Sentiment Analysis on Twitter through Topic-based Lexicon Expansion

Zhixin Zhou, Xiuzhen Zhang, and Mark Sanderson

Department of Computer Science and IT,
RMIT University, Melbourne, VIC 3000
{zhixin.zhou,xiuzhen.zhang,mark.sanderson}@rmit.edu.au
http://www.rmit.edu.au

Abstract. Supervised learning approaches are domain-dependent and it is costly to obtain labeled training data from different domains. Lexicon-based approaches enjoy stable performance across domains, but often cannot capture domain-dependent features. It is also hard for lexicon-based classifiers to identify the polarities of abbreviations and misspellings, which are common in short informal social text but usually not found in general sentiment lexicons. We propose to overcome this limitation by expanding a general lexicon with domain-dependent opinion words as well as abbreviations and informal opinion expressions. The expanded terms are automatically selected based on their mutual information with emoticons. As there is an abundant amount of emoticon-bearing tweets on Twitter, our approach provides a way to do domain-dependent sentiment analysis without the cost of data annotation. We show that our technique leads to statistically significant improvements in classification accuracies across 56 topics with a state-of-the-art lexicon-based classifier. We also present the expanded terms, and show the most representative opinion expressions obtained from co-occurrence with emoticons.

1 Introduction

Both machine-learning and lexicon-based approaches have been adopted to do sentiment analysis on Twitter. Machine-learning approaches to sentiment analysis usually require annotated text and are known to be domain-dependant. Annotation is generally costly to obtain, but lack of labeled data in the target domain can lead to deteriorated classification performance. It has also been shown in existing work [14] that better performance is achieved when using all words as features. As such, the feature list of a supervised classifier often contains spurious patterns which are difficult to make sense of by a human reader. As has been indicated by Thewall et.al [17], supervised classifiers may harness nonsentiment features and falsely identify sentiment.

Lexicon-based approaches rely on opinion lexicons to classify text. Words from such lexicons are direct indicators of sentiment and transparent to human readers, and the polarities of most opinion words are not domain-dependent. As such, these approaches can achieve stable performance across domains. However,

for more accurate classification, it is desirable to capture contextual polarities of words [3], especially so when dealing with short social text such as Twitter tweets, where single-sentence status updates are common and the number of features are rather limited. For example, the word "big" implies positive emotion in "My new office is BIG!!" but negative emotion in "it's too big to fit into your pocket". Another challenge for lexicon-based approaches on short social text is to identify the polarity of informal expressions of sentiments, such as abbreviations and misspellings. In a tweet positive expressions may include "+1", "hear hear". A negative expression of opinion can be "O come on", "lmao" (laugh my a** off). Such informal expressions may evolve and emerge over time, and to the best of our knowledge, no existing opinion word lexicon provides sufficient coverage of these informal expressions.

In addition to opinion lexicons, emotions can also be expressed through emoticons. Unlike literal words, emoticons usually have a stable polarity across domains and have been widely used for sentiment classification. In fact, the default twitter search allows users to add emoticons to the query to find positive or negative tweets, and the returned results usually contains emoticons. However, the majority of tweets do not have emoticons. Our statistics show that only 9.40% (7.37% positive and 2.03% negative) of the tweets in the Microblog Track 2011 collection have at least one emoticon[1].

Insufficient lexicon coverage and the limitations of using emoticons have motivated this study. We propose an automatic lexicon expansion technique to improve the coverage of the sentiment lexicon employed by the classifier, by measuring the mutual information between potentially sentiment-bearing words and the emoticons. We specifically study the following research questions,

1. Can emoticon-aided lexicon expansion improve the performance of a sentiment classifier?
2. Can topic-biased emoticon-aided lexicon expansion improve the performance of sentiment classification?

Our expansion technique is based on a simple intuition. For tweets that look like "+1 :)", or "hear hear :P", we may use the polarity implied by the emoticons ":)" and ":P" to infer that "+1" and "hear hear" are positive. Specifically, we use the point-wise mutual information (PMI) between each word (or symbol) and a known set of emoticons (see Table 1) to measure the sentiment polarity of the word. Our technique differs from Turney et.al [20] not only in that we use positive emoticon groups and negative emoticon groups as references for PMI calculation, but also in the way we deal with negation. In their study, negations in the text are not handled. As such, *"the online service was not excellent at all"* would be treated as evidence that *online service* is a positive phrase. In our appoach we apply negation detection mechanism to flip such cases so that *"I don't like their online service :("* would be counted as a co-occurrence of *like* and *:)*.

---

[1] We matched all tweets against the emoticons from Table 1 in this paper

To answer the second research question, we use hashtags to create a collection of tweets from 56 topics. Hashtags are a type of metadata used in Twitter community to add additional context to tweets, by prefixing a word with a hash symbol, such as *#twitter*. Tweets with the same hashtag can often be considered to be about the same coarse "topic", though in many cases not a topic in the common sense. For instance, hashtags such as *#justsaying* and *#nowplaying* are not typical topics. Nonetheless, a hashtag groups tweets of a similar concern, thus enforcing a coarse semantic relation between the tweets. Intuitively, topic-based lexicon expansion is more difficult when such semantic relations are weaker. Therefore, evaluation conducted on this hashtag-based topic collections should be more rigorous than one based on a human-labeled topic collection. We show that per-topic expansion leads to significantly better performance than global expansion done on the combined set of tweets from all topics.

2 Related Work

Sentiment classification models on Twitter can be broadly categorised as machine-learning and lexicon-based approaches, though some algorithms [22] have elements of both. Machine-learning approaches typically require labeled training data, and often use text features as well as emoticons as features to train the model. To reduce the cost of obtaining labeled training data, some sentiment-bearing tokens (e.g. emoticons, sometimes even hashtags) with known sentiment polarities have been used to automatically collect training instances. Go et.al [8] use emoticons as "noisy labels" to obtain instances, and classify with various classifiers, among which MaxEnt has achieved an accuracy of 83% on their test set. Pak et.al [12] also use emoticons as labels for training data and has built a multinomial Naïve Bayes classifier based on N-gram and POS-tags as features. Davidov et.al [6] utilize 50 Twitter tags and 15 smileys as sentiment labels. Liu et.al [9] utilize both manually labeled data and noisy labeled data for training, where emoticons are used to smoothen a supervised classification model.

Recently, lexicon-based approaches [16, 18] have gained popularity. Lexicon-based approaches can achieve stable performance across domains, and the features used are more transparent to a human user. They typically employ opinion word lexicons, such as SentiWordNet [1] and General Inquirer lexicon [5], to match against the text to be classified. The presence of annotated (typically with opinion scores or scales) lexical items (opinion words/phrases) are processed with linguistic rules to compute an overall semantic orientation of the document.When such a pre-defined lexicon is not present, other unsupervised methods [20] can be adopted to automatically construct a lexicon.

Various studies aim to solve the problem of adapting a machine-learning model for classification in a new domain. Such techniques are often referred to as transfering learning techniques[2], and the majority [21, 7, 4] of them require at least a small amount of labeled data from the target domain. The approaches

---

[2] The terminologies *domain adaptation* and *transfer learning* are often used interchangeably, and in this study we stick to the latter.

that do not require labeled data in the target domain include the Structural Correspondence Learning (SCL) algorithm by Blitzer et.al [3] and the dimensionality reduction approach by Pan et.al [13]. These approaches generally utilize the common features between the source domain and the new domain to establish a link for knowledge transfer. In [3] the link is built through the pivot features and in [13] the link is built through a latent space that minimizes the difference between different domain distributions.

To do domain-specific classification with lexicon-based classifiers several approaches have been developed. Ponomareva et.al [15] use a graph to model a group of labeled and unlabeled documents, and update the sentiment scores of unlabeled documents based on nearest documents. This approach operates at document level, therefore the classification process is no longer transparent to a human user. Domain-specific lexicon expansion is another way to adapt a lexicon-based classifier to different domains. The approach by Thelwall et.al [17] requires human intervention to annotate the corpus and do a small amount of term selection for different domains. Choi et.al [5] adapt an existing lexicon to a given domain through an optimization framework, where phrase-level subjectivity annotation is required. Turney et.al used two reference words, *excellent* and *poor*, to represent the two extremes of opinions, and use pointwise mutual information (PMI) to calculate the semantic orientation of words. Becker's work [2] is the most related to ours. Their study is also built around Turney's idea [20] of using PMI to construct a lexicon, but instead of computing PMI between a word and *reference words*, they compute the PMI between a word and *sentiments*. The sentiments are tagged by a polarity classifier before PMI calculation.

While emoticons have been used in sentiment analysis on twitter [8, 12, 6, 9], and PMI calculation has been used to construct opinion word lexicons [20, 2], no existing work uses emoticon as reference tokens to do domain-specific PMI-based expansion. Also, when computing PMI, existing studies do not consider negation handling, therefore the cooccurrence of a negated word/phrase and a reference word [20] (or sentiments [2]) will give misleading information on the polarity of the word.

3   Emoticon-based Sentiment Lexicon Expansion

3.1   Classification Framework

While we focus on lexicon-based classifiers in this study due to the stability of their performance across domains and the transparency of the classification process, the classification results of a machine learning approach is also presented as a reference.

SentiStrength [18] (SS) is a state-of-the-art lexicon-based classifier. Incorporating a booster word list, an emoticon list, an idiom list, a negation word list, a question word list, a slang list and a general opinion word list, it further applies linguistic rules to compute the overal sentiment polarity. The opinion word scores are integers ranging from [-5, -1] for negative words and [1,5] for positive

words, where -1 and 1 denotes neutral words. The core of the SentiStrength lexicon is a general opinion word list of 298 positive and 465 negative terms, some of which include wild cards. For example, *abandon** would match all words that start with *abandon*. For details of its algorithm please refer to [19]. SentiStrength is designed to report binary (positive or negative), trinary (positive, negative or neutral) and single scale (-4 to +4) results. We use its binary output, as our collection comprising of positive and negative tweets. Before expanding the SentiStrength lexicon, linear scaling has been performed to transform the semantic orientation score calculated from PMI to the SentiStrength scales. We merge our expanded terms with the SentiStrength terms without further modification. We disabled the emoticon word list in SentiStrength, since we are using the emoticons as class labels. In our pilot study, SentiStrength has achieved 90.32% when using emoticons, but its performance dropped to 76.92% when the emoticon list was disabled.

The Naïve Bayes Multinomial (NBM) classifier uses all words as features without stemming, nor normalization to lower case. Naïve Bayes has achieved superior performance at a per-topic average of 93.15%. Despite the high accuracy, the features used are hard to make sense of. In our experiments, NBM is confined to use only opinion words as features to make a relative fair comparison. The original feature list includes the words from SentiWordnet [1], which comprises of 21,109 opinion words, many times larger than the opinion lexicon used by SentiStrength.

3.2   Lexicon Expansion

For lexicon-based classification on Twitter, a widely acknowledged problem is the word mismatch between tweet content and general opinion word lexicons. For example, words with repeated letters are commonly seen in tweets, such as *huuuungry*. Some studies [8] replace all repeated letters with two repeated letters in each word, and leave it to the classifier to leverage these features. While being effective in this particular case, this approach is insufficient to tackle many other forms of informal spellings, such as *gr8t*. In fact, these informal expressions (including informal spellings and abbreviations) are evolving over time, making it difficult for any rule-based approach to adapt to the changes.

In this study, we propose to use mutual information between sentiment tokens[3] and emoticons to adjust the sentiment strength scores of words in a general sentiment lexicon and add unseen variations of sentiment words into the lexicon. There are an abundant amount of emoticon-bearing tweets on Twitter, forming the basis of our technique. With our expansion technique, the polarities of the informal expressions as well as the less common emoticons can be automatically computed.

---

[3] A token can be a word, an abbreviation, an emoticon not found in our emoticon list shown in Table 1, or any other text segment potentially bearing sentiment

Emoticon	Polarity	Emoticon	Polarity	Emoticon	Polarity	Emoticon	Polarity
:)	+	:-)	+	;(	-	:/	-
:>	+	;)	+	:(	-	:-(	-
;-)	+	;>	+	:[	-	:<	-
B-)	+	8-)	+	8-(	-	:-o	-
B->	+	8->	+	:-&	-	:/)	-
:->	+	:-)))	+	:-c	-	:-C	-
:D	+	:-D	+	:-<	-	;-C	-
:-P	+	^_^	+	:-|	-	:'-(	-
^.^	+	^_^'	+	:~-(	-	-_-	-
D=	+	:p	+	>_<	-	=_=	-
;D	+						

Table 1. List of Emoticons used (Manually selected from http://en.wikipedia.org/wiki/List_of_emoticons, based on frequency and clarity)

The following formula is used to calculate the point-wise mutual information (PMI) between any two words,

$$PMI(word_1, word_2) = log_2 \left(\frac{p(word_1, word_2)}{p(word_1)p(word_2)} \right) \quad (1)$$

The semantic orientation (SO) of a word is given by Equation 2,

$$SO(token) = PMI(token, +ve) - PMI(token, -ve) \quad (2)$$

where $+ve$ and $-ve$ represents positive and negative emoticons respectively.

Our lexicon expansion technique is based upon the following assumption: The sentiment orientations of emoticons such as ":)" and ":(" are relatively stable across all tweets. Therefore, positive and negative emoticons can then be used to represent the two extremes of opinions. Turney's approach also implicitly assumes that words in the context of the reference word tend to share its polarity. This second assumption is not valid in sentences with negations, for example, "I don't like this guy:(". As such, in tweets with negation words, we flip the polarity of the emoticon before calculating mutual information. Also, we map all negative emoticons to ":(" and all positive emoticons to ":)". The sentiment orientation (SO) of any sentiment token is then calculated as,

$$SO(token) = log_2 \left(\frac{hits(token, +ve) \times hits(-ve)}{hits(token, -ve) \times hits(+ve)} \right) \quad (3)$$

In Equation 3, $hits(+ve)$ and $hits(-ve)$ are global counts of positive and negative emoticons respectively. $hits(token, +ve)$ represents the number of co-occurrences of the token and positive emoticons, or the token and negative emoticons in the presence of negation. In the above-mentioned example, the co-occurrence of *like* and *:(* is counted as a hit between *like* and *:)* due to the presence of the negation word *don't*. The list of regular expression patterns for negation detection are shown in Table 2.

Pattern	Example	Frequency
`[Nn][Oo][Tt]`	not	3.83%
`[Dd][Oo][Nn]'*[Oo]*[Tt]+`	don't	3.10%
`[Cc][Aa][Nn]+'*[Oo]*[Tt]+`	can't	1.70%
`[Aa][Ii][Nn]'*[Tt]`	ain't	0.97%
`[Ii][Ss][Nn]'*[Oo]*[Tt]`	isn't	0.30%
`[Hh][Aa][Vv][Ee][Nn]'*[Oo]*[Tt]`	haven't	0.29%
`[Ww][Oo][Uu][Ll][Dd][Nn]'*[Tt]`	wouldn't	0.20%
`[Cc][Oo][Uu][Ll][Dd][Nn]'*[Tt]`	couldn't	0.16%
`[Hh][Aa][Ss][Nn]'*[Oo]*[Tt]`	hasn't	0.05%
`[Bb][Aa][Rr][Ee][Ll][Yy]`	barely	0.03%
`[Hh][Aa][Rr][Dd][Ll][Yy]`	hardly	0.02%
`[Hh][Aa][Dd][Nn]'*[Oo]*[Tt]`	hadn't	0.02%
`[Ww][Uu][Dd][Nn][Tt]`	wudn't	0.00%
`[Cc][Uu][Dd][Nn][Tt]`	cudn't	0.00%

Table 2. Stats on negation patterns collected on the whole Microblog Track collection. Column Frequency shows the percentage of tweets in which the pattern has a match

Not all words bear sentiments. We use the part-of-speech tags of words to filter out potential opinion words and only include nouns, adjectives, verbs, adverbs, abbreviations emoticons and interjections. Part-of-speech tagging is done with the TwitterNLP package [11]. This set of tags was chosen after experimenting with differnt combinations of tag sets in preliminary experiments, and using this set to do classification has led to high classification accuracy, while still keeping meaningful words.

4   Experiment

4.1   Dataset

Our experiments are based on a 56-topic collection of tweets from the TREC Microblog Track 2011 and the Stanford Sentiment140 collection. This dataset is generated by aggregating opinion-bearing tweets with 56 popular hashtags in the combined collection. The tweets do not come with labels – we use the emoticons as their labels, as has been done in previous studies. In our experiments on this dataset, 10-fold cross validation is carried out and the lexicon expansion was done on the training set only. The emoticons used for lexicon expansion are not used as features by any classifier in the classification process. The collection is generated via three steps,

1. Merge all tweets from TREC Microblog Track 2011 collection [10] and the Stanford Sentiment 140 collection [8].
2. Filter out all tweets that do not bear any emoticon from Table 1.
3. Group the remaining tweets by hashtag and keep the groups with at least 100 tweets.

Topic	# Pos	# Neg	Topic	# Pos	# Neg	Topic	# Pos	# Neg
glee	99	17	follow	255	14	nowfollowing	165	0
2	214	152	followfriday	2529	143	agoodboyfriend	119	6
bgt	133	114	bieberd3d	484	166	neversaynever	484	74
e3	119	136	asot400	190	135	nowwatching	131	5
fb	934	1101	februarywish	228	25	oneofmyfollowers	109	51
1	425	187	followback	111	3	iranelection	79	432
f1	122	152	iphone	96	236	icantdateyou	93	79
humor	143	1	improudtosay	395	16	idontunderstandwhy	48	82
fml	10	99	iremember	153	150	jedwardlipstick	125	16
jfb	379	9	justsaying	103	13	marsiscoming	205	54
music	162	15	musicmonday	383	38	myweakness	217	52
bsb	131	164	nowplaying	590	70	neversaynever3d	456	37
fail	72	400	random	90	30	spymaster	99	70
ff	2098	130	purpleglasses	189	61	questionsidontlike	16	89
np	606	72	seb-day	428	71	shoutout	464	29
nw	139	12	squarespace	282	656	superbowl	233	58
tcot	140	34	twitteroff	152	22	teamfollowback	499	22
trackle	16	145	twitition	193	2	twitter	111	75
tfb	117	2	wheniwaslittle	120	75			

Table 3. Statistics of the 56-topic collection

The whole collection includes 16,683 positive and 6,099 negative tweets. Full details of the collection are shown in Table 3.

4.2   General Lexicon vs. Global Expansion

In this experiment we aim to answer the first research question by contrasting the performance of SentiStrength (SS) and Naïve Bayes Multinomial (NBM) before and after doing global expansion (GE). With GE, tweets from all topics are merged into a single collection, on which lexicon expansion is based upon. The expanded lexicon is then used to do per-topic classification. For NBM, words from SentiWordNet are used as the original feature list, and SS uses its own lexicon. After expansion, the expanded terms are added to the feature list of NBM, and also added to the lexicon of SS.

The classification results are shown in Table 4. A paired t-test was done across the 56 topics to show statistically significant improvement with SentiStrength, with a p value of 1.6e-5. This indicates that PMI-based lexcion expansion does indeed lead to a better lexicon for classification with SentiStrength. With NBM however, the classification performance after global lexicon expansion is higher but not statistically significant.

4.3   Topic-based Expansion vs. Global Expansion

We contrast the performance of SS and NBM with Global Expansion (GE) against Topic-based Expansion (TE) to answer the second research question. As

Classifier	No Expansion (%)			With GE (%)			With TE (%)		
	All	+	-	All	+	-	All	+	-
NBM	82.38	83.77	38.62	82.61	83.43	41.50	82.72	83.62	40.02
SS	76.92	84.33	44.77	**79.99**	**87.51**	**48.99**	**85.24**	**88.29**	**58.03**

Table 4. Classification with no lexicon expansion, Global Expansion and Topic-based Expansion. Columns *All, + and* - shows the overall accuracy, positive precision and negative precision respectively. NBM refers to Naïve Bayes Multinomial, and SS refers to SentiStrength.

is shown in Table 4, topic-based expansion has lead to a significantly better clas-sifcation results with SentiStrength. A paired t-test done across 56 topics shows significant improvement with a p value of 3.97e-9. In fact, the performance of SentiStrength with topic-based expansion has even exceeded NBM with statistical significance (p value = 1.69 e-45 against NBM with GE and 2.89 e-45 agaist NBM with TE). Note though, NBM and SS are not using the same feature list. NBM is using an opinion word lexicon of more than 21,109 words while the size of the SS's core lexicon is only around 2,000 words.

Figure 1 shows the similarity between expanded terms from different topics. Cosine similarity was measured between each pair of topics to compare how similar the 50 most opinionated terms are. The subjectivity of a term is evaluated by PMI. As is shown in the heatmap, the terms are quite different. The highest similarity was found between *justsaying* and *agoodboyfriend*, with a similarity score of only 0.47, which indicates highly divergent extensions to the original lexicon.
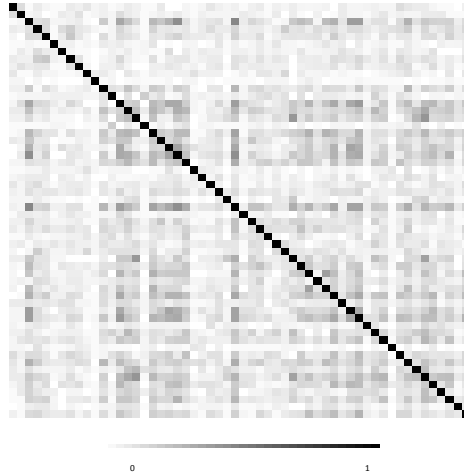

Fig. 1. Cosine similarity between the expanded terms (top 50) from each topic. The topics on the X axis (from left to right) and Y axis (from top to bottom) are shown in the same order as in Table 3

Topic	Improvement	#pos	#neg	Topic	Improvement	#pos	#neg
questionsidontlike	+88.97%	16	89	music	-14.07%	165	15
fml	+42.77%	10	99	follow	-9.94%	255	14
iphone	+40.78%	96	236	humor	-7.69%	143	1
trackle	+36.12%	16	145	musicmonday	-3.62%	383	38
e3	+34.34%	119	136	tfb	-3.40%	117	2
fail	+31.72%	72	400	neversaynever3d	-2.69%	484	74
spymaster	+29.51%	99	70	idontunderstandwhy	-2.17%	48	82
squarespace	+26.06%	282	656	seb-day	-1.04%	428	71
tcot	+26.06%	140	34	random	-0.71%	90	30
justsaying	+24.25%	103	13	nowfollowing	-0.45%	165	0

Table 5. The 10 topics that benefit most and the 10 topics that suffer most from lexicon expansion.

Topics on which classification performance have been affected most by topic-based expansion, both negatively and positively, are shown in Table 5. Abbreviations like *fml* (f*** my life), *tcot* (The changing of times), *tfb* (teamfollowback are not self-explanatory. Both conventional topics such as *iphone* and *e3*, and hashbag-based topics such as *questionsidontlike* and *fml* benefit hugely from topic-based lexicon expansion. Among the topics that suffer most, both conventional topics (*music*) and hashtag-based topics (*tfb*) are found too.

Table 6 provides two samples of lexicon expansions. As is shown in the table, our expansion technique can effectively identify the polarities of 1) abbreviations. For instance, *lmao* (laugh my a** off), *lmfao* (laugh my f***ing a** off), *smh* (shaking my head) are associated with scores of the proper polarities. 2) informal words such as *yup*, *ugh*. 3) topic-specific words including *gay* from topic *icantdateyou* and *making, want* from topic *iphone*. When mentioning *gay* in topic *icantdateyou*, people are often making jokes, for example: *@MissSarahDan hahahahaha that's exactly what I was thinking :) or #icantdateyou cause you're gay! Lol XxX*. The word *making* would be considered neutral in most cases, but is often mentioned in the contexts such as *making it hard to look at any screen* and *why are you making me wait until noon for #iPhone OS 3.0?* in the topic *iphone*, thus having a strongly negative PMI score. Similarly, the word *want* is frequently used in the pattern of *I want ..., but ...*, showing a need that has not been satisfied. As such, the seemingly neutral word has been assigned a strongly negative score, becoming a useful feature in the context of topic *iphone*.

Finally, we contrast our negation-aware PMI expansion to traditional PMI expansion that does not detect negations. As is shown in Table 7, no statistically significant difference have been observed between the two runs, though the average accuracy is slightly higher when using negation handling. The number of tweets per topic in this collection is limited compared to the data that a commercial company may own, which may be the reason why the effect of negation-aware PMI expansion is not clearly shown.

Word	PMI	Word	PMI	Word	PMI	Word	PMI
expanded terms for *icantdateyou*				expanded terms for *iphone*			
love	19.27	wear	-18.83	cant	20.66	want	-20.36
dont	-18.50	smarter	-18.50	cool	20.17	seems	-19.55
haha	18.27	lmao	18.27	guess	-19.55	yup	19.43
URLSTRING	18.27	talk	-18.09	explain	19.43	worked	19.43
look	-18.09	gay	17.85	care	19.43	finish	19.43
can't	-2.24	how	-2.24	says	-19.36	didn't	-19.36
go	-2.24	am	2.09	being	-18.87	work	-18.87
don't	-1.88	have	-1.82	making	-18.87	said	-18.55
got	-1.82	call	-1.82	trying	-18.55	crashes	-18.55
date	-1.82	still	-1.82	already	-18.55	hmm	-18.55
will	1.76	want	1.76	seriously	-18.55	wants	-18.55
do	1.57	know	-1.56	keeps	-18.55	went	-18.14
being	1.35	back	1.35	gonna	-18.14	oh	-18.14
when	-1.24	even	-1.24	small	-18.14	crashing	-18.14
been	-1.24	only	-1.24	let	-18.14	cut	-18.14
wanna	-1.24	waiting	-1.24	where	-18.14	looks	-18.14
try	-1.24	play	-1.24	would	-18.14	means	-18.14
come	-1.24	see	-1.24	gone	-18.14	ugh	-18.14
smh	-1.24	be	1.18	lost	-18.14	stuck	-18.14
get	1.09	not	-1.01	feel	-18.14	omg	-18.14
are	-0.97	already	-0.82	not	6.00	can't	3.76
is	-0.77	then	0.76	don't	2.88	looking	2.88
oh	0.76	cant	0.76	happy	2.88	think	2.88
lmfao	0.76	make	0.76	find	2.88	can	2.52
aint	0.76	bigger	0.76	did	2.30	again	2.30

Table 6. Expanded term samples. Top 50 from two topics ordered by the absolute values of PMI

5   Conclusions

In this study we propose a structured approach to domain-dependent sentiment analysis through lexicon expansion aided by emoticons. Our experiments have shown that emoticon-aided lexicon expansion does improve the performance of a state-of-the-art lexicon-based classifier, and topic-based expansion outperforms global expansion. This indicates that our technique is effective in domain-dependant sentiment classification.

It has been observed from the experiments that our technique is able to identify the polarities of abbreviations and informal expressions, as well as topic-specific words, making it particularly useful in classifying short informal text such as tweets. The abundant amount of emoticon-bearing tweets forms a solid basis of the application of our approach. With our approach, emoticons are only used in the expansion phase. With the better coverage of terms and expressions in social text, our approach is particularly useful in improving the classification of tweets that do not contain emoticons.

	Handle Negation (%)			Ignore Negation (%)		
Classifier	All	+	-	All	+	-
SS	85.24	88.29	58.03	84.95	85.73	71.91

Table 7. PMI-based expansion with and without negation detection. Columns *All, +* *and -* shows the overall accuracy, positive precision and negative precision respectively

References

1. S. Baccianella et.al. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, 2010.
2. L. Becker et.al AVAYA: sentiment analysis on twitter with self-training and polarity lexicon expansion. In *SemEval*, 2013.
3. J. Blitzer et.al Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, 2007.
4. E. Bonilla et.al Multi-task gaussian process prediction. 2008.
5. Y. Choi and C. Cardie. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *EMNLP*, 2009.
6. D. Davidov et.al Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010*, 2010.
7. J. Davis and P. Domingos. Deep transfer via second-order markov logic. In *ICML*, 2009.
8. A. Go et.al Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 2009.
9. K.L. Liu et.al Emoticon smoothed language models for twitter sentiment analysis. In *AAAI*, 2012.
10. I. Ounis et.al Overview of the trec-2011 microblog track. In *TREC 2011*, 2011.
11. O. Owoputi et.al Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of NAACL-HLT*, 2013.
12. A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, 2010.
13. S. J. Pan et.al Transfer learning via dimensionality reduction. In *AAAI*, 2008.
14. B. Pang et.al Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*, 2002.
15. N. Ponomareva and M. Thelwall. Do neighbours help?: an exploration of graph-based algorithms for cross-domain sentiment classification. In *Proceedings of the 2012 Joint Conference on EMNLP and CoNLL*, 2012.
16. M. Taboada et.al Lexicon-based methods for sentiment analysis. *Computational linguistics*, 2011.
17. M. Thelwall and K. Buckley. Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *JASIST*, 2013.
18. M. Thelwall et.al Sentiment strength detection for the social web. *JASIST*, 2012.
19. M. Thelwall et.al Sentiment strength detection in short informal text. *JASIST*, 2010.
20. P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL*, 2002.
21. D. Zhang et.al Sentiment detection with auxiliary data. *Information retrieval*, 2012.
22. L. Zhang et.al Combining lexiconbased and learning-based methods for twitter sentiment analysis. *HP Laboratories, Technical Report HPL-2011*, 2011.