AUTHOR(S): MUHAMMAD, A., WIRATUNGA, N. and LOTHIAN, R.

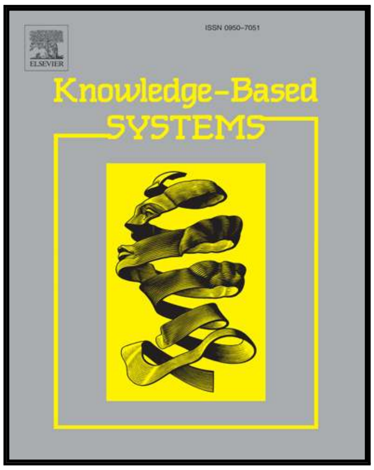
TITLE: Contextual sentiment analysis for social media genres.

YEAR: 2016

Publisher citation: MUHAMMAD, A., WIRATUNGA, N. and LOTHIAN, R. 2016. Contextual sentiment analysis for social media genres. Knowldege-based systems [online], 108, pages 92-101. Available from: https://dx.doi.org/10.1016/j.knosys.2016.05.032

OpenAIR citation: MUHAMMAD, A., WIRATUNGA, N. and LOTHIAN, R. 2016. Contextual sentiment analysis for social media genres. Knowledge-based systems, 108, pages 92-101. Held on OpenAIR [online]. Available from: https://openair.rgu.ac.uk

Accepted Manuscript



Contextual Sentiment Analysis for Social Media Genres

Aminu Muhammad, Nirmalie Wiratunga, Robert Lothian

PII: S0950-7051(16)30114-9
DOI: 10.1016/j.knosys.2016.05.032
Reference: KNOSYS 3529

To appear in: *Knowledge-Based Systems*

Received date: 15 November 2015
Revised date: 12 May 2016
Accepted date: 13 May 2016

Please cite this article as: Aminu Muhammad, Nirmalie Wiratunga, Robert Lothian, Contextual Sentiment Analysis for Social Media Genres, *Knowledge-Based Systems* (2016), doi: 10.1016/j.knosys.2016.05.032

Contextual Sentiment Analysis for Social Media Genres

Aminu Muhammad, Nirmalie Wiratunga and Robert Lothian

School of Computing Science and Digital Media,
Robert Gordon University, Aberdeen, Scotland

Abstract

The lexicon-based approaches to opinion mining involve the extraction of term polarities from sentiment lexicons and the aggregation of such scores to predict the overall sentiment of a piece of text. It is typically preferred where sentiment labelled data is difficult to obtain or algorithm robustness across different domains is essential. A major challenge for this approach is accounting for the semantic gap between prior polarities of terms captured by a lexicon and the terms' polarities in a specific context (contextual polarity). This is further exacerbated by the fact that a term's contextual polarity also depends on domains or genres in which it appears. In this paper, we introduce SmartSA, a lexicon-based sentiment classification system for social media genres which integrates strategies to capture contextual polarity from two perspectives: the interaction of terms with their textual neighbourhood (local context) and text genre (global context). We introduce an approach to hybridise a general purpose lexicon, SentiWordNet, with genre-specific vocabulary and sentiment. Evaluation results from diverse social media show that our strategies to account for local and global contexts significantly improve sentiment classification, and are complementary in combination. Our system also performed significantly better than a state-of-the-art sentiment classification system for social media, SentiStrength.

Preprint submitted to Elsevier *May 16, 2016*

1. Introduction

Sentiment analysis concerns the study of opinions expressed in text. The task of sentiment analysis comprises of the extraction of opinion polarity (positive or negative), the target or specific aspects of the target to which the opinion refers, the holder of the opinion, and the time at which the opinion was expressed [1]. Aggregation of sentiment polarity scores from a resource such as a sentiment lexicon is typically used to classify opinionated text into sentiment classes. As a result, several general purpose sentiment lexicons have been developed and made public for research, e.g., General Inquirer [2], Opinion Lexicon [3] and SentiWordNet (SWN) [4]. However, the performance of lexicon-based sentiment analysis still remains below acceptable levels. This is because the polarity with which a sentiment-bearing term appears in text (i.e. contextual polarity) can be different from its prior polarity offered by a lexicon. Two forms of semantic difference seem to contribute to this semantic gap. First, the difference in *local context*, arising from the interaction of the term with its textual neighbourhood. For example, the prior polarity of 'good' is positive, however, such polarity is changed in 'not good'. Second, the difference in *global context* arising from the difference in the typical sentiment polarity of a term captured by a lexicon and the term's domain- or genre-specific polarity. For example, in the text 'the movie sucks', although the term 'sucks' seems highly sentiment-bearing, this may not be reflected in a general purpose sentiment lexicon. Also, as sentiment lexicons are static resources, they need to be equipped with a strategy to adapt to changing vocabulary and sentiment over time - a characteristic of social media.

In this paper, we propose an approach to account for local and global contexts in social media genres. First, we introduce strategies to account for sentiment modifiers: negations, intensifiers/diminishers, and discourse structures. Here, we leverage the fine-grained sentiment information offered by SWN. To account for discourse structures, we introduce heuristic-based discourse parsing and weighting based on the Rhetorical Structure Theory (RST) [5]. RST posits that text can be broken into non-overlapping spans in a tree-like structure with relations that may exist between any two adjacent spans. Each text span can either have the status of the central focal point of the writer's message (i.e. nucleus) or a supporting message that helps in understanding the nucleus (i.e. satellite). As our approach is heuristic-based, we avoid the need for parsers trained with text untypical of social media, yet maintain the theoretical framework of RST. Our strategies to account for local context also incorporate non-lexical modifiers commonly used to express or emphasise sentiment in social media: capitalisation, sequence of repeated character, and emoticons. Second, we introduce an approach to hybridise general purpose lexicons with genre-specific sentiment polarities (global context) and vocabulary. The main contributions of this paper are as follows:

- We introduce a set of strategies relevant to both the social media and a high-coverage lexicon (SWN) that adjusts term prior polarity based on local context. These include strategies for negation, intensification/diminishing, discourse structure, and non-lexical modifiers.
- We introduce a strategy to adapt a lexicon to a domain by facilitating genre-specific vocabulary enhancement using distant-supervised learning.
- We provide a comparative analysis with state-of-the-art systems.

To the best of our knowledge, this is the first time SWN, together with the proposed contextual analysis are applied to sentiment classification of social media. The rest of the paper is organised as follows. Related work is presented next in Section 2, followed by our system (SmartSA) in Section 3. Evaluation results are presented and discussed in Section 4, followed by conclusions and future work in Section 5.

2. Related Work

The task of sentiment classification involves the labelling of text with sentiment class. Several methods have been employed for the task, drawing from both supervised/unsupervised machine learning and lexicon-based unsupervised strategies. Inspired by the field of topic-based text classification, supervised methods make use of machine learning algorithms trained with sentiment-labelled data to predict sentiment class of unlabelled test documents. Although this method was shown to work well in sentiment classification, it becomes problematic when reliable and sufficient training data are difficult to obtain. This is particularly the case for the non-review-based social media where content is not associated with ratings that could be exploited as “noisy” labels. A solution to the problem of labelled data acquisition is the use of unsupervised topic modelling approaches. These typically involve the use of probabilistic topic detection methods to detect both topic and sentiment from a collection of unlabelled documents.

Machine learning sentiment classifiers tend to be highly domain/genre specific, performing well on the domain/genre of training but poorly on a different domain/genre. However, social media text is diverse in domains and genre ranging from political to lifestyle discussions with short messages (e.g., tweets) and lengthy posts (e.g., blogs). Therefore, a system for analysing social media text needs to maintain consistent performance across domains/genres. This is a characteristic of the lexicon-based methods to sentiment classification. In this paper, we adopt the lexicon-based methods, hence, we concentrate on these methods in the rest of this related work section.

2.1. Lexicon-based Methods

A lexicon-based sentiment analysis begins with the creation of a list of words associated with their sentiment polarity values (i.e. a sentiment lexicon), or the adoption of an existing one, from which the sentiment scores of terms are extracted and aggregated to predict sentiment of a given piece of text. Sentiment lexicons are either manually or semi-automatically generated from generic knowledge sources. Manually generated lexicons are obviously more accurate, however, they tend to have relatively low term coverage. In contrast, semi-automatically generated lexicons, such as by expanding a small set of seed words within a large corpus [6] or by dictionary propagation [4], have a high coverage of over 20,000 words. Moving away from traditional lexicons that tend to capture individual terms, SenticNet has been introduced based on the idea of integrating concepts with common-sense knowledge [7]. SenticNet is a graph-structured resource with concepts as nodes and common-sense relationship between concepts as edges. Thus, when a concept extracted from a test text is triggered within SenticNet, common-sense knowledge associated with that concept can be exploited to enrich the machine's assessment of the problem being solved. Another resource with similar structure to SenticNet is WordNet [8], a machine readable dictionary that provides definitions of disambiguated word senses and establishes several relationships among them. These word senses were assigned quantified positive, negative and neutral polarity scores using an automated process to form the sentiment lexicon, SWN [4]. In this work, we use SWN as a general-purpose sentiment lexicon motivated by its relative high coverage of terms and its fine-grained sentiment information at word-sense level rather than term level.

A baseline lexicon-based classifier predicts the polarity class of a document using the aggregate of polarities of the terms contained in the document. With SWN, the sentiment dimension (positive or negative) that has the highest aggregate score becomes the sentiment class for the document [9–12]. This approach is inadequate for an effective sentiment analysis because the prior polarities of terms offered by a lexicon can be different from the contextual polarities of the terms. Such a difference, for instance, can arise due to the effect of linguistic rules such as negation or domain-specific term semantics that are not captured in a lexicon [13].

2.2. Contextual analysis

This involves the adjustment of a term's prior polarity to reflect its polarity in a specific context. For example, the text *"I don't like the idea of smoking in general"* may be classified as positive because it is dominated by positive terms (*'like'* and *'idea'*). However, the appearance of the negation (*'don't'*) in the linguistic context of both terms rendered the text to be negative. In a contextual analysis strategy, the polarities of terms that are under the influence of negation are switched to the opposite sentiment dimension [14, 15]. Similarly, polarity strength of terms that are under the influence of intensifiers (e.g., *'very'*, *'highly'*) or diminishers (e.g., *'slightly'* and *'a-little-bit'*) are increased and decreased respectively. Negation analysis is a particular challenge as the polarity of negated terms do not always translate to its opposite. For instance, whereas "It is *not good*" is more or less the same as "It is *bad*", "It is *not excellent*" is more positive than "It is *horrible*". Consequently, a shift approach was proposed as a preferred alternative to sentiment inversion for negation [16]. Here, the prior polarity of sentiment terms that are under the influence of negation is reduced by a certain weight, but the negation terms were not considered to bear sentiment of their own. However, a recent study suggests that negation terms are not just modifiers of sentiment but also indicators of sentiment [17]. In SWN, negation terms are associated with polarity scores. Thus, a strategy can be introduced to treat negation both as sentiment-bearing and as sentiment modifier for other terms.

Sentiment lexicons are typically generated independently of their target application. Thus, they tend to capture knowledge that is applicable across diverse domains (i.e. they are general-purpose). Not surprisingly deviations are common, especially on social media genres, due to variability in vocabulary usage resulting in poor sentiment coverage. Contextual deviations are also common, for instance where the sentiment polarities of terms differ from the domain-specific use of the terms. The poor sentiment coverage can be improved using a lexicon expansion strategy. In [18], a general-purpose lexicon has been expanded with Twitter-oriented sentiment-bearing terms extracted based on their mutual information with emoticons. In [19], a sentiment lexicon derived from SenticNet [20] was expanded with additional terms from WordNet-Affect [21] using a term-level polarity classifier trained on the vocabulary intersection of both resources. An evaluation of the expanded lexicon on a sentiment classification task using the polarity scores of concepts along with their emotion information, acquired from WordNet-Affect, shows a performance improvement over the baseline approaches [22].

To address contextual deviations, strategies are developed to adapt a sentiment lexicon to an application domain, often utilising a sentiment-labelled dataset from the domain. [23] use the integer linear programming method to adapt a general-purpose lexicon to a domain by utilising sentiment expressions from the domain to derive the most likely polarity of each term (positive, neutral, negative, or negator) for the given domain. Similarly, a domain-specific sentiment lexicon has been adapted to another domain using the information bottleneck framework [24]. In order to address the need for labelled data researchers have utilised alternative knowledge sources. For instance, in review domains star-rating knowledge can be usefully adopted to generate labelled data. This approach has been successfully used to adapt a general-purpose lexicon derived from SenticNet to a domain-specific one [25]. Here, the labelled dataset helped identify ambiguous sentiment-bearing terms which are disambiguated using contextual information and the word sense knowledge from WordNet and ConceptNet. Our approach also exploits a general purpose sentiment lexicon combined with labelled domain data. However, instead of limiting our supervision to domains with star ratings, we adopt distance supervision strategies that are scalable for social media content. A further difference is that by using SWN as our lexicon, a relatively high-coverage lexicon, we are able to utilise fine-grained sentiment information at the word-sense level rather than term level.

In summary, social media is characterised by diversity in domains/genres and the lack of training data, making the lexicon-based approaches better suited for sentiment classification. However, accounting for contextual polarity, improving lexicon coverage, and domain/genre adaptation remain challenges for sentiment analysis. In this paper, we address these challenges by leveraging high-coverage and fine-grain sentiment information from SWN, introducing relevant contextual analysis strategies, and adaptation to social media genres. Next, we present our sentiment classification system and its associated strategies.

3. SmartSA System

SmartSA uses a hybrid lexicon which captures both general purpose knowledge from SWN as well as genre-specific knowledge (global context) to determine polarity scores for pre-processed documents (see Figure 1). Thereafter, the polarities are adjusted to account for the effect of local context on terms. Here, we introduce strategies for negation, intensification/diminishing, discourse analysis, capitalisation, the use repeat letters/characters, and emoticons. Sentiment class for a given document is determined by the maximum of the contextually modified scores. Details of these operations are presented next.

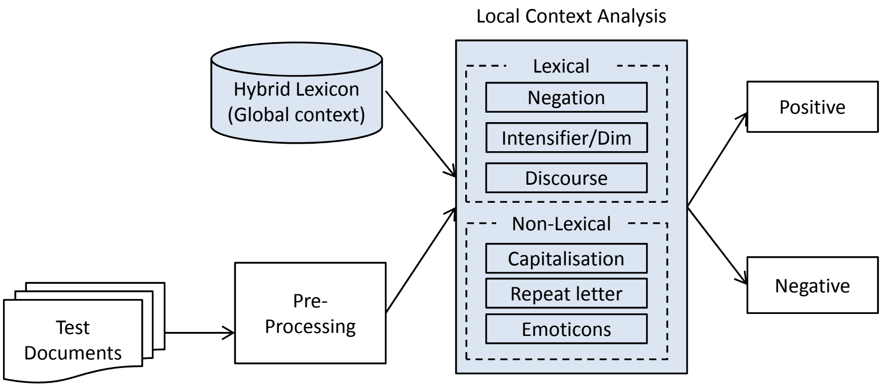


Figure 1: SmartSA

3.1. Pre-Processing

Prior to the main task of sentiment classification, it is essential to apply text pre-processing operations to transform input text into unit terms and associated information. We use the TweetNLP [26] library for tokenisation and PoS tagging. As the word lemma is required to extract scores from SWN, we apply lemmatisation[1] to convert each token to its dictionary equivalent.

[1]We use the lemmatizer from the Stanford CoreNLP library [27]

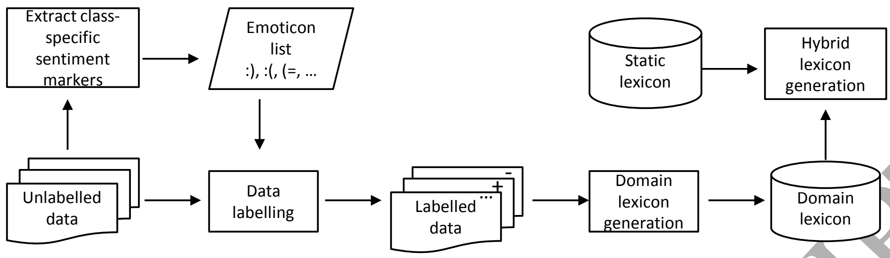


Figure 2: Diagram showing stages involved in the hybrid lexicon generation

3.2. Global Context

When SWN is used as the sole sentiment lexicon, only sentiment-bearing terms that have an entry in the lexicon contribute towards analysis. This means many domain-specific terms are likely to be ignored. Similarly, some terms might have their sentiment context misrepresented in the lexicon, as it only captures general purpose usage of terms. To address this limitation, we introduce a strategy to hybridise SWN with terms and sentiment context extracted from the domain of application.

The process of generating the hybrid lexicon is shown in Figure 2. First, a domain-focused lexicon is generated from data extracted from the target domain and labelled using distant supervision approach. Next, the hybrid lexicon is generated by combining the sentiment scores from the domain-focused lexicon with existing scores in SWN. We look at each of these in turn.

3.3. Data Labelling: Distant Supervision

Distant supervision offers an automated approach to assigning sentiment class labels to documents. It uses the presence of class-specific emoticons in a document as evidence for its true class. For example, a smiley-face emoticon would, according to distant supervision, be considered to be expressing positive sentiment and, as such, evidence for labelling the related content as belonging to the positive class. Accordingly, given a dataset and a lexicon of class-specific emoticons, we can assign such noisy labels to all documents that contain them in order to generate a labelled dataset for supervised learning tasks. To minimise

Table 1: Datasets and sizes

Dataset	#Pos.(#Neg.)	Avg. Sentences	Avg. words
Twitter	10,000	1.96	16.84
Digg	5,222	6.69	78.71
MySpace	292	4.36	12.80

the level of potential noise, a reasonable strategy is needed to process documents containing emoticons from both positive and negative classes. In this work, we avoid documents with mixed emoticons.

We generate three distant-supervised datasets from varying web social communication settings (see Table 1): from blog messages (Digg and MySpace samples made available by cyberemotions.eu) to micro-blogs (twitter sample made available by sentiment140.com). Unlike Twitter dataset which contain short documents due to its character limit, with Digg and MySpace, we confine the labelling to the sentences that contain emoticons rather than the whole document. Such a sentence-level labelling is more intuitive since emoticons often apply only to the sentence in which they appear. With both these datasets, there were many more positive (almost 80%) compared to negative emoticons present.

Accordingly, a balanced sample from this extremely skewed distribution was used to create the distant-supervised datasets. The main difference between the Digg and MySpace datasets is in their sizes (Digg with 5,222 and MySpace with 292 positive/negative messages). Twitter, unlike with the other two datasets, contained over a million distant-supervised tweets. We sampled 10,000 from each class (positive and negative) to generate a suitably sized dataset in this work.

3.4. Domain Lexicon Generation

The domain-focused lexicon associates a positive and negative score to each unique term in the distant-supervised dataset. Key to this generation is to capture association of a term, t_i, to a class, c_j, given a set of distant-supervised documents, D. Although there are many metrics to quantify the term-class association, in this work we use the simple metric shown to be effective for low-frequency, social media terms [28]. This is given in Equation 1.

$$ds(t_i, c_j) = \frac{\mathrm{TF}(t_i, D_{c_j})}{\mathrm{TF}(t_i, D)} \tag{1}$$

Where, D_c is the subset of D labelled as class c, $\mathrm{TF}(t_i, D)$ is the term frequency of t_i in D, and $ds(t_i, c_j)$ is the domain-focused association of t_i with c_j.

3.5. Hybrid Lexicon Generation

Scores from SWN and domain-focused (DF) lexicons for each term t_i are combined to form the hybrid score for the term (Algorithm 1). When t_i appears in both lexicons, a weighted average of positive and negative scores supplied by both lexicons is calculated using α and β as mixing parameters for positive and negative scores respectively. This weighting favours scores from one lexicon over the other. So $\alpha = 0.5$ would lead to equal weighting of positive scores from SWN and DF, whilst α= 0 will ignore positive score from SWN (see step 3). The use of different mixing parameters is likely to separately address possible bias towards a sentiment dimension (usually positive) due to the observation that people tend to use positive terms in a more frequent and diverse manner (Pollyanna hypothesis) [29]. We determine optimal values for the mixing parameters, α and β, as the combination that produces the highest performance on a development dataset.

When only one lexicon, SWN or DF, contains scores for t_i, such scores are fully used without an aggregation (see steps 5 and 7). Thereafter, the new scores for t_i (i.e. t_i^+ and t_i^-) are added to the hybrid lexicon, H (step 9). Finally, H is returned as the output.

3.6. Local Context

In social media, two types of modifiers affect term polarity in context: *lexical* and *non-lexical* valence shifters. Lexical valence shifters are in the form of dictionary recognisable words, whereas non-lexical valence shifters are other

Algorithm 1 Generate Hybrid Lexicon

INPUT: SWN, SentiWordNet
DF, Domain-focused Lexicon
α, β Unifying weight
OUTPUT: H, Hybrid lexicon

for all $t_i \in (SWN \cup DF)$ **do**
  if $t_i \in SWN \cap DF$ **then**
    $t_i^+ \leftarrow \alpha \times (t_i^+ \in SWN) + (1-\alpha) \times (t_i^+ \in DF)$; $t_i^- \leftarrow \beta \times (t_i^- \in SWN) + (1-\beta) \times (t_i^- \in DF)$
  else if $t_i \in SWN$ **then**
    $t_i^+ \leftarrow (t_i^+ \in SWN)$; $t_i^- \leftarrow (t_i^- \in SWN)$
  else
    $t_i^+ \leftarrow (t_i^+ \in DF)$; $t_i^- \leftarrow (t_i^- \in DF)$
  end if
  $H.AddEntry(t_i^+, t_i^-)$
end for
Return H

word inflexions and artificial symbols that affect the expression of sentiment, such as repeating a letter or character, capitalisation for emphasis and the use of emoticons. Crucial to implementing any scores adjustment strategy is the identification of the term, or a group of terms, affected by a modifier in text (i.e the scope of the modifiers). Ideally, it is the task of a dependency parser to identify modifiers in text and the terms they modify. However with the attendant non-standard spelling and grammar of social media, standard parsers often fail to produce satisfactory results. Thus, we adopt window-based approaches, whereby modifiers are assumed to affect terms within a specific text window [30–32].

3.7. Lexical Valence Shifters

Lexical valence shifters are typically used to increase sentiment (i.e. intensifiers, e.g., *'very'*, *'highly'*); decrease sentiment (i.e. diminishers, e.g., *'slightly'*, *'somewhat'*) or negate sentiment (i.e. negation terms, e.g., *'not'*, *'never'*). These terms are associated with sentiment scores in SWN. For example, positive and negative scores of the adverb 'very' are 0.25 and 0.0 respectively, thus, the term always contributes positively. However, this term can also contribute negatively, for example, in 'very bad'. Therefore, it is important to determine the polarity contribution likely to be made and modify scores accordingly.

3.7.1. Negation

Negation is a common linguistic phenomenon that affects expressions of sentiment in a profound way. Utilising the positive and negative scores for negation terms in SWN, we introduce a strategy in which negation is considered as a sentiment diminisher rather than a complete inverter of sentiment (i.e. the shift approach). We also integrated polarity scores of negation terms into the sentiment aggregation process, thus, capturing the concept that negation is both a modifier and sentiment-bearing. In our approach, considering that modifiers tend to affect the dominant polarity when a term is negated, we ignore this polarity. For instance, in Figure 3, examples (a) and (b), the contextual polarity of the phrases 'not good' and 'not excellent' becomes negative after the shift operation. The relative intensities of their polarity is also maintained (i.e. 'not good' is more negative than 'not excellent'). However, for the negation of terms that are more negative than positive, utilising scores of negation terms will produce an undesired result. For instance, *'not angry'* will still remain overall negative (Figure 3, example c). In such cases, we exclude scores of negation terms from the aggregation (such as in example d). We use a text window size of three terms before and after a negation term to establish the scope of the negation. Our negation detection is based on the list of negation terms in [31], extended to include scenarios when an apostrophe is omitted or misplaced for the terms such as in *don't*, *wouldn't*, *couldn't* and *can't*.

3.7.2. Intensification/Diminshing

Intensifiers and diminishers are linguistic constructs used to increase and decrease the sentiment or emotional charge of terms. In SmartSA, the value of the dominant polarity of terms that are within the scope of the intensifier is increased (or decreased in the case of diminisher) relative to the strength of the

	Before Adjustment			After Adjustment			
(a)	not	good	→	not	good	:	sum
pos:	0.000	0.638		0.000	~~0.638~~	:	0.000
neg:	0.625	0.125		0.625	0.125	:	0.750
							aggregate (pos-neg)=-0.750
(b)	not	excellent	→	not	excellent	:	sum
pos:	0.000	1.000		0.000	~~1.000~~	:	0.000
neg:	0.625	0.000		0.625	0.000	:	0.625
							aggregate (pos-neg)=-0.625
(c)	not	angry	→	not	angry	:	sum
pos:	0.000	0.307		0.000	0.307	:	0.307
neg:	0.625	0.500		0.625	~~0.500~~	:	0.625
							aggregate (pos-neg)=-0.318
(d)	not	angry	→	not	angry	:	sum
pos:	0.000	0.307		~~0.000~~	0.307	:	0.307
neg:	0.625	0.500		~~0.625~~	~~0.500~~	:	0.000
							aggregate (pos-neg)=0.307
(e)	really	awful	→	really	awful	:	sum
pos:	0.438	0.250		~~0.438~~	0.250	:	0.250
neg:	0.065	0.542		~~0.065~~	0.542×150%	:	0.878
							aggregate (pos-neg) =-0.628

Figure 3: Examples: Negation and Intensification Adjustments

intensifier (or diminisher), as illustrated in Figure 3, (e). We use a lexicon of intensifiers and diminshers, where each term is annotated with intensification or diminishing strength. For instance, the intensification strength of 'extremely' is 100% while that of 'very' is 25% increase in dominant polarity. Unlike with negation, it was observed that although intensifiers/diminshers might be associated with sentiment scores, it is better for sentiment analysis when they are decoupled from such scores and strictly treated as modifiers [12]. Thus, we do not include sentiment scores from intensifiers/diminishers in the aggregation process.

3.7.3. Discourse structure

The main idea behind harnessing discourse structure for sentiment analysis is that since discourse structure of a piece of text can specify segments of the text that are more (or less) important to the writers message, it can also be exploited to associate weights to the segments. Consequently, sentiment terms that occur within the important segments will have higher weights. This will potentially lead to an improved sentiment analysis. Working with this notion, in SmartSA, we use regular expressions to identify the occurrence of discourse markers and apply a weight to their scope. Here, the scope of a discourse marker is the two text segments involved in the relation the marker represents. We use the rule-based algorithm in [33] to split up text into discourse segments using lists of discourse markers per relation [34]. We extend this list to include social media variation of the markers, such as *'cos'*, *'bcos'* or *'bc'* for the marker 'because'. Next, amongst the two segments in a relation, we need to differentiate between the central focal point of the writer's message (nucleus) and the supporting text (satellite). To this end, we utilise the contextual information derived from a corpus study of distributional environments for discourse markers [34]. This information specifies the nucleus/satellite of a relation in reference to a given segment containing a discourse marker of the specified relation (this can be the segment before or after). This is usually influenced by the position of the given marker within its segment: beginning, middle or end.

After the discourse segmentation and the identification of nucleus/satellite segments, we apply a weight corresponding to the potential effect of each segment for sentiment analysis. Considering that, similar to the role of intensifiers/diminshers, the effect of discourse increases/decreases sentiment, we mapped this on the effect of typical intensifier/diminisher (i.e. 50% increase/decrease). Although, [5] identified 24 generic discourse relations, not all are relevant for sentiment analysis. Thus, here we concentrate on the subset of 13 relations identified to be useful for sentiment analysis [34]. We heuristically group the discourse relations according to their potential effect, with respect to sentiment expression, to their nucleus or satellite as follows.

Group 1: No Effect on Nucleus, Decrease Satellite. These are the relations of *concession* and *background*. Concession holds between conflicting information present in nucleus and satellite segments whereby the writer clearly favours the nucleus, though not denying the satellite. Therefore, it is worthwhile for a sentiment analysis system to concentrate on the sentiment expressed in the nucleus of this relation while suppressing the satellite. For example, in [**although** I dont like the series,]$_{\mathrm{S}}$ [I really enjoyed this episode]$_{\mathrm{N}}$, the writer seems to promote the positive sentiment (really enjoy) within the nucleus segment (denoted by the subscript N) despite the negative sentiment (dont like) of the satellite segment (denoted by the subscripts S). In this example, the relation is signalled by the discourse marker *although* (denoted in bold font). For background, the satellite provides a context based on which the information provided in the nucleus can be better understood. The sentiment expressed in this context can be in conformity with that expressed in the nucleus or otherwise. However, since the nucleus is the focal point of the relation, it is more reliable to concentrate on the sentiment it conveys and suppress the sentiment in the satellite which can be tangential with the overall sentiment of both. For example, in [I was happy the laptop was working]$_{\mathrm{S}}$ [**but 3 days later** it stopped]$_{\mathrm{N}}$, the focus is on the negative sentiment within the nucleus (stopped) despite the positive sentiment in the satellite (happy).

Group 2: Decrease Nucleus, Decrease Satellite. These are the relations of *condition*, *circumstance* and *purpose*. Condition presents a hypothetical future whereby the realisation of the nucleus depends on the realisation of the satellite. However, both nucleus and satellite are unrealised. Thus, for the purpose of sentiment analysis, such a situation can be given low weight. For instance, in [**if** the world ends on december 2,]$_{\mathrm{S}}$ [i'm gonna be so disappointed]$_{\mathrm{N}}$, despite the negatively charged terms in both segments (world ends, disappointed), the text still seems to remain largely neutral. For circumstance, the satellite sets the framework within which the reader is expected to interpret the nucleus. It tends to soften both the nucleus and satellite. For example, the statement: [The animal is dangerous]$_{\mathrm{N}}$ [**when** left in hunger]$_{\mathrm{S}}$, though dominated by negative terms (dangerous, hunger) is still of mild sentiment. Similarly in purpose, the satellite presents a situation to be realised through the activity in the nucleus, as in the example: [the quality of the food should be improved]$_{\mathrm{N}}$ [**so as** to improve sales]$_{\mathrm{S}}$.

Group 3: No effect on Nucleus, Increase Satellite. These are *elaboration*, *evaluation*, *re-statement*, *summary* and *cause/result* relations. Elaboration exists between a nucleus and a satellite when the satellite presents additional information to better understand the nucleus. Thus, sentiment present in the satellite tends to be supportive of the nucleus. It also tends to be more verbose, increasing the chance of containing sentiment-bearing terms. For example, in [**in addition** to the location,]$_{\mathrm{N}}$ [the food also tastes good]$_{\mathrm{S}}$, the sentiment expressed within satellite (good) also applies to the nucleus. Re-statement tends to function in a similar manner as elaboration. The satellite is the paraphrase of the nucleus. Thus, sentiment within the satellite is important as it is also applicable to the nucleus. In the evaluation relation, the satellite tends to contain an opinion regarding the nucleus. This is directly relevant for sentiment analysis as it signals a reliable location for opinions. For example, [Now **it seems** action of Yadav]$_{\mathrm{N}}$ [have back fired]$_{\mathrm{S}}$, the evaluation marker (it seems) signals the appearance of the sentiment-charged term (back fired) in the satellite. In summary relation, the satellite provides concise and overall information the writer meant to convey from an often lengthier nucleus. Opinion expressed in the satellite is thus representative of the text and can be given high weights. Finally, the cause/result signifies relation between satellite and nucleus whereby the information given in the satellite is the cause of the information present in the nucleus. Both segments tend to present the same sentiment orientation, with the satellite being central to believing the nucleus. For example, in the text: [I always eat in that restaurant]$_{\mathrm{N}}$ [**because** of its friendly staff]$_{\mathrm{S}}$, the positive justification in the satellite (friendly staff) adds strength to the overall sentiment of the text.

3.8. Non-lexical Modifiers

In addition to lexical valence shifters, non-lexical modifiers are also commonly used to increase sentiment in social media. These modifiers manifest in the form of term inflexion with a sequence of repeating characters/letters, capitalization and the appearance of emoticons.

3.8.1. Capitalisation

The informal social media communication present the convention of term capitalisation for emphasis. This is often used to emphasise sentiment or emotion expressions. Therefore, we introduce an approach in which capitalisation is treated as the intensification of the capitalised term. This adjustment is applied only if the rest of the sentence is not capitalised because in such cases the capitalisation may not be for emphasis but writing style. We use the intensification strength of 'very', being an average and the most occurring lexical intensifier in our datasets. For example, the sentence "saw this last night...AMAZING!" becomes "saw this last night...very amazing!". We do not extend the intensification to the neighbouring terms because capitalisation is also often used for abbreviations and acronyms.

3.8.2. Repeated Letter/Character

Repeat of the same letter or character is another phenomenon used to express emphasis in social media. In SmartSA, when a sequence of three or more letters is detected, the target term is identified by first reducing the number of the letter to a maximum of two and then checked with SWN. If the intermediate word is not found, the repeated letters are further reduced to one letter, one sequence at a time. We consider a sequence of repeated letters as an intensification of not just the affected term but also its context. This is because, unlike with capitalisation, a sequence of repeated letters is mainly for emphasis, and sometimes the affected term is not sentiment-bearing (e.g., "Mannnnnn, I loved this show"). The occurrence of three or more consecutive exclamation or question marks or a mixture of both is also treated as the sentiment intensification of context using the intensification weight of the word 'very'.

3.8.3. Emoticons

In the informal social media, emoticons are often used to express sentiment for either the whole document or individual sentences. We use regular expressions to identify positive and negative emoticons based on the list of emoticons provided in [30]. If one or more positive (or negative) emoticons are found in a sentence, the sentence is simply assigned the scores of the emoticon (i.e. pos=1.0, neg=0.0 for positive emoticon; pos=0.0 and neg=1.0 for negative emoticon). We restrict the context of emoticons to sentence level, as sentiment can change from one sentence to another [35].

3.9. SmartSA *Algorithm*

The classifier is shown in Algorithm 2. It takes as input the document to be classified, the hybrid lexicon, and the lists of lexical valence shifters and emoticons. Each sentence contained in the document is checked for the occurrence of an emoticon. If present, the sentence carries sentiment scores of the emoticon without further analysis of the sentence's text (steps 3-4). Otherwise, the sentence's text is scanned for terms that contain repeating letters or characters of question/exclamation marks. These are converted to their dictionary equivalents (step 8) and appended with a typical intensifier 'very' (step 9). Next, sentiment scores for each term are extracted from the hybrid lexicon. Terms that are selectively capitalised within the sentence are intensified using the intensification weight of a typical intensifier (i.e. 25%). Thereafter, score adjustments, based on the occurrence a lexical valence shifters, are applied to the context of the term (i.e. its neighbourhood) in steps 16-22. Each sentence is assigned the total adjusted scores of its terms. Likewise, each document is assigned the total scores of its sentences. Lastly, the document class is returned as positive, if its total positive score is greater than or equal to its total negative score. Otherwise, the class is returned as negative.

Algorithm 2 SmartSA

INPUT:	H,	Hybrid Lexicon
	LexValShifters{}	list of Negation, Intensifiers/Diminishers and discourse markers
	Emoticons{}	List of positive and negative emoticons
	Doc,	Document to be classified
OUTPUT:	Class,	Sentiment class for Doc

```
Initialise Doc⁺, Doc⁻, Sent⁺, Sent⁻
for all Sentence ∈ Doc do
    if ContainSingleType(Emoticon{}) then
        Sent⁺+ ← EmoticonType⁺;    Sent⁻+ ← EmoticonType⁻
    else
        for all t ∈ Sentence do
            if t.hasRepeatCharacter then
                convertStandard(t, SWN)
                appendIntensifier('very', t)
            end if
            Retrieve t⁺ and t⁻ from H
            if t.isCaps AND ¬sentence.isCaps then
                applyAdjustment(25%, t)
            end if
        end for
        for all mod ∈ LexValShifters{} do
            if mod ∈ sentence then
                modType ← getType(mod)
                context ← getContext(mod, modType, sentence)
                ApplyAdjustment(modType, context)
            end if
        end for
        Sent⁺+ ← sum (t⁺ ∈ sentence),    Sent⁻+ ← sum (t⁻ ∈ sentence)
    end if
    Doc⁺+ ← Sent⁺,    Doc⁻+ ← Sent⁻
end for
if Doc⁺ ≥ Doc⁻ then
    Return Positive
else
    Return Negative
end if
```

4. Evaluations

We conduct a comparative study to evaluate the role of accounting for both local and global contexts of terms, as proposed in this work, for sentiment classification of social media text. The aim of the evaluation is two-fold. First to investigate the performance of our system compared to the baseline and state-of-the-art classifiers. Second, to evaluate the contribution of each individual component of our system. To this end, we compare the following systems:

1. Base: Basic sentiment classification algorithm using SWN, as discussed in Section 2.
2. Base+LC: An extension of the Base algorithm with accounting for local context (i.e. the SmartSA Algorithm but using SWN instead of the hybrid lexicon).
3. Base+GC: An extension of the Base algorithm with accounting for global context (i.e. the baseline aggregation but using the hybrid lexicon).
4. SmartSA: The classifier introduced in this paper (i.e. Algorithm 2).
5. State-of-the-art machine learning algorithms: We use three commonly used sentiment classification algorithms: Support Vector Machines (SVM), Naïve Bayes (NB) and Logistic Regression (LR). These classifiers are trained with the distant-supervised data (see Table 1) and tested with the hand-labelled data. We used the combination of uni-gram, bi-gram and tri-gram as training features after tokenisation using the TweetNLP tool.
6. SentiStrength, a state-of-the-art sentiment classifier for social media.

All algorithms are tested using human labelled datasets from the three social media platforms, introduced earlier containing: 2587 positive and 843 negative Twitter[2]; 107 positive and 221 negative Digg[3]; and 400 positive and 105 negative

[2]Test data from SemEval 2015, task 10B

[3]Acquired from Cyberemotions.eu

MySpace[4] examples. For the parameters α and β in hybrid lexicon generation, we randomly split the labelled data into 40% for a development sample on which parameters are learnt, and 60% for test. We report the average over 20 such experiments in our results to mitigate against a possible sensitivity of the learnt values on a specific development sample. As is typical with unbalanced datasets [30, 36] we present results based on the average value of the F1-score for positive and negative classes to quantify classification quality. Class-based precision (P) and recall (R) are also reported. Finally, we measure statistical the significance of difference in F1-score using the Chi-square test.

4.1. Results and Discussion

Table 2 shows sentiment classification results on Twitter, Digg and MySpace datasets. Asterisk (*) and bold font indicate the best performance on a dataset and a significant difference from Base respectively. Combining Local and global contexts (SmartSA) performs best on Twitter and Digg datasets. Likewise, both Base+LC and Base+GC significantly improve upon Base on these datasets. Base+LC performs better than Base+GC on Digg while the reverse is the case on Twitter. This can be attributed to the fact that lexical modifiers, which Base+LC accounts for, are more likely to appear in Digg than in Twitter due to the lengthier documents in Digg. Also, the number of distant-supervised tweets, which is about double the size of Digg (see Table 1), makes it more likely for Twitter to have a more reliable domain-focused lexicon, hence the good performance in Base+GC (i.e. global context). This observation is obvious on the smaller MySpace dataset, where Base+GC performed less well than Base (difference of 5.59%), and SmartSA performed just marginally better than the Base.

Overall, the SmartSA approach performs better than all of the supervised machine learning algorithms (SVM, NB and LR) on all three datasets, providing an improvement of 1.03% on Twitter, 5.83% on Digg and 1.99% on MySpace

[4] Also from Cyberemotions.eu

Algorithm	Positive			Negative			Avg F1
	P	R	F1	P	R	F1	
Twitter							
SVM	67.40	33.20	44.49	55.60	82.70	66.49	55.49
NB	65.60	67.30	66.44	65.20	63.50	64.34	65.39
LR	74.70	81.00	77.72	78.20	71.40	74.65	**76.19**
Base	84.90	75.88	80.14	44.19	58.60	50.38	65.26
Base+LC	87.64	79.51	83.38	51.06	65.60	57.42	**70.40**
Base+GC	68.20	74.50	71.21	76.20	70.20	73.08	**72.15**
SmartSA	74.80	79.00	76.84	79.70	75.60	77.60	**77.22***
SentiStrength	86.20	84.20	85.19	54.70	58.60	56.58	**70.87**
Digg							
SVM	35.10	49.70	41.14	69.90	55.00	61.56	51.35
NB	35.30	49.70	41.28	70.10	55.50	61.95	51.62
LR	45.80	72.20	56.05	81.70	58.20	67.98	62.02
Base	37.44	75.24	50.00	85.56	53.85	66.10	58.05
Base+LC	43.00	83.33	56.73	90.67	59.44	71.81	**64.27**
Base+GC	81.50	44.60	57.65	53.30	89.20	66.73	62.19
SmartSA	87.10	49.00	62.72	59.20	95.10	72.97	**67.85***
SentiStrength	45.60	81.90	58.68	90.60	64.20	75.15	**66.87**
MySpace							
SVM	79.20	100	88.40	0.00	0.00	0.00	44.20
NB	86.90	43.30	57.8	26.60	73.50	39.06	48.43
LR	91.00	70.80	79.64	37.50	67.80	48.29	63.97
Base	88.67	82.48	85.46	32.04	43.94	37.06	61.26
Base+LC	89.31	83.33	86.22	35.00	47.37	40.26	63.24
Base+GC	61.90	86.60	72.20	58.80	29.10	38.93	55.57
SmartSA	77.20	90.30	83.24	61.70	40.20	48.68	**65.96**
SentiStrength	91.80	90.50	91.15	52.80	56.80	54.73	**72.94***

Table 2: Classification results from Twitter, Digg and MySpace datasets

when compared with the best-performing classifier, LR. This confirms the superiority of our lexicon-based approach using a hybrid lexicon generated from distant supervision over the machine learning approaches to sentiment classification. The improvement is more pronounced on the Digg dataset. This dataset differs from the other two in documents size and the composition of positive/negative documents (it is composed of mostly negative documents). While its lengthier document size might have helped in local context analysis, the hybrid lexicon is likely to have helped address the typical limitation of general-purpose lexicons (such as SWN) which tend to have many true positives but with lower precision rates on the negative class.

Compared to the state-of-the-art system, SentiStrength, SmartSA performed best on Twitter and Digg datasets. It can be noted that SentiStrength integrates strategies to account for local context. Thus, its lower performance is largely due to its reliance on a static lexicon, whereas SmartSA adapts vocabulary and sentiment context of the target genre. However, SmartSA performed less well than SentiStrength on the MySpace dataset, apparently due to the small amount of distant-supervised data from this genre. This lead to the question: can we use distant-supervised data from one domain to generate a hybrid lexicon on another domain? We discuss these experiments next.

4.2. Transferability Across Social Media Platforms

Table 3 shows results of transferring a hybrid lexicon across social media platforms (the plus sign, +, indicates improvement while the minus sign, −, indicates decline over using within platform/genre distant-supervised data). For Twitter, using its own genre for distant supervision (i.e. within platform) is better than either using Digg posts or MySpace messages (77.22 Vs 65.12 and 63.73). However, with the other smaller distant-supervised datasets (Digg and MySpace) we see significant improvements when they are augmented or replaced with the larger Twitter distant-supervised dataset. For instance, with Digg an increase of over 5% is observed when using a distant-supervised Twitter dataset. Whilst with MySpace an impressive 10% improvement is observed

Algorithm	Positive			Negative			Avg F1
	P	R	F1	P	R	F1	
Twitter as Distant-supervised dataset:							
Digg	70.90	58.80	64.29	77.10	85.70	81.17	72.73^{+}
MySpace	63.40	93.60	75.60	79.00	36.30	49.74	62.67^{-}
Digg as Distant-supervised dataset:							
MySpace	86.20	90.40	88.25	56.80	48.50	52.32	70.29^{+}
Twitter	74.30	64.10	68.82	56.40	67.40	61.41	65.12^{-}
MySpace as Distant-supervised dataset:							
Twitter	46.10	73.30	56.60	84.30	61.10	70.85	63.73^{-}
Digg	44.50	55.40	49.36	84.80	77.40	80.93	65.15^{-}
All genres as source							
Twitter	73.40	76.10	74.73	76.40	73.80	75.08	74.91^{-}
Digg	70.40	73.10	71.72	73.40	70.60	71.97	71.85^{+}
MySpace	90.40	93.00	91.68	68.40	51.20	58.56	75.12^{+}

Table 3: Transferability of hybrid lexicon across social media genres

with a distant-supervised dataset formed by combining data from all platforms. These results indicate that where the within platform dataset is small or unavailable, using data from a different platform is advantageous. However, the results on MySpace raises the question of which platforms are compatible with each other, considering that the Digg generated lexicon compares favourably over the Twitter lexicon even though the size of the distant-supervised Twitter dataset is a magnitude larger than the Digg dataset.

5. Conclusions and Future Work

In this paper, we presented SmartSA, a sentiment classification system for social media genres. The system incorporates strategies to account for contextual polarities of terms to improve classification accuracy. We confirm previous research that identifies the usefulness of local context in negation, intensifiers, diminishers, discourse structure and other non lexical modifiers. Another aspect of the semantic gap is the difference in vocabulary coverage and term usage between a lexicon and its domain of application. We presented a novel approach to capture this global context through the generation of a hybrid lexicon that enhances a general purpose lexicon (SWN) with domain knowledge for sentiment classification. We demonstrated how distant supervision can be exploited for this purpose. Experimental evaluation shows that the approach is effective, and better than state-of-the-art machine learning sentiment classification trained on the same dataset from which the domain knowledge is extracted in our approach (i.e. distant-supervised data). Our system also achieved better classification performance than a state-of-the-art lexicon-based classifier, SentiStrength. Combining all distant-supervised data from the three domains leads to an overall significant performance improvement with the hybrid lexicon, confirming the transferability of the lexicon across social media platforms. This also suggests that combining distant-supervised data from multiple social media platforms may help, especially where there is not sufficient data from a target platform. However, we also observed that there is compatibility problems between genres that warrants further investigation. In future work, we will explore how characterising a dataset might help towards addressing these problems.

References

[1] B. Liu, Sentiment Analysis and Subjectivity, 2nd Edition, Chapman and Francis, 2010, Ch. Handbook of Natural Language Processing, pp. 627–666.

[2] P. J. Stone, D. C. Dexter, S. S. Marshall, O. M. Daniel, The General Inquirer: A Computer Approach to Content Analysis, MIT Press, Cambridge, MA, 1966.

[3] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.

[4] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proceedings of the Annual Conference on Language Resouces and Evaluation, 2010, pp. 2200 – 2204.

[5] W. Mann, S. Thompson, Rhetorical structure theory: Toward a functional theory of text organization., Text 8 (3) (1998) 243–281.

[6] S. M. Mohammad, S. Kiritchenko, X. Zhu, Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets, in: Proceedings of the 7th int'l workshop on Semantic Evaluation Exercises (SemEval-2013), Atlanta, Georgia, USA, 2013, pp. 321 – 327.

[7] E. Cambria, D. Olsher, D. Rajagopal, Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis, in: Twenty-eighth AAAI conference on artificial intelligence, 2014, pp. 1515–1521.

[8] C. Fellbaum, WordNet: An Electronic Lexical Database, MIT Press, 1998.

[9] B. Heerschop, F. Goossen, A. Hogenboom, F. Frasincar, U. Kaymak, F. de Jong, Polarity analysis of texts using discourse structure, in: Proceedings of the 20th ACM int'l conference on Information and knowledge management CIKM'11, ACM, Glasgow UK, 2011, pp. 1061–1070.

[10] A. Devitt, K. Ahmad, Sentiment polarity identification in financial news: A cohesion-based approach, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, 2007, pp. 984–991.

[11] K. Denecke, Using sentiwordnet for multilingual sentiment analysis, in: ICDE Workshop, 2008, pp. 507–512.

[12] A. Muhammad, N. Wiratunga, R. Lothian, R. Glassey, Contextual sentiment analysis in social media using high-coverage lexicon, in: Research and Development in Intelligent Systems XXX, Springer, 2013, pp. 79–93.

[13] E. Cambria, Affective computing and sentiment analysis, IEEE Intelligent Systems 31 (2) (2016) 102–107.

[14] L. Polanyi, A. Zaenen, Contextual valence shifters, in: Computing attitude and affect in text: Theory and applications, Springer, 2006, pp. 1–10.

[15] A. Kennedy, D. Inkpen, Sentiment classification of movie reviews using contextual valence shifters, Computational intelligence 22 (2) (2006) 110–125.

[16] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational Linguistics 37 (2011) 267–307.

[17] C. Potts, On the negativity of negation, in: N. Li, D. Lutz (Eds.), Proceedings of Semantics and Linguistic Theory 20, CLC Publications, Ithaca, NY, 2011, pp. 636–659.

[18] Z. Zhou, X. Zhang, M. Sanderson, Sentiment analysis on twitter through topic-based lexicon expansion, in: Databases Theory and Applications, Springer, 2014, pp. 98–109.

[19] S. Poria, A. Gelbukh, E. Cambria, P. Yang, A. Hussain, T. S. Durrani, Merging senticnet and wordnet-affect emotion lists for sentiment analysis, in: Proceedings of the 2012 IEEE int'l conference on Signal Processing (ICSP), Vol. 2, IEEE, 2012, pp. 1251–1255.

[20] E. Cambria, R. Speer, C. Havasi, A. Hussain, Senticnet: A publicly available semantic resource for opinion mining., in: AAAI fall symposium: commonsense knowledge, Vol. 10, 2010, pp. 15–18.

[21] C. Strapparava, A. Valitutti, Wordnet-affect: an affective extension of wordnet, in: In Proceedings of the 4th Int'l Conference on Language Resources and Evaluation, 2004, pp. 1083–1086.

[22] S. Poria, A. Gelbukh, E. Cambria, D. Das, S. Bandyopadhyay, Enriching senticnet polarity scores through semi-supervised fuzzy clustering, in: Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction, SENTIRE 2012, at 12th IEEE Int'l Conf. on Data Mining Workshops (ICDMW), Belgium, IEEE, 2012, pp. 709–716.

[23] Y. Choi, C. Cardie, Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification, in: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, ACL, 2009, pp. 590–598.

[24] W. Du, S. Tan, X. Cheng, X. Yun, Adapting information bottleneck method for automatic construction of domain-oriented sentiment lexicon, in: Proceedings of the third ACM international conference on Web search and data mining, ACM, 2010, pp. 111–120.

[25] A. Weichselbraun, S. Gindl, A. Scharl, Enriching semantic knowledge bases for opinion mining in big data applications, Knowledge-based systems 69 (2014) 78–85.

[26] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, M. Heilman, D. Yogatama, J. Flanigan, N. A. Smith, Part-of-speech tagging for twitter: annotation, features, and experiments, in: Proceedings of the 49th Annual Meeting of the ACL, HLT '11, ACL, Stroudsburg, PA, USA, 2011, pp. 42–47.

[27] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, D. McClosky, The Stanford CoreNLP natural language processing toolkit, in: Association for Computational Linguistics (ACL) System Demonstrations, 2014, pp. 55–60.

[28] A. Muhammad, N. Wiratunga, R. Lothian, A hybrid sentiment lexicon for social media mining, in: proceedings of the 2014 IEEE 26th International Conference onTools with Artificial Intelligence (ICTAI), IEEE, 2014, pp. 461–468.

[29] J. Boucher, C. E. Osgood, The pollyanna hypothesis, Journal of Verbal Learning and Verbal Behavior 8 (1) (1969) 1–8.

[30] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, Journal of the American Society for Information Science and Technology 61 (12) (2010) 2444–2558.

[31] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment strength detection for the social web, Journal of the American Society for Information Science and Technology 63 (1) (2012) 163–173.

[32] A. Hogenboom, P. van Iterson, B. Heerschop, F. Frasincar, U. Kaymak, Determining negation scope and strength in sentiment analysis, in: Proceedings of the IEEE Int'l Conference on Systems, Man and Cybernetics, Anchorage, Alaska, USA, 2011, pp. 2589–2594.

[33] D. Marcu, The rhetorical parsing of unrestricted texts: A surface-based approach, Computational linguistics 26 (3) (2000) 395–448.

[34] D. Das, Computational analysis of text sentiment: A report on extracting contextual information about the occurrence of discourse markers, Tech. rep., Simon Fraser University (2010).

[35] A. Andreevskaia, S. Bergler, M. Urseanu, All blogs are not made equal: Exploring genre differences in sentiment tagging of blogs., in: ICWSM, 2007.

[36] Y. Li, K. Bontcheva, H. Cunningham, Using uneven margin svm and perceptron for information extraction, in: Proceedings of the Conference on Natural Language Learning (CONLL'05), 2005, pp. 72–79.