Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context

Bagus Setya Rintyarna*

Department of Informatics,
Institut Teknologi Sepuluh Nopember,
Surabaya, Indonesia
and
Department of Electrical Engineering,
Universitas Muhammadiyah Jember,
Jember, Indonesia
Email: bagus.setya@unmuhjember.ac.id
*Corresponding author

Riyanarto Sarno and
Chastine Fatichah

Department of Informatics,
Institut Teknologi Sepuluh Nopember,
Surabaya, Indonesia
Email: riyanarto@if.its.ac.id
Email: chastine@if.its.ac.id

Abstract: Commonly, product review analysis includes extracting sentiment from product documents. The contextual aspect contained in a review document has potential to improve results obtained by the sentiment analysis task. In this regard, this paper proposes an approach that takes into account both local and global context. The main contribution of this work is threefold. Firstly, local context is defined and the graph-based word sense disambiguation (WSD) method is extended to assign the correct sense of a word in the context of a sentence. Secondly, global context is defined for addressing contextual issues related to the specific domain of a review document by using an improved SentiCircle-based method. Thirdly, a weighted mean-based strategy to determine sentiment value at document level is presented. Several experiments were conducted to assess the proposed method. Overall, the proposed method outperformed the baseline method in the metrics of precision, recall, F-measure and accuracy.

Keywords: sentiment analysis; local context; global context; word sense disambiguation; WSD; SentiCircle; decision sciences.

Reference to this paper should be made as follows: Rintyarna, B.S., Sarno, R. and Fatichah, C. (2020) 'Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context', *Int. J. Information and Decision Sciences*, Vol. 12, No. 1, pp.75–101.

Biographical notes: Bagus Setya Rintyarna received his MKom from Institut Teknologi Sepuluh Nopember Surabaya majoring in Computer Science in 2012. Currently, he is a doctorate student in the same institution. His research interests is natural language processing and intelligence systems.

Riyanarto Sarno received his MSc and PhD in Computer Science from the University of Brunswick Canada in 1988 and 1992. Currently, he is a Professor in Institut Teknologi Sepuluh Nopember Surabaya. His research includes internet of things, enterprise computing, information management, intelligent systems and smart grids.

Chastine Fatichah received her MKom in Computer Science from Universitas Indonesia, Depok, Indonesia. She received her PhD from Tokyo Institute of Technology, Japan in 2012. Her research interests include medical image processing, soft computing, and data mining. She is currently a member of International Association of Engineers (IAENG). She is also Co-chief of postgraduate program of Informatics Department in Institute Teknologi Sepuluh Nopember Surabaya.

1 Introduction

Sentiment analysis (SA), sometimes also referred to as opinion mining or polarity classification, is a recent research topic at the crossroads of information retrieval and computational linguistics aimed at extracting sentiment from text data obtained from online activities. SA is basically a text classification task aimed at classifying a piece of text as positive, neutral or negative (Coutinho and Figueiredo, 2015). Sentiment is people's attitude (Medhat et al., 2014) or emotion polarity (Cui et al., 2016) toward an entity within a given text. Since emotion plays an important role in decision-making (Sarno et al., 2016), SA is a useful tool (Dat et al., 2017) in a broad range of applications (Giatsoglou et al., 2016), including product review analysis. One of the most popular approaches of dealing with SA tasks is the lexicon-based approach, which relies on a corpus or sentiment lexicon (Giatsoglou et al., 2016). A sentiment lexicon is a pre-built collection of words with their corresponding prior sentiment value (Baccianella and Esuli, 2010). Several sentiment lexicons are publicly available for this purpose, such as Opinion Lexicon, General Inquirer and SentiWordNet (Muhammad et al., 2016).

Lexicon-based SA involves extracting prior sentiment values from the lexicon and aggregating the score at sentence or document level in order to classify text data into two sentiment classes, i.e., positive or negative (Giatsoglou et al., 2016). However, the sentiment value of a word/term appearing in different texts can vary because of their different contextual meaning (Pamungkas et al., 2017). Different meanings arise from the interaction between a word and the particular neighbourhood in the sentence (Hung and Chen 2016). To obtain better results from lexicon-based SA, an approach to capture contextual sentiment value is needed. In this study, this issue is referred to as *local context*.

Moreover, the specific text domain also potentially affects the polarity of the prior sentiment value of a word. For product reviews, this domain can be defined according to the product type in a product review, such as computer, Smartphone or automobile. For instance, while 'long battery life' in the camera review domain implies positive polarity, 'long response time' in the computer review domain implies negative polarity. The word 'long' in these two different local neighbourhoods may have the same meaning (being or indicating a relatively great or greater than average duration or passage of time, or a duration as specified), which means that its prior sentiment value is the same. Still, from the example it is obvious that both words have different polarity. Saif et al. (2016) call this contextual issue contextual semantics while in the present work it is referred to as global context.

In this work, we propose a contextual approach in order to achieve better performance of lexicon-based SA. Referring to the previous paragraphs, the motivation for this research comprised:

1 Taking into account that the local neighbourhood of the word in the sentence can be used to tune the prior sentiment value provided by a sentiment lexicon. In this research, we call this issue local context. This problem has been investigated in a previous research (Muhammad et al., 2016), where local context was defined as the influence of negation, intensifiers and discourse, however, without considering the relationship with neighbouring words. The state of the art study on context derived from neighbouring words (Sharma et al., 2017) did not take into account the potential effect of different product review domains on the sentiment value of words.

2 Capturing the potential change of prior sentiment value in different product review domains. We call this issue global context. A previous study (Saif et al., 2016) proposed an approach to deal with this problem, however, without considering the latent effect of local context on the results of lexicon-based SA.

This study addresses the potential performance reduction of SA caused by both local and global context. In the first phase of the proposed method, local context is captured using a weighted graph-based word sense disambiguation (WSD) method (Sinha and Mihalcea, 2007) to obtain the contextual sentiment value of a word from its prior sentiment value as provided by a general-purpose sentiment lexicon; in this research, we employed SentiWordNet (Baccianella and Esuli, 2010). In the second phase of the proposed method, potential change of the sentiment value with respect to different product review domains is detected by using an extended senti-median method (Saif et al., 2016) as well as providing similarity-based pivot word determination for the method. Since we wanted to evaluate our approach at the document level, we also propose a weighted mean-based formula to aggregate sentiment values from sentence-level to document-level.

The rest of this paper is organised into the following sections: Section 2 describes related works on SA, including contextual approaches of SA that have been done previously. Section 3 explains the approach proposed in this work including the algorithms and formulas that are introduced. In Section 4 the scenario used in the experiment in this research is described and the results of the method are compared with a baseline method. Finally, Section 5 summarises the result of the work that was done.

2 Related work

In this section, we explain related work, including supervised machine-learning approaches, unsupervised lexicon-based approaches and contextual analysis with lexicon-based approaches. In terms of machine learning, SA is a text classification task that classifies a piece of text data into labelled text data by employing a machine-learning algorithm (Pratama and Sarno, 2015). Several machine-learning approaches have been employed for SA tasks. An ensemble classifier method has been proposed by (Onan et al., 2016). The classifiers employed were logistic regression, linear discriminant analysis, naïve Bayes and Bayesian logistic regression, with support vector machine as the base learner. The method was designed to give higher weight to classifiers with better predictive performance. The appropriate value of the weight for each classifier is assigned by employing the multi-objective differential evolution (MODE) algorithm. Compared to several baseline methods, such as majority voting, random subspace, bagging and AdaBoost, the method improved performance on nine public datasets.

The statistical features of unigrams and bigrams and their combination have been used to classify sentiment in financial news (Yazdani et al., 2017). By taking benefit from traditional feature weighting methods (binary, TF, TF-IDF), a new feature space with different dimension was generated. The experiment used Google Finance news and was conducted by using SVM with two kernel methods, i.e., a linear and a Gaussian radial basis function (RBF). The result of the experiment indicated that the combination of unigrams and bigrams along with the TF-IDF feature weighting method increased the sentiment classification accuracy for financial news.

An ensemble technique was employed to enhance deep learning in SA (Araque et al., 2017). As the baseline model, this work employed a word-embedding model and a linear machine-learning algorithm. This baseline was then combined with other surface classifiers in two ensemble techniques, namely a fixed-rule model and a meta-classifier model. The approach also employed a combination of ensemble features, namely surface features, generic automatic word vectors and affect word vectors. The results of the experiment that the method outperformed the baseline model in terms of the F1 score.

A combination of multiple classifiers with new term weighting has been proposed by (Abdel Fattah, 2015). By utilising six traditional term-weighting schemes as the baseline model, the new term-weighting scheme attempted to provide more informative terms by investigating their occurrence in both class space and document space. In the classification step, the study employed support vector machine, probabilistic neural network and Gaussian mixture model. The combination of multiple classifiers made use of simple voting and Borda count. By using movie reviews and a multi-domain sentiment dataset, the result of the experiment showed that the proposed new term-weighting scheme outperformed traditional approaches.

In another machine-learning approach (Chen et al., 2017), sentence type classification was proven to increase SA performance. In the first step, a target expression from opinionated sentences was extracted by introducing the BiLSTM-CRF model, which is a deep neural sequence model. Target extraction regards sentences as tokens labelled with IOB format (inside, outside, beginning). Lastly, a sentiment classification model called 1d-CNN was described to forecast the polarity of non-target, one-target and multi-target sentences individually. From the evaluation of the proposed method on five different benchmark datasets, namely MR, SST-1, SST-2 and CR, it was concluded that sentence type classification using the BiLSTM-CRF model has potential to increase the results of SA at sentence level.

A dimensionality reduction technique, integrated sentiment analysis (ISA), has been proposed to make SA fast, scalable, efficient and accurate (Ceron et al., 2016). This notion to reduce the dimensionality of the stem so it needs no complete length is derived from the coarsened exact matching algorithm. Experimental analysis on several datasets, including the large movie review dataset, the INVALSI dataset and the Expo2015 dataset, showed that iSA outperformed individual classification employing several other machine-learning techniques, such as SVM and Random Forest.

Since dimensionality reduction potentially helps achieving better classification accuracy, (Yousefpour et al., 2017) presented a new reduced feature vector from a unigram-based and part-of-speech based feature set by proposing an ordinal-based integration of different feature vectors (OIFV). The framework consists of two phases:

1 feature representation

2 feature subset selection employing feature ranking (document frequency, chi-square, information gain, standard deviation, weighted log likelihood ratio), integrating features and wrapping.

Method validation by using two benchmark datasets, movie review and Amazon product reviews, indicated that the part-of-speech pattern performed better compared to the unigram-based feature.

In another work (Khan et al. 2016), SVM was employed in an attempt to enhance machine-learning-based SA using features generated from the general-purpose sentiment lexicon SentiWordNet in a framework called semi-supervised subjective feature weighting and intelligent model selection (SWIMS). After passing on only the subjective features by filtering out objective features, the features are assigned to term#POS pairs. Basically, SWIMS is a feature-weighting mechanism that employs min-max-normalised-SentiMI. The result is multiplied with term presence. SVM light is then used to classify the review data in a scenario of ten-fold cross-validation in which for every fold 90% of the data are used to be training data and the rest is employed as testing data. The fold that achieves best accuracy is selected as the model for the training data. Validation using three kinds of benchmark datasets, namely the large movie review dataset, the Cornell movie review dataset and the multi-domain sentiment dataset, indicated that the proposed framework outperformed other SA techniques.

An improved one versus one (OVO) strategy for multi-class sentiment classification was presented by Liu et al. (2017). OVO is a practical and effective strategy to convert multi-class classification into multiple binary classification. The research step involved:

1 choosing an important bag-of-words document feature by employing the IG algorithm

2 determining the sentiment category of the test text according to the enhanced OVO strategy and the SVM algorithm.

Evaluation and analysis using the movie review dataset indicated that the improved OVO strategy achieved better results than existing multiclass classification techniques.

To capture the problem of domain dependency that usually emerges in SA tasks, Wu et al. (2017) have proposed a unified framework to train domain specific classifiers to be able to perform better in different domains. The idea was to provide general-purpose sentiment knowledge by integrating sentiment information from three sentiment lexicons (SentiWordNet, MPQ and Bing Liu's Lexicon) to obtain a general-purpose sentiment lexicon that is more complete and accurate. A term with a higher positive sentiment score is regarded as positive and a term with a higher negative sentiment score is regarded as negative. All neutral words from SentiWordNet were removed. Unfortunately, polysemic words that convey different sentiment scores were also removed. The proposed method outperformed domain-specific sentiment lexicons as well as decrease the dependence on labelled data.

The previous paragraphs described related studies employing supervised approaches that make use of machine-learning algorithms. Although this seems to work well in sentiment classification tasks, when sufficient trained sentiment-labelled data are difficult to obtain it becomes dubious (Muhammad et al. 2016). To cope with the task of providing training data, which can be labour-intensive, unsupervised approaches of SA tasks based on sentiment lexicons is promising (Medhat et al., 2014). In the following paragraph, we will illustrate related works that employ unsupervised approaches, including contextual studies on unsupervised approaches.

An unsupervised lexicon-based SA approach has been presented by Vilares et al. (2017). The concept of compositional operations was introduced and syntactic information in the form of universal dependencies was exploited. The operation is executed by calculating sentiment orientation of each node in the dependency tree. The prior sentiment value from the sentiment lexicon is used as the initial value of the node. An algorithm to update sentiment orientation was proposed and applied when traversing the parse tree post-order. An evaluation of the proposed method confirmed that it worked better compared to existing unsupervised methods.

An SA based method has been proposed (Liu et al., 2017) to provide a ranking of products by analysing online reviews. Making use of a sentiment dictionary called HowNet, an algorithm was developed to identify sentiment orientation concerning a product feature. Since the product feature itself had been extracted from the review (Automobile Home), the algorithm was only concerned with the extraction of the sentiment orientation by detecting the sentiment words and counting the frequency of both positive and negative sentiment words. A sentence with no sentiment word was regarded as neutral. In the last step, the sentiment value identified was converted into an intuitionistic fuzzy number to build a computer aided computer purchase decision.

A novel unsupervised lexicon-based SA method is described in Fernández-Gavilanes et al. (2016). This work included sentiment lexicon building by adapting a non-supervised method called PolarityRank. From the assumption that traditional lexicon-based schemes are over-simplistic, this approach made use of the concept of sentiment propagation to determine sentiment polarity. First, an undirected graph consisting of a set of nodes and a set of undirected edges is built from a set of texts in a particular domain. For the nodes, adverbs are filtered out because they do not provide any sentiment value. Once the graph has been built, sentiment propagation is initiated. To augment sentiment propagation, this approach identifies negation and intensification. Results of the experiment on the Cornel movie review dataset, the Obama-McCain debate and the SemEval-2015 dataset indicated that the method was competitive and robust.

Sentiment lexicons are an important source for SA, especially for lexico-based SA methods. An approach of developing an adaptive sentiment lexicon has been proposed by Keshavarz and Abadeh (2017) in an attempt to find an optimum sentiment lexicon. A classic genetic algorithm was incorporated to find a lexicon that optimises accuracy and minimises classification errors. A chromosome represents a word and its value represents the sentiment value. A fitness function was formulated to evaluate the chromosome. After certain convergence criteria are met, the iteration to find optimum fitness is stopped. Applying the method on six datasets, the accuracy achieved was higher than 80%.

The present work considered two contextual issues, namely local context and global context. We overviewed both issues in the previous section. Previous studies that took into account the local context issue are Sumanth and Inkpen (2015), Hung and Chen (2016) and Sharma et al. (2017), while Saif et al. (2016) handled global context in a Twitter dataset. Actually, both definitions have been previously adopted by Muhammad et al. (2016), which explains local context as the context that emerges because of the interaction between words in their local neighbourhood, i.e., the sentence. Yet, the work seems to focus on analysing the effect of negation, intensifiers and discourse. For example, the word 'good' implies different polarity than the phrase 'not good'. How to pick the correct sense of a word is not discussed in the work. For taking into account global context, the work enhanced a genre-specific vocabulary utilising distant supervised learning.

Sumanth and Inkpen (2015) assessed how a WSD system positively influences the performance of an SA task on micro-post data. To extract the feature vector representation of the text data from the general-purpose sentiment lexicon of SentiWordNet, a WSD method called Babelfy was utilised. The experiment was conducted using the SemEval-2013 dataset in a supervised environment. Utilising WEKA, the work experimented with the Random Forest machine-learning algorithm. The results indicated an enhancement of performance over the baseline method (unigram feature).

Likewise, Hung and Chen (2016) adopted three WSD techniques to extract feature representation of Word of Mouth (WOM) documents, i.e., 25,000 review documents from Hotels.com. For the ground truth, 1- or 2-star reviews were labelled as negative documents, while 4- and 5-star reviews were labelled as positive documents. Reviews with 3-star were ignored. The experiment was conducted in a supervised environment of SWM and J48 decision tree using WEKA. The work reported an improved SA result.

In another study (Sharma et al., 2017), the effect of neighbouring words was investigated. The WSD algorithm was employed but, from the example presented, rather as a POS tagger than as a method to determine the contextual sense of a word. In the work, to update native scores of individual words, two factors were considered, i.e.,

1 distance of neighbouring word

2 score of neighbouring word.

To optimise the score, a genetic algorithm was employed. The experiment used data collected from Tripadvisor and indicated a small increase in accuracy. This work considered only local context.

Saif et al. (2016) proposed to cope with contextual semantics (global context) in the domain of a Twitter dataset. The introduced a novel method, SentiCircles, which was adopted in the present work. However, how to provide the prior contextual sentiment value before adjustment by SentiCircles is not discussed. In the experiment, three datasets were used, i.e., OMD, HCR and STS-Gold. Several scenarios were investigated. The results confirmed that SentiCircles outperformed two baselines methods, i.e., lexicon labelling method and SentiStrength.

Another work that considered global context is the work of Lek et al. (2014). Domain-sensitive sentiment as defined in this work is the same as global context. Not utilising a sentiment lexicon, the work proposes to first extract all sentiment words from a very large corpus of product reviews in various domains along with their aspects. Since the dataset has pro/con sections, the prior orientation of sentiment words can be assigned in advance. The authors propose a formula to specify the sentiment value of every sentiment word based on the probability of their appearance in pro/cons sections.

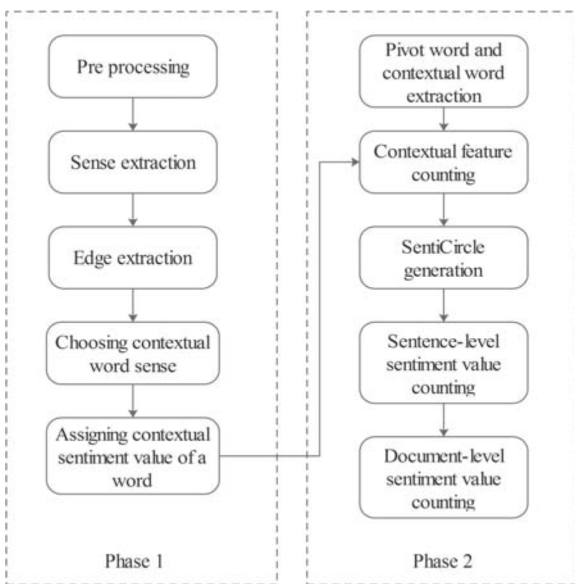
The present work considered both local and global context and also developed an extension of several previous works concerning WSD tasks (Sinha and Mihalcea, 2007) and contextual semantics (Saif et al., 2016) to investigate potential reinforcement of both contextual issues in SA in product reviews.

3 Proposed method

The proposed method consists of two phases:

1. Contextual sentiment value determination to deal with the contextual feature in the form of the contextual prior sentiment value of a word brought by a certain sense of the word contained in a sentence.
2. Sentiment value update to deal with a new sentiment value of the word affected by specific aspects of the review domain.

Figure 1 Proposed method



Both phases consist of several steps. For clarification, all of the steps are presented in Figure 1. The aim of the first phase is to pick the correct sentiment value of words from SentiWordNet. SentiWordNet is based on the WordNet lexical database. Every word in the WordNet database can have multiple senses. In SentiWordNet, every sense has its own sentiment value. Most works that utilised SentiWordNet picked the sentiment value of the first sense or the average value of all senses (Hung and Chen 2016). The second phase deals with global context. The sentiment value that was previously assigned to a word in Phase 1 needs to be adjusted with respect to the domain. Since there are many notations used in this section, we provide a list of notations with their definitions in Table 1.

Table 1 List of notations

Notations	*Definition*
Swup	similarity using Wu and Palmer formula
LCS	least common subsumer of two word senses in WordNet taxonomy
Slch	similarity using Leacock and Chodorow formula
D	maximum depth of WordNet taxonomy
Simres	similarity using Resnik formula
IC	information content of two word senses
Simlin	similarity using Lin formula
Simlin	similarity using Jiang and Conrath formula
S_k	review sentence with index-*k*
w_i	word with index-*i*
ws_i^j	word sense of word *i* with index-*j*
cs_i	prior sentiment value of w_i
$cspos_i$	positive value of cs_i
$csneg_i$	negative value of cs_i
$csneu_i$	neutral value of cs_i
c_i	contextual word that has contextual sentiment value of w_i
x_i	*x* value of c_i in Cartesian coordinates
y_i	*y* value of c_i in Cartesian coordinates
p_k	pivot word of review sentence S_k
wcp_i	selected sense of w_i as candidate pivot word
w_d	first sense of domain word
r_i	feature of c_i
θ_i	feature of c_i
N	number of word in the whole domain of a product review
Nc_i	number of c_i in the whole domain of a product review
UOS_k	sentiment value of review sentence S_k
wS_k	weight of S_k based on its similarity between p_k and *wd*
α	neutral region threshold

3.1 Contextual prior sentiment value calculation

The first phase in this work comprises contextual sentiment value determination. This phase deals with determining the correct sense of a word in a review document. The contextual sentiment value is determined by a certain sense of the word with respect to the context.

The latest version of SentiWordNet is SentiWordNet 3.0, which is the result of automatic annotation of terms contained in WordNet 3.0 (Baccianella and Esuli, 2010). WordNet itself organises its collection in terms of a synonym set (henceforth: synset), a group of words that have the same sense. Every synset corresponds to a certain gloss that represents the sense of a word. Every synset is enclosed in braces and associated with a certain part of speech (POS) of the noun (*n*), verb (*v*), adjective (*adj.*) or adverb (*r*). A word in the WordNet collection can have more than one sense.

In this phase, a graph-based WSD technique (Sinha and Mihalcea, 2007) is adapted and extended to pick the correct sense of a word with respect to its local context in order to pick the correct prior sentiment value. We incorporated the Adapted Lesk similarity algorithm (Banerjee and Pedersen, 2002), which has the highest coverage of POS compared to other similarity algorithms (Wu and Palmer etc.). Adapted Lesk is an extension of the Lesk Algorithm. Unlike Lesk, which restricts its comparison to the glosses of the word being disambiguated, Adapted Lesk provides a richer source of information by comparing the glosses of the word that have a direct relation (in WordNet structure) with the word being disambiguated. It was proven (Rintyarna and Sarno, 2016) that Adapted Lesk improved the performance of this graph-based WSD technique in terms of precision by 19% compared to original Lesk (1986) in an individual similarity experiment for WSD. Both Lesk as used in Sinha and Mihalcea (2007) and Adapted Lesk as used in this work have the highest coverage across POS tag(s) since they do not rely on the graph structure of WordNet but on the gloss of the word being disambiguated. All steps in phase 1 will be described in the following paragraphs.

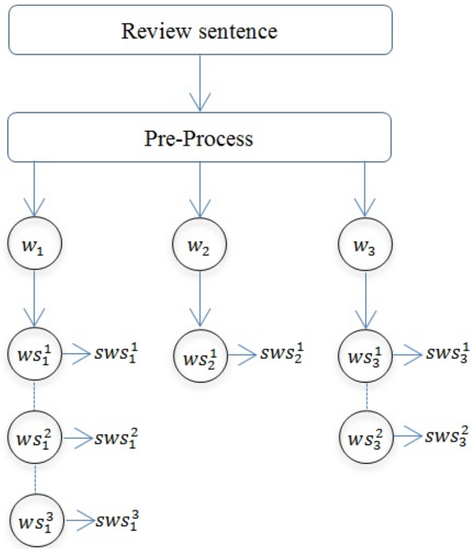
3.1.1 Pre-processing

This step consists of stop word removal, POS tagging and filtering. POS tagging is needed to assign the correct part of the speech of the word to retrieve the suitable gloss of the word from WordNet since WordNet organises its collection with specific speech parts. In this work we used the library of Stanford CoreNLP 3.6.0 (Manning et al., 2014). We also employed filtering to keep all nouns, verbs and adjectives. Later, nouns, verbs and adjectives are treated as candidate vertices of the graph that will be generated for every sentence of the review document.

3.1.2 Sense extraction

After the words have been POS tagged and filtered, their senses are extracted from the WordNet database. Every word sense is treated as a candidate vertex of the graph. Since a graph will be generated for every sentence in the review document, supposing we have a sentence S to disambiguate, consisting of n words after pre-processing (w_1, w_2, w_3, …, w_n), the word senses picked from WordNet Database are represented as ws_i^j, meaning the senses of word *i* with index-*j*. In the next step, the edge between ws_i^j is extracted. Lastly, every ws_i^j is associated with its prior sentiment value s_i^j, which is picked from a general-purpose sentiment lexicon of SentiWordNet. Phase 1 of this research is aimed at choosing the suitable sws_i^j for a word w_i by taking into account its contextual meaning in the sentence (local context). The notation is given in Figure 2.

Figure 2 Example of extracted sense and their corresponding sws_i^j (see online version for colours)



3.1.3 Edge extraction

This step involves counting the similarity measure between concepts (word senses) extracted in the previous step. This step adopts the work of Sinha and Mihalcea (2007). We improved the performance of the method by incorporating the Adapted Lesk similarity algorithm to achieve better results. In the performance matrix evaluation, we employ several similarity algorithms, including Adapted Lesk, in order to evaluate the performance of the graph-based WSD method in extracting local context using different similarity algorithms.

Once the similarity measure between the word senses has been counted, an edge is generated between corresponding nodes (*ws*). A weight is then assigned to this edge according to its similarity value. As was done by Sinha and Mihalcea (2007), in this step we employ several similarity measure algorithms, like those used by Sinha and Mihalcea (2007), i.e., Wu and Palmer, Leacock and Chodorow, Resnik, Lin, Jiang and Conrath and Adapted Lesk (Banerjee and Pedersen, 2002). We incorporated Adapted Lesk (Banerjee and Pedersen, 2002) to achieve better results. Adapted Lesk is an extension of the Lesk algorithm that provides a richer hierarchy of semantics using the glosses of the WordNet lexical database. In the process of disambiguation, Adapted Lesk expands the comparison of the words being disambiguated to words that are not directly connected with them. Adapted Lesk utilises all possible relations between nouns, verbs and adjectives in the WordNet taxonomy. In the implementation, we adapted WordNet Similarity for Java (WS4J) from Hideki Shima to calculate the similarity algorithm.

Wu and Palmer (1994) introduced a function called least common subsumer (LCS) to measure the similarity between two concepts. In this study, both concepts are word senses. Given two word senses, ws_a^c and ws_b^d, the similarity between them, *Swup*, can be computed using equation (1) is the nearest ancestor of both word senses in the WordNet taxonomy.

$$Swup = \frac{2 * Depth(LCS)}{Depth\left(ws_a^c\right) + Depth\left(ws_b^d\right)} \tag{1}$$

Utilising Leacock and Chodorow's (Leacock et al., 1998), the similarity between two word senses can be calculated using equation (2). In the equation, $l(ws_a^c, ws_b^d)$ is the length of the shortest path of both word senses, while D is the maximum depth of the WordNet taxonomy.

$$Slch = -\log \frac{l\left(ws_a^c, ws_b^d\right)}{D} \tag{2}$$

Resnik (1995) defines similarity as a function of the information content, *IC*, of *LCS* of both word senses, as shown in equations (3) and (4). In equation (4), *it*, *if* and *N* are term frequency, inherited frequency and total number of terms in the taxonomy of WordNet respectively, while *c* indicates the *LCS* of both senses.

$$Simres\left(ws_a^c, ws_b^d\right) = IC\left(LCS\left(ws_a^c, ws_b^d\right)\right) \tag{3}$$

where

$$IC(c) = -\log \frac{tf(c) + if(c)}{N} \tag{4}$$

Using a similar formula as Wu and Palmer's, Lin (1998) introduced similarity between senses as a function of information content, as presented in equation (5).

$$Simlin = \frac{2 * IC\left(LCS(c)\right)}{IC\left(ws_a^c\right) + IC\left(ws_b^d\right)}. \tag{5}$$

Likewise, Jiang and Conrath (1997) define similarity as a function of information content, as shown in equation (6).

$$Simjcn = \frac{1}{IC\left(ws_a^c\right) + IC\left(ws_b^d\right) - 2 * IC(LCS(c))} \tag{6}$$

For a simple illustration of this step, suppose that after the filtering step 3 words (w_1, w_2 and w_3) remain, where w_1 has three senses (ws_1^1, ws_1^2, ws_1^3), w_2 has one sense (ws_2^1) and w_3 has two senses (ws_3^1, ws_3^2). These word senses become a vertice of the graph. An edge is then generated between the word senses. The weight of the edge $e_{a,b}^{c,d}$ is assigned by counting the similarity between word senses ws_a^c and ws_b^d by employing the WordNet Similarity Algorithm, as previously explained. The last step is to count the indegree score by employing the indegree Algorithm. In an undirected graph, we can define the indegree as the number of edge ends adjacent to a vertex. The sense chosen as the contextual word sense is the sense with the highest indegree score [equation (10)]. Since edges are generated between all word senses (vertex), we can represent the value of their weight

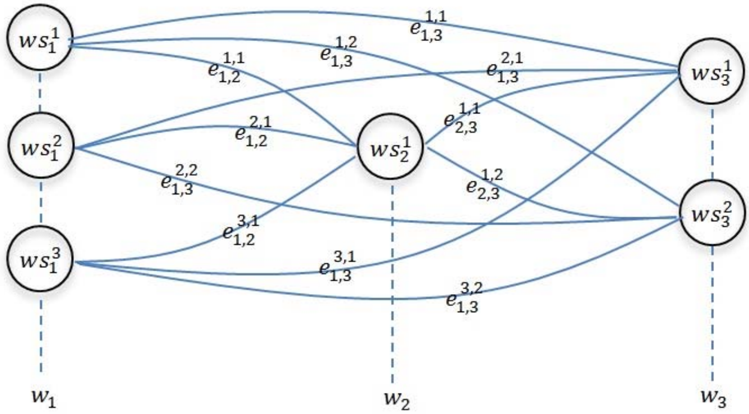
$e_{a,b}^{c,d}$ as the mean weight between ws_a^c and ws_b^d. A simple illustration of this is presented in Figure 3. From the figure, we can obtain the indegree score of ws_2^1, i.e., $In(ws_2^1) = e_{2,3}^{1,1} + e_{2,3}^{1,2} + e_{1,2}^{1,1} + e_{1,2}^{2,1} + e_{1,2}^{3,1}$. For counting the indegree score, we use Algorithm 1. The contextual word sense is the word sense with the highest indegree score.

Algorithm 1 Indegree score calculation

```
Input:
int N = sentence.length; /*Length of the sentence after pre-processed*/
string[] w;
string w[i] = word#pos#senseid //pos : part of speech
string ws_row,ws_ column, e;
Output:
score //Indegree Score
Process
for (int row = 0; ws_b < w.length – 1; row++) {
ws_row = w[row];
for (int column = row + 1; column < w.length; column++) {
ws_column = w[column];
e = sim (ws_row, ws_column); //calculate indegree score
}
}
function sim(string wsa, string wsb){
length_senseid_wsa = wsa.split(“#”);
length_senseid_wsb = wsb.split(“#”);
score = “”;
for(int a=1; a<= length_senseid_wsa[2]; a++){
for(int b=1; b<= length_senseid_wsb[2]; b++){
score = (length_senseid_wsa[a], length_senseid_wsb[b]); //calculate similarity
}
score.= score;
}
return score;
}
```

Before counting the indegree score, we normalise the weight of the edges. The node with the highest indegree score is then selected and the sense it corresponds to is assigned to the word as the sense of the word in the particular context of the sentence (local context). The prior sentiment value corresponding to this sense is then picked from SentiWordNet and assigned as the prior sentiment value (cs_i) of contextual sentiment value c_i. The previously explained step is illustrated in Figure 3.

Figure 3 Edge extraction (see online version for colours)



Source: Adapting the work of Sinha and Mihalcea (2007)

3.1.4 Assigning contextual sentiment value

The last step in Phase 1 is assigning the contextual prior sentiment value (cs_i) of word w_i, where w_i is the highest indegree score of w_i and $f(m_i)$ is a function to retrieve three sentiment values from SentiWordNet, namely *cspos*, *csneg* and *csneu*, as shown in equation (7). In equation (8), $In\left(ws_i^j\right)$ is the indegree score of ws_i^j computed using Algorithm 1, where ws_i^j are the word senses of w_i with index j, where n is the number of word senses of w_i.

$$cs_i = f\left(m_i\right) \tag{7}$$

where

$$m_i = \arg\max{}_j \; In\left(ws_i^j\right)_{j=1,2,\ldots n} \tag{8}$$

Once all contextual prior sentiment values have been calculated, the sentiment value at sentence level is computed and denoted as *Spos*, *Sneg* and *Sneu*. If n is the number of words in a sentence, every sentence will have three sentiment values, computed as:

$$Spos = \sum\nolimits_{i=1}^{n} cspos_i \tag{9}$$

$$Sneg = \sum\nolimits_{i=1}^{n} csneg_i \tag{10}$$

$$Sneu = \sum\nolimits_{i=1}^{n} csneu_i \tag{11}$$

In this work, we ignore *Sneu* and apply Rule (12) to determine the orientation of the sentiment at sentence level (OS_k) for the local context experiment. To test our proposed method for SA to capture local context, in the performance evaluation step we present SWSD-WU, SWSD-LEACOCK, SWSD-LIN, SWSD-JIANG and SWSD-ADT (WU, LEACOCK, LIN, JIANG and ADT point as the similarity measures employed for extracting local context).

$$OS_k = \begin{cases} positive\ if\ |Spos| > |Sneg| \\ negative\ if\ |Sneg| > |Snog| \\ neutral\ if\ Spos| = Sneg| \end{cases} \tag{12}$$

3.2 Sentiment value update (phase 2)

This phase is based on the SentiCircle Median approach (Saif et al. 2016). We extended the approach by providing a similarity-based method to extract pivot word from a review sentence to select certain aspects that tune the orientation of the sentiment value of contextual words (c_i) in a specific review document domain. We also consider the local context (see the previous phase) based on the weighted graph-based WSD approach (Sinha and Mihalcea 2007) to adjust the prior sentiment value for contextual sentiment value c_i.

3.2.1 Pivot word and contextual word extraction

The pivot word p_k is a word with the POS tag 'noun' that represents the effect of the specific domain on the contextual word sense in a review sentence represented by SentiCircles. So, rather than using all words from a sentence in the review document with part of speech (common noun, proper noun, pronoun) like was done in Saif et al. (2016), we used all words with the POS tag 'noun' taken from a sentence in the review document and propose to select one of them by counting the similarity between the first sense of domain word and the selected sense of candidate pivot word wcp_i using Wu and Palmer's algorithm, i.e., equation (14). We choose candidate pivot word cp with maximum similarity as pivot word p. Suppose the first sense of the domain word is wd and the selected sense of the candidate pivot word with index-i is wcp_i, the pivot word is determined by wcp_i with maximum similarity as expressed in equation (13). We label the other words in the review sentence with POS tags 'verb', 'adjective' or 'noun' as contextual words. LCS is the least common subsumer of wd and wcp_i in the WordNet taxonomy.

$$p_k = \arg\max_{wcp_i} Sim\left(wd, wcp_i\right)_{i=1,2,\ldots n} \tag{13}$$

where

$$Sim\left(wd, wcp_i\right) = \frac{2 * Depth(LCS)}{Depth(wd) + Depth\left(wcp_i\right)} \tag{14}$$

3.2.2 Contextual feature counting

To update contextual prior sentiment value cs_i with respect to a specific domain, we count contextual feature r_i and θ_i to generate a WSD based on SentiCircle. We count r_i, which represents word degree of correlation (WDOC) between pivot word p and contextual word c_i. Like Saif et al. (2016), we compute WDOC using equation (15).

$$r_i = WDOC(p_k, c_i) = f(c_i, p_k)\log\frac{N}{Nc_i} \tag{15}$$

Since we work on a review dataset that contains multiple sentences in a single review document, we compute $f(c_i, p)$ in a sentence-based analysis across the review document. As we create a SentiCircle to represent the global context, we compute N as the total number of words and N_{c_i} as the total number of c_i. Meanwhile, for the angle of adjustment θ_i, we extend the SentiCircle by providing a graph-based approach (see the previous phase) to extract contextual prior sentiment value cs_i of contextual term c_i based on its local context. The formula can be seen in equation (16). We provide Rule 11 to determine cs_i.

$$\theta_i = cs_i * \pi rad \tag{16}$$

where

$$cs_i = \begin{cases} cspos_i \ \ if \ |cspos_i| > |csneg_i| \\ csneg_i \ \ if \ |csneg_i| > |cspos_i| \end{cases} \tag{17}$$

3.2.3 SentiCircle generation

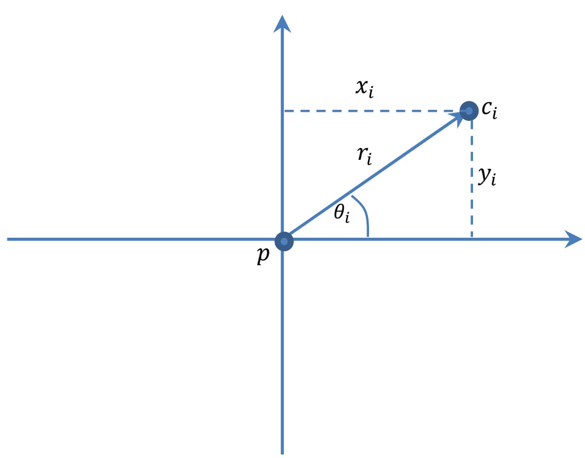
For generating the SentiCircle to adjust the sentiment value of a word with respect to a specific domain, we count x_i and y_i for every contextual term c_i extracted from a sentence in the review document using the simple trigonometric formulas in equation (18) and equation (19):

$$x_i = r_i \cos\theta_i \tag{18}$$

$$y_i = r_i \sin\theta_i \tag{19}$$

Every contextual term c_i is represented by its x_i and y_i value in the Cartesian coordinate system as shown in Figure 4.

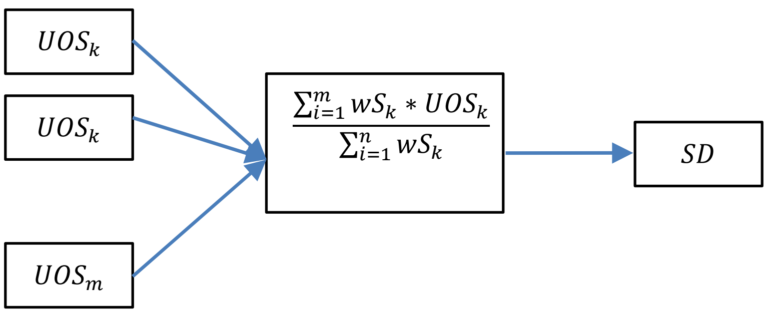
Figure 4 Senticircle representation in Cartesian coordinate system (see online version for colours)



3.2.4 Sentiment identification of a review document

The last step of phase 2 is identifying the sentiment orientation of the review document (at the document level). Different from the method by Saif et al. (2016), which handles single sentences from a Twitter dataset, a review document commonly consists of more than one single sentence. Suppose a review document contains m review sentences $S(S_1, S_2, S_3, \ldots, S_m)$, we generate a SentiCircle for every sentence S_k and we determine its updated sentiment value UOS_k. Following Saif et al. (2016), UOS_k is computed by calculating the senti-median of c_i. To calculate the result on a document level, we need to aggregate the results of the sentence level into a document level result. We propose a weighted-mean based formula to count the results of the SA at the document level. This strategy is described in Figure 5.

Figure 5 Document level sentiment identification strategy (see online version for colours)



We view pivot term p_k as the term that represents the influence of the domain on contextual term c_i in terms of its contextual sentiment value. Every sentence contains one pivot term. To take into account the effect of different pivot terms on the overall sentiment value at a document level, we compute SD by averaging UO_k weighed $\left(w_{s_k}\right)$ by the similarity of its pivot term with the domain counted by using Wu and Palmer's algorithm, as expressed in equation (20). Given that the similarity of the pivot term of sentence S_k is Sim_{s_k}, we propose to compute the weight and the sentiment value of the document with equations (21) and (22).

$$SD = \frac{\sum_{i=1}^{n} ws_i * UOS_k}{\sum_{i=1}^{n} wS_k} \tag{20}$$

where

$$wS_k = \frac{1 - Sim_{S_k}}{\sum_{i=1}^{n} Sim_{S_k}} \tag{21}$$

$$Sim_{S_k} = \frac{2 * Depth(LCS)}{Depth(wd) + Depth(p_k)} \tag{22}$$

with LCS is the least common subsumer of wd and p_k. Lastly, the sentiment orientation of the document is determined by using Rule (23), where α is the neutral region threshold.

$$OD = \begin{cases} positive\ if\ \ SD > \alpha \\ negatuve\ if\ \ SD < \alpha \\ neutral\ if\ \ \alpha - \leq sd \leq \alpha \end{cases} \tag{23}$$

To represent our proposed method, which handles both local and global context, we present SSENTI-WU, SSENTI-LEACOCK, SSENTI-LIN, SSENTI-JIANG and SSENTI-ADT (WU, LEACOCK, LIN, JIANG and ADT point as the similarity measures employed in extracting local context).

3.3 Experimental design

An experiment was conducted to classify review documents using the proposed method into one of three categories, i.e., positive documents, negative documents and neutral documents. For the evaluation, we calculated the matrix of precision, recall, F-measure and accuracy as shown in equations (24) to (27). As the baseline method we used SentiCircle Median (Saif et al., 2016).

$$Prec = \frac{t_p}{t_p + f_p} \tag{24}$$

$$Rec = \frac{tp}{tp + fn} \tag{25}$$

$$Fmeans = 2.\frac{Prec.Rec}{Prec + Rec} \tag{26}$$

$$Acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{27}$$

Figure 6 Excerpt from the Amazon review dataset collected


```
{
"reviewerID": "AWJ0WZQYMYFQ4",
"asin": "120401325X",
"reviewerName": "JM",
"helpful": [4, 4],
"reviewText": "Item arrived in great time and was in
perfect condition. However, I ordered these buttons
because they were a great deal and included a FREE
screen protector. I never received one. Though its not
a big deal, it would've been nice to get it since they
claim it comes with one.",
"overall": 4.0,
"summary": "Cute!",
"unixReviewTime": 1382313600, "reviewTime": "10 21,
2013"
}
```


Source: McAuley et al (2015)

We used Amazon review data provided by McAuley et al. (2015) in the experiment. This dataset contains Amazon online product reviews spanning from May 1997 to July 2014. An excerpt of the dataset and its metadata is shown in Figure 6. Five domains of the dataset were employed, i.e.,

1 automotive
2 beauty
3 books
4 electronics
5 movies.

To provide the ground truth, following Fang and Zhan (2015), the online product reviews with an overall score (rating) of 4–5 were labelled as positive. We labelled the online product reviews with an overall score (rating) of 1–2 as negative documents while the rest was labelled as neutral.

4 Results and discussions

Evaluation was conducted with several datasets from several product review domains in order to validate the result of the proposed method. The proposed method was then compared with the state of the art in contextual SA, i.e., the Senti-Median Method, as the baseline method in terms of performance measure as well as the only other method that considers local context.

4.1 Results

The graph-based WSD approach to extract local context was implemented with several similarity algorithms, as explained in Chapter 3.1. The results in comparison with the baseline method are presented for every similarity algorithm. The performance evaluation measures used in this work were precision, recall, F-measure and accuracy.

4.2 Evaluation results of proposed method considering local context

In the following subsection, we show the evaluation results of the proposed approach only considering local context of review documents. Every table represents the performance of the proposed method in one review dataset domain. The average performance increase over all dataset domains is shown in Figure 7.

Table 2 Local context performance for book dataset

	BASELINE	*SWSD-WU*	*SWSD-LEACOCK*	*SWSD-RES*	*SWSD-LIN*	*SWSD-JIANG*	*SWSD-ADT*
PREC	0.28	0.34	0.39	0.32	0.31	0.33	0.31
REC	0.28	0.46	0.52	0.36	0.34	0.40	0.35
FMEAS	0.28	0.39	0.45	0.34	0.32	0.36	0.33
ACC	0.56	0.76	0.85	0.76	0.72	0.82	0.74

Table 3 Local context performance for beauty dataset

	BASELINE	*SWSD-WU*	*SWSD-LEACOCK*	*SWSD-RES*	*SWSD-LIN*	*SWSD-JIANG*	*SWSD-ADT*
PREC	0.24	0.30	0.31	0.30	0.31	0.33	0.31
REC	0.27	0.40	0.38	0.38	0.41	0.40	0.40
FMEAS	0.25	0.34	0.34	0.34	0.35	0.36	0.35
ACC	0.62	0.68	0.70	0.69	0.68	0.74	0.69

Table 4 Local context performance for automotive dataset

	BASELINE	*SWSD-WU*	*SWSD-LEACOCK*	*SWSD-RES*	*SWSD-LIN*	*SWSD-JIANG*	*SWSD-ADT*
PREC	0.34	0.31	0.31	0.31	0.31	0.31	0.31
REC	0.38	0.39	0.40	0.37	0.37	0.40	0.40
FMEAS	0.36	0.35	0.35	0.34	0.34	0.35	0.35
ACC	0.71	0.74	0.76	0.71	0.72	0.77	0.76

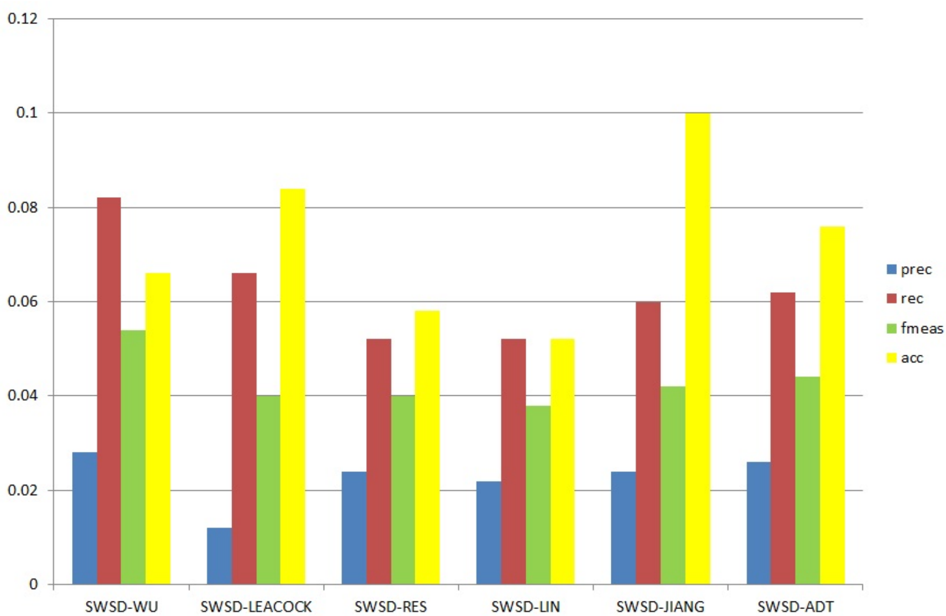
Table 5 Local context performance for movie dataset

	BASELINE	*SWSD-WU*	*SWSD-LEACOCK*	*SWSD-RES*	*SWSD-LIN*	*SWSD-JIANG*	*SWSD-ADT*
PREC	0.39	0.39	0.28	0.44	0.40	0.39	0.39
REC	0.38	0.41	0.31	0.46	0.41	0.39	0.40
FMEAS	0.38	0.40	0.30	0.45	0.41	0.39	0.39
ACC	0.62	0.68	0.64	0.70	0.68	0.71	0.70

Table 6 Local context performance for electronics dataset

	BASELINE	*SWSD-WU*	*SWSD-LEACOCK*	*SWSD-RES*	*SWSD-LIN*	*SWSD-JIANG*	*SWSD-ADT*
PREC	0.35	0.40	0.37	0.35	0.38	0.36	0.41
REC	0.34	0.40	0.37	0.34	0.38	0.36	0.41
FMEAS	0.34	0.4	0.37	0.34	0.38	0.36	0.41
ACC	0.67	0.65	0.65	0.61	0.64	0.64	0.67

Figure 7 Average increase of local context performance over all dataset domains (see online version for colours)



4.3 Evaluation results of proposed method considering both local and global context

In the following tables (Tables 7 to 11) we present the performance evaluation matrixes of the proposed method considering both local and global context. The average performance increase of the proposed method compared to the baseline method on the overall dataset domain is given in Figure 8.

Table 7 Local and global context performance on book dataset

	BASELINE	SSENTI-WU	SSENTI-LEACOCK	SSENTI-RES	SSENTI-LIN	SSENTI-JIANG	SSENTI-ADT
PREC	0.28	0.45	0.31	0.26	0.28	0.29	0.28
REC	0.28	0.41	0.33	0.30	0.33	0.33	0.33
FMEAS	0.28	0.43	0.32	0.29	0.30	0.30	0.30
ACC	0.56	0.89	0.64	0.64	0.89	0.89	0.89

Table 8 Local and global context performance on beauty dataset

	BASELINE	SSENTI-WU	SSENTI-LEACOCK	SSENTI-RES	SSENTI-LIN	SSENTI-JIANG	SSENTI-ADT
PREC	0.24	0.26	0.26	0.37	0.32	0.21	0.23
REC	0.27	0.33	0.30	0.36	0.34	0.33	0.32
FMEAS	0.25	0.29	0.28	0.36	0.33	0.27	0.27
ACC	0.62	0.73	0.61	0.67	0.75	0.75	0.74

Table 9 Local and global context performance on beauty dataset

	BASELINE	SSENTI-WU	SSENTI-LEACOCK	SSENTI-RES	SSENTI-LIN	SSENTI-JIANG	SSENTI-ADT
PREC	0.34	0.35	0.34	0.34	0.34	0.34	0.30
REC	0.38	0.36	0.37	0.32	0.35	0.47	0.30
FMEAS	0.36	0.35	0.35	0.33	0.34	0.39	0.30
ACC	0.71	0.86	0.75	0.73	0.89	0.89	0.87

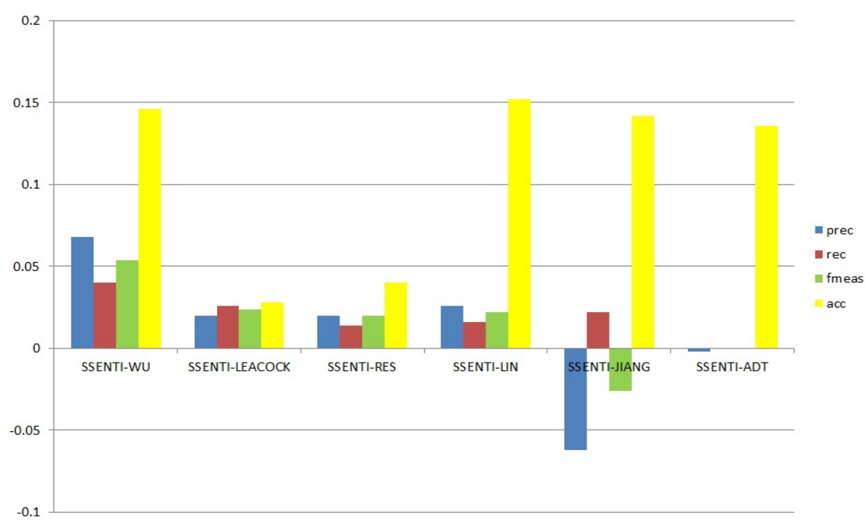
Table 10 Local and global context performance on movie dataset

	BASELINE	SSENTI-WU	SSENTI-LEACOCK	SSENTI-RES	SSENTI-LIN	SSENTI-JIANG	SSENTI-ADT
PREC	0.39	0.42	0.35	0.32	0.44	0.25	0.39
REC	0.38	0.37	0.36	0.33	0.37	0.30	0.35
FMEAS	0.38	0.39	0.35	0.32	0.41	0.27	0.37
ACC	0.62	0.72	0.62	0.63	0.71	0.66	0.68

Table 11 Local and global context performance on electronics dataset

	BASELINE	SSENTI-WU	SSENTI-LEACOCK	SSENTI-RES	SSENTI-LIN	SSENTI-JIANG	SSENTI-ADT
PREC	0.35	0.46	0.44	0.41	0.35	0.20	0.39
REC	0.34	0.38	0.42	0.41	0.34	0.33	0.35
FMEAS	0.34	0.42	0.43	0.41	0.34	0.25	0.37
ACC	0.67	0.71	0.70	0.71	0.70	0.70	0.68

Figure 8 Average increase of both local and global context performance over all dataset domains (see online version for colours)



4.4 Analysis and discussions

We summarise the performance results as follows:

1. The proposed approach increased almost all performance measures (precision, recall, F-measure and accuracy) compared with the baseline method.
2. For local context the best performance was achieved by SWSD-WU with a 2.8% increase for precision, an 8.2% increase for recall and a 5.5% increase for F-measure. Meanwhile, SWSD-JIANG worked best for accuracy, indicated by a 10% increase compared with the baseline method (Figure 7).
3. For both local and global context, the best algorithm was SSENTI-WU, indicated by a 6.8% rise for precision, a 4% rise for recall and a 5.4% rise for F-measure compared with the baseline method. Meanwhile, SSENTI-LIN worked best for accuracy, denoted by a 15.2% increase compared with the baseline method (Figure 8).
4. SSENTI-JIANG experienced a precision decrease by 6.2% and F-measure by 2.6%, while SSENTI-ADT decreased precision by 0.2%.
5. The performance evaluation of the proposed method was conducted on five different product review domains from Amazon collected by Julian McAuley. The results may be different for other domains.
6. Adapted Lesk (SWSD-ADT) obtained the best performance. It achieved a 2.6, 6.2, 4.4 and 7.6 increase of precision, recall, F-measure and accuracy, respectively, in local context evaluation. Unfortunately, for both local and global context (SSENTI-ADT), the performance increase for precision, recall, F-measure and accuracy was –0.2%, -2.1×10^{-15}, -1.1×10^{-15} and 13.6, respectively.

5 Conclusions

Local and global context in review documents may influence the results of SA tasks. Capturing both contextual issues is important to increase the performance of lexicon-based SA. In this paper, we proposed an approach based on a graph-based WSD and the Senti-Median Method to deal with both contextual aspects. The results of the performance evaluation for five review domains led us to conclude that considering both local and global context in unsupervised lexicon-based SA tasks at the document level of product reviews increased the results for precision, recall, F-measure and accuracy.

To get better results for precision, recall F-measure and accuracy, future work will focus on extracting the sentiment value at the aspect level while still considering local and global context as well as proposing an approach to extract aspects and aggregating the results at the document level.

Acknowledgements

The authors would like to thank Institut Teknologi Sepuluh Nopember as well as Universitas Muhammadiyah Jember for supporting this work.

References

Abdel Fattah, M. (2015) 'New term weighting schemes with combination of multiple classifiers for sentiment analysis', *Neurocomputing*, Vol. 167, pp.434–442 [online] http://dx.doi.org/10.1016/j.neucom.2015.04.051.

Araque, O. et al. (2017) 'Enhancing deep learning sentiment analysis with ensemble techniques in social applications', *Expert Systems with Applications*, Vol. 77, pp.236–246 [online] http://dx.doi.org/10.1016/j.eswa.2017.02.002.

Baccianella, S. and Esuli, F.S. (2010) 'SentiwordNet 3.0 : an enhanced lexical resource for sentiment analysis and opinion mining', *Proceedings of the 9th Conference on Language Resources and Evaluation*, Vol. 9, pp.2200–2204.

Banerjee, S. and Pedersen, T. (2002) 'An Adapted lesk algorithm for word sense disambiguation using wordnet', *Computational Linguistics and Intelligent Text Processing*, Vol. 2276, pp.136–145 [online] http://dx.doi.org/10.1007/3-540-45715-1_11.

Ceron, A., Curini, L. and Iacus, S.M. (2016) 'ISA: a fast, scalable and accurate algorithm for sentiment analysis of social media content', *Information Sciences*, Vol. 367–368, pp.105–124 [online] http://dx.doi.org/10.1016/j.ins.2016.05.052.

Chen, T. et al. (2017) 'Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN', *Expert Systems with Applications*, Vol. 72, pp.221–230 [online] http://dx.doi.org/10.1016/j.eswa.2016.10.065.

Coutinho, D.P. and Figueiredo, M.A.T. (2015) 'Text classification using compression-based dissimilarity measure', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 29, No. 5, pp.1–19.

Cui, Z., Shi, X. and Chen, Y. (2016) 'Neurocomputing Sentiment analysis via integrating distributed representations of variable-length word sequence', *Neurocomputing*, Vol. 187, pp.126–132 [online] http://dx.doi.org/10.1016/j.neucom.2015.07.129.

Dat, N.D. et al. (2017) 'Sting algorithm used english sentiment classification in a parallel environment', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 31, No. 7, pp.1–30.

Fang, X. and Zhan, J. (2015) 'Sentiment analysis using product review data', *Journal of Big Data* [online] http://dx.doi.org/10.1186/s40537-015-0015-2.

Fernández-Gavilanes, M. et al. (2016) 'Unsupervised method for sentiment analysis in online texts', *Expert Systems with Applications*, 1 October, Vol. 58, pp.57–75.

Giatsoglou, M. et al. (2016) 'Sentiment analysis leveraging emotions and word embeddings', *Expert Systems with Applications*, Vol. 69, pp.214–224 [online] http://linkinghub.elsevier.com/retrieve/pii/S095741741630584X.

Hung, C. and Chen, S-J. (2016) 'Word sense disambiguation based sentiment lexicons for sentiment classification', *Knowledge-Based Systems*, Vol. 0, pp.1–9 [online] http://linkinghub.elsevier.com/retrieve/pii/S0950705116302453.

Jiang, J.J. and Conrath, D.W. (1997) 'Semantic similarity based on corpus statistics and lexical taxonomy', *Proceedings of International Conference Research on Computational Linguistics, (Rocling X)*, pp.19–33 [online] http://arxiv.org/abs/cmp-lg/9709008 (accessed 12 July 2017).

Keshavarz, H. and Abadeh, M.S. (2017) *Knowledge-Based Systems ALGA: Adaptive Lexicon Learning using Genetic Algorithm for Sentiment Analysis of Microblogs*, 15 April, Vol. 122, pp.1–16.

Khan, F.H., Qamar, U. and Bashir, S. (2016) 'SentiMI : Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection', *Applied Soft Computing Journal*, Vol. 39, pp.140–153 [online] http://dx.doi.org/10.1016/j.asoc.2015.11.016.

Leacock, C., Miller, G.A and Chodorow, M. (1998) 'Using corpus statistics and WordNet relations for sense identification', *Computational Linguistics*, Vol. 24, No. 1, pp.147–165 [online] http//:papers2://publication/uuid/F8B8B5B7-CEDF-44B0-A548-FD5849FDAB85 (accessed 12 July 2017).

Lek, H.H., Chiang, D. and Poo, C. (2014) 'Automatic generation of an aspect and domain sensitive sentiment lexicon', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 23, No. 4, pp.1–21.

Lesk, M. (1986) 'Automatic sense disambiguation using machine readable dictionaries', *Proceedings of the 5th Annual International Conference on Systems Documentation – SIGDOC '86*, pp.24–26.

Lin, D. (1998) 'An information-theoretic definition of similarity', *Proceedings of ICML*, pp.296–304.

Liu, Y., Bi, J-W. and Fan, Z-P. (2017) 'A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm', *Information Sciences* [online] http://linkinghub.elsevier.com/retrieve/pii/S002002551631043X (accessed 12 July 2017).

Manning, C.D. et al. (2014) 'The stanford CoreNLP natural language processing toolkit', in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pp. 55–60 [online] http://aclweb.org/anthology/P14-5010 (accessed 15 November 2016).

McAuley, J., Pandey, R. and Leskovec, J. (2015) 'Inferring networks of substitutable and complementary products', *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785–794 [online] http://arxiv.org/abs/1506.08839.

Medhat, W., Hassan, A. and Korashy, H. (2014) 'Sentiment analysis algorithms and applications : a survey', *Ain Shams Engineering Journal*, Vol. 5, No. 4, pp.1093–1113 [online] http://dx.doi.org/10.1016/j.asej.2014.04.011.

Muhammad, A., Wiratunga, N. and Lothian, R. (2016) 'Contextual sentiment analysis for social media genres', *Knowledge-Based Systems*, Vol. 108, pp.92–101 [online] http://dx.doi.org/10.1016/j.knosys.2016.05.032.

Onan, A., Korukoğlu, S. and Bulut, H. (2016) 'A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification', *Expert Systems with Applications*, 15 November, Vol. 62, pp.1–16.

Pamungkas, E.W., Sarno, R. and Munif, A. (2017) 'B-BabelNet: business-specific lexical database for improving semantic analysis of business process models', *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, Vol. 15, No. 1, pp.07–414.

Pratama, B.Y. and Sarno, R. (2015) 'Personality classification based on twitter text using naive Bayes KNN and SVM', *2015 International Conference on Data and Software Engineering (ICoDSE)*, pp.170–174.

Resnik, P. (1995) 'Using information content to evaluate semantic similarity in a taxonomy', *Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI'95*, Vol. 1, No. 1, p.6 [online] http://arxiv.org/abs/cmp-lg/9511007 (accessed 16 May 2017).

Rintyarna, B.S. and Sarno, R. (2016) 'Adapted weighted graph for word sense disambiguation', in *4th International Conference on Information and Communication Technology (ICoICT)*, pp.60–64.

Saif, H. et al. (2016) 'Contextual semantics for sentiment analysis of Twitter', *Information Processing and Management*, Vol. 52, No. 1, pp.5–19 [online] http://dx.doi.org/10.1016/j.ipm.2015.01.005.

Sarno, R., Munawar, M.N. and Nugraha, B.T. (2016) 'Real-time electroencephalography-based emotion recognition system', *International Review on Computers and Software*, Vol. 11, No. 5, pp.456–465.

Sharma, S. et al. (2017) 'A context-based algorithm for sentiment analysis', *International Journal of Computational Vision and Robotics*, Vol. 7, No. 5, pp.558–573.

Sinha, R. and Mihalcea, R. (2007) 'Unsupervised graph-basedword sense disambiguation using measures of word semantic similarity', International *Conference on Semantic Computing (ICSC'07)*, pp.363–369.

Sumanth, C. and Inkpen, D. (2015) 'How much does word sense disambiguation help in sentiment analysis of micropost data?', *6Th Workshop on Computational Approaches To Subjectivity, Sentiment and Social Media Analysis (Wassa)*, p.115.

Vilares, D., Gómez-rodríguez, C. and Alonso, M.A. (2017) 'Knowledge-based systems universal, unsupervised (rule-based), uncovered sentiment analysis', *Knowledge-Based Systems*, Vol. 118, pp.45–55 [online] http://dx.doi.org/10.1016/j.knosys.2016.11.014.

Wu, F., Huang, Y. and Yuan, Z. (2017) 'Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources', *Information Fusion*, Vol. 35, pp.26–37 [online] http://dx.doi.org/10.1016/j.inffus.2016.09.001.

Wu, Z. and Palmer, M. (1994) 'Verb semantics and lexical', *Proceeding of the 32nd Annual Meeting of The Association for Computational Linguistics*, pp.133–138.

Yazdani, S.F. et al. (2017) 'Sentiment classification of financial news', *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 31, No. 3, pp.1–34.

Yousefpour, A. et al. (2017) 'Ordinal-based and frequency-based integration of feature selection methods for sentiment analysis', *Expert Systems with Applications*, 1 June, Vol. 75, pp.80–93.