumberic/ unstructured sentences)*	*Dealing with stemming*	*Handling of idiomatic sentences*	*Abbreviation*	*Spelling correction*	*Change in strength of opinion the presence of long tail words or short words*	*Results*
1	UzZaman and Khan (2005)	✓	✓	×	×	✓	✓	×	–
2	Willett (2006)	×	×	✓	×	×	×	×	–
3	Whittle et al. (2010)	×	✓	×	×	×	×	×	$A = 0.58$
4	Brody and Diakopoulos (2011)	✓	✓	×	×	×	×	×	H.A. = 0.676
5	Eisenstein (2013)	✓	✓	×	×	×	×	×	–
6	Liu et al. (2012)	✓	✓	×	×	×	✓	×	–

3 *Normalisation (Removal of Noisy Words (Brody and Diakopoulos, 2011) and smileys/elongation of words)*: Many social sites give the user a template of images to show the happiness, sadness, etc. in the form of smileys as happy-☺, sad-☹. Apart from these people these days express their extreme happiness and sadness by stretching the word in format as grtttttttttttttttt instead of very much greater, lovelyyyyyyyyyyy are termed as noisy words.

4 *Stemming*: Stemming is a part of morphological analysis of the text which is to find a root word-like automatic is the root word for automation, the cat is the root word of cats, etc.

5 *Change in strength of opinion in the presence of long tail words or short words*: The presence of long-tail words give either intense high or low feelings about any entity. After normalisation it often loses its intensity which hinders the actual opinion.

For example, the word *Happpppppppyyyyyyyy* should be normalised to *very happy* instead of happy.

After recapitulating Table 8, it comes forward that the area of SA still lacking from various idiomatic phrases used in the textual data and the intensity change due to the short- and long-tail words into actual consideration. Researchers need to focus in handling text in its natural form. Thus, better algorithm is needed to handle slangs, noisy words, etc.

6 Conclusion

Finally, we have analysed the contribution of various researchers and were able to drill out research gap for the future researchers. The paper describes the basic building blocks of SA, i.e., knowledge acquisition, selection of levels of granularity and extraction. The paper presents research at various granularity levels. It was observed that the goal of the research in the area is to build a sentiment analyser which work like common sense. As a human being can understand a natural text with his commonsense, a state-of-the-art sentiment analyser should also work like that.

The various research gaps identified are multi-lingualism, fake review detection, temporal aspect of the SA generation and processing natural text. It is concluded that the importance of including factual information is also felt. There are huge factual data available on the web which can be used to more fruitful results. The researchers should try to extract sentiments based on factual information along with subjectivity instead of completely rejecting sentences containing factual data. Another most important factor is to work with multi-lingualism in a single sentence. The real-time SA is done based on time value of any blog, post, review, etc. These factors may uplift the research in this area and also refine the performance of decision support system.

References

Agarwal, R. and Srikant, R. (1994) 'Fast algorithms for mining association rules in large databases', *Proceedings of 20th International Conference on VLDB*, Vol. 1215, pp.487–499.

An, N.T.T. and Hagiwara, M. (2014) 'Adjective-based estimation of short sentence's impression', *International Conference on Kansei Engineering and Emotion Research*, 11–13 June, Keer, Linköping, pp.1–16.

Baccianella, S., Esuli, A. and Sebastiani, F. (2010) 'SENTIWORDNET 3.0: an enhanced lexical resource for sentiment analysis and opinion mining', *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, European Language Resources Association (ELRA), Valletta, Malta, pp.2200–2204.

Banea, C., Mihalcea, R. and Wiebe, J. (2011) 'Multilingual sentiment and subjectivity analysis', *Multilingual Natural Language Processing*, Vol. 6, pp.1–19.

Boiy, E. and Moens, M-F. (2009) 'A machine learning approach to sentiment analysis in multilingual web texts', *Journal Information Retrieval*, Vol. 12, No. 5, October, pp.526–558.

Bollegala, D., Weir, D. and Carroll, J. (2013) 'Cross-domain sentiment classification using a sentiment sensitive thesaurus', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 8, August, pp.1719–1731.

Brill, E. (1994) 'Some advances in transformation-based part of speech tagging', *Proceedings of the 12th National Conference on Artificial Intelligence*, AAAI Press, Menlo Park, CA, pp.722–727.

Brody, S. and Diakopoulos, N. (2011) 'Cooooooooooooooollllllllllllll!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs', *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, pp.562–570.

Chamlertwat, W., Bhattarakosol, P., Rungkasiri, T. and Haruechaiyasak, C. (2012) 'Discovering consumer insight from twitter via sentiment analysis', *Journal of Universal Computer Science*, Vol. 18, No. 8, pp.973–992.

Che, W., Zhao, Y., Guo, H., Su, Z. and Liu, T. (2015) 'Sentence compression for aspect-based sentiment analysis', *Transactions on Audio Speech and Language Processing*, Vol. 1, No. 99, p.1.

Denecke, K. (2008) 'Using SentiWordNet for multilingual sentiment analysis', *Proceedings of IEEE 24th International Conference on Data Engineering Workshop*, Cancun, pp.507–512.

Dias, G., Hasanuzzaman, M., Ferrari, S. and Mathet, Y. (2014) 'TempoWordNet for sentence time tagging', *Proceedings of the 23rd International Conference on World wide Web Companion (WWW'14)*, Seoul, Korea, ACM, pp.833–838.

Ding, X., Liu, B. and Yu, P.S. (2008) 'A holistic lexicon-based approach to opinion mining', *Proceedings of International Conference on Web Search and Data Mining*, New York, NY, USA, pp.231–240.

Ding, X., Liu, B. and Zhang, L. (2009) 'Entity discovery and assignment for opinion mining applications', *Proceedings of the 15th ACM SIGKDD*, NY, USA, pp.1125–1134.

Eisenstein, J. (2013) 'What to do about bad language on the internet', *Proceedings of NAACL-HLT*, Georgia, pp.359–369.

Esuli, A. and Sebastiani, F. (2006) 'SentiWordNet: a publicly available lexical resource for opinion mining', *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, Genova, IT, pp.417–422.

Etzioni, O., Cafarella, M. and Downey, D. (2004) 'Web scale information extraction in KnowItAll', *Proceedings of the 13th international conference on World Wide Web*, NY, USA, pp.100–110.

Florian, R., Ittycheriah, A., Jing, H. and Zhang, T. (2003) 'Named entity recognition through classifier combination', *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL*, PA, USA, Vol. 4, pp.168–171.

Fukuhara, T., Nakagawa, H. and Nishida, T. (2007) 'Understanding sentiment of people from news articles: temporal sentiment analysis of social events", *Proceedings of the International Conference on Weblogs and Social Media ICWSM*, (*2007*).

Hatzivassiloglou, V. and McKeown, K.R. (1998) 'Predicting the semantic orientation of adjectives', *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, PA, USA, pp.174–181

Hatzivassiloglou, V. and Wiebe, J.M. (2000) 'Effects of adjective orientation and gradability on sentence subjectivity', *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, Vol. 1, pp.299–305.

Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U. and de Jong, F. (2014) 'Multi-lingual support for lexicon-based sentiment analysis guided by semantics', *Decision Support Systems*, Vol. 62, June, pp.43–53.

Hung, C., Tsai, C-F. and Huang, H. (2012) 'Extracting word-of-mouth sentiments via SentiWordNet for document quality classification', *Recent Patents on Computer Science*, Vol. 5, No. 2, pp.145–152.

Jindal, N. and Liu, B. (2007) 'Review spam detection', *Proceeding of the 16th International Conference on World Wide Web, Banff*, Alberta, Canada, pp.1189–1190.

Kucuktunc, O., Barla Cambazoglu, B., Weber, I. and Ferhatosmanoglu, H. (2012) 'A large-scale sentiment analysis for yahoo! Answers', *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, NY, USA, pp.633–642.

Larsen, B. and Aone, C. (1999) 'Fast and effective text mining using linear-time document clustering', *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, pp.16–22.

Lim, E-P., Nguyen, V-A., Jindal, N., Liu, B. and Lauw, H.W. (2010) 'Detecting product review spammers using rating behaviors', *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, NY, USA, pp.939–948.

Lin, Z., Tan, S., Cheng, X., Xu, X. and Shi, W. (2012) 'Effective and efficient? Bilingual sentiment lexicon extraction using collocation alignment', *Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM)*, NY, USA, pp.1542–1546.

Liu, B. (2012) *Sentiment Analysis and opinion Mining*, Morgan & Claypool, May 2012, p.167.

Liu, F., Weng, F. and Jiang, X. (2012) 'A broad-coverage normalization system for social media language', *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, Jeju, Republic of Korea, pp.1035–1044.

Liu, L., Kang, J., Yu, J. and Wang, Z. (2005) 'A comparative study on unsupervised feature selection methods for text clustering', *Proceedings of the 2005 12th IEEE International Conference on Natural Language Processing and Knowledge Engineering (IEEE NLP-KE '05)*, Wuhan, China, pp.597–601.

Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C. (2011) 'Learning word vectors for sentiment analysis', *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp.142–150.

McCallum, A. and Nigam, K. (1998) 'A comparison of event models for Naive Bayes text classification', *Proceedings of the AAAI- Workshop on Earning for Text Categorization*, Vol. 752, pp.41–48.

McCord, M. and Chuah, M. (2011) 'Spam detection on twitter using traditional classifiers', *Proceedings of 8th International Conference on Autonomic and Trusted Computing*, Banff, Canada, pp.175–186.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. (1990) 'WordNet: an online lexical database', *International Journal of Lexicography*, Vol. 3, No. 4, pp.235–244.

Moghaddam, S. and Popowich, F. (2010) *Opinion Polarity Identification through Adjectives*, CoRR, abs/1011.4623.

Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M. and Ghosh, R. (2013) 'Spotting opinion spammers using behavioral footprints', *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Illinois, USA, pp.632–640.

Mukherjee, A., Liu, B. and Glance, N. (2012) 'Spotting fake reviewer groups in consumer reviews', *International World Wide Web Conference (WWW-2012)*, Lyon, France, 2012, pp.191–200.

Nitin, J. and Liu, B. (2008) 'Opinion spam and analysis', *Proceedings of the 2008 International Conference on Web Search and Data Mining*, ACM, New York, USA, pp.219–230.

O'Connor, B., Balasubramanyan, R., Routledge, B.R. and Smith, N.A. (2010) 'From tweets to polls: linking text sentiment to public opinion time series', *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington DC, pp.1–8.

Pang, B. and Lee, L. (2004) 'A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts', *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Article No. 271, PA, USA, pp.271–278.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up? Sentiment classification using machine learning techniques', *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, Vol. 10, pp.79–86.

Pontiki, M., Galanis, D., Pavlopoulos, I., Papageorgiou, H., Androutsopoulos, I. and Manandhar, S. (2014) 'SemEval 2014 Task 4: aspect based sentiment analysis', *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) at (COLING 2014)*, Dublin, Ireland, pp.27–35.

Razavi, A.H., Matwin, S., De Koninck, J. and Amini, R.R. (2013) 'Dream sentiment analysis using second order soft co-occurrences (SOSCO) and time course representations', *Journal of Intelligent Information Systems*, pp.393–414.

Riloff, E. and Wiebe, J. (2003) 'Learning extraction pattern for subjective expressions', *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*, Sapporo, Japan.

Srivastava, R., Bhatia, M.P.S., Srivastava, H.K. and Sahu, C.P. (2010) 'Exploiting grammatical dependencies for fine-grained opinion mining', *Proceedings of IEEE International Conference on Computer & Communication*, September, Allahabad, Uttar Pradesh, pp.768–775.

Thelwall, M., Buckley, K. and Paltoglou, G. (2011) 'Sentiment in twitter events', *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 2, February, pp.406–418.

Turney, P.D. (2001) 'Mining the Web for synonyms: PMI-IR versus LSA on TOEFL', *Proceedings of the 12th European Conference on Machine Learning*, Springer-Verlag, Berlin, pp.491–502.

Turney, P.D. (2002) 'Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews', *Proceedings of the Association for Computational Linguistics (ACL)*, Stroudsburg, PA, USA, pp.417–424.

UzZaman, N. and Khan, M. (2005) 'T12: an advanced text input system with phonetic support for mobile devices', *2nd International Conference on Proceedings of Mobile Technology, Applications and Systems, 2005*, Guangzhou, pp.1–7.

Wang, G., Xie, S., Liu, B. and Yu, P.S. (2012) 'Identify online store review spammers via social review graph', *Journal ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 3, No. 4, September, pp.1–21, Article No. 61.

Wang, J-z., Yan, Z., Yang, L.T. and Huang, B-x. (2015) 'An approach to rank reviews by fusing and mining opinions based on review pertinence', *Information Fusion*, Vol. 23, May, pp.3–15.

Whittle, J., Simm, W., Ferrario, M-A., Frankova, K., Garton, L., Woodcock, A., Nasa, B., Binner, J. and Ariyatum, A. (2010) 'VoiceYourView: collecting real-time feedback on the design of public spaces', *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, NY, USA, pp.41–50.

Willett, P. (2006) 'The porter stemming algorithm: then and now', *Electronic Library and Information Systems*, 2006, Vol. 40, No. 3, pp.219–223.

Wilson, T., Wiebe, J. and Hoffmann, P. (2005) 'Recognizing contextual polarity in phrase-level sentiment analysis', *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, pp.347–354.

Xie, S-x. and Wang, T. (2014) 'Construction of unsupervised sentiment classifier on idioms resources', *Journal of Central South University*, Vol. 21, No. 4, April, pp.1376–1384.

Yi, J. and Niblack, W. (2005) 'Sentiment mining in webfountain', *Proceedings of the 21st International Conference on Data Engineering (ICDE)*, Tokyo, Japan, pp.1073–1083.

Zhang, T., Damerau, F. and Johnson, D. (2002) 'Text chunking based on a generalization of winnow', *Journal of Machine Learning Research*, Vol. 2, March, pp.615–637.

Zhao, W. and Zhou, Y. (2009) 'A template-based approach to extract product features and sentiment words', *Proceedings of International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE 2009)*, September, Dalian, China, pp.1–5.

Zhu, J., Zhang, C. and Ma, M.Y. (2012) 'Multi-aspect rating inference with aspect-based segmentation', *IEEE Transactions on Affective Computing*, Vol. 3, No. 4, pp.469–481.

Zhu, S., Liu, Y., Liu, M. and Tian, P. (2009) 'Research on feature extraction from Chinese Text for opinion mining', *Proceedings of International Conference on Asian Languages Processing*, Singapore, pp.7–10.

Websites

http://onesecond.designly.com/ (Accessed June, 2014).

https://class.coursera.org/nlp/lecture/142 (Accessed May, 2014).