AUTHOR(S):

TITLE:

YEAR:

Publisher citation:

OpenAIR citation:

Emotion-Aware Polarity Lexicons for Twitter Sentiment Analysis

Anil Bandhakavi, Nirmalie Wiratunga, Stewart Massie and Deepak P.

Abstract Theoretical frameworks in psychology map the relationships between emotions and sentiments. In this paper we study the role of such mapping for computational emotion detection from text (e.g. social media) with a aim to understand the usefulness of an emotion-rich corpus of documents (e.g. tweets) to learn polarity lexicons for sentiment analysis. We propose two different methods that leverage a corpus of emotion-labelled tweets to learn word-polarity lexicons. The proposed methods model the emotion corpus using a generative unigram mixture model (UMM), combined with the emotion-sentiment mapping proposed in Psychology for automated generation of word-polarity lexicons that capture emotion-rich vocabulary. We comparatively evaluate the quality of the proposed mixture model in learning emotion-aware sentiment lexicons with those generated using supervised latent dirichlet allocation (sLDA) and word-document frequency (WDF) statistics. Sentiment analysis experiments on benchmark Twitter data sets confirm the quality of our proposed lexicons. Further a comparative analysis with sLDA, WDF based emotion-aware lexicons and standard sentiment lexicons that are agnostic to emotion knowledge suggest that the proposed lexicons lead to a significantly better performance in both sentiment classification and sentiment intensity prediction tasks.

1 Introduction

Sentiment analysis concerns the computational study of natural language text (e.g. words, sentences and documents) in order to identify and effectively quantify its polarity (i.e positive or negative) [28]. Sentiment lexicons are the most popular resources used for sentiment analysis, since they capture the polarity of a large col-

Anil Bandhakavi, Nirmalie Wiratunga, Stewart Massie

School of Computing, Robert Gordon University, Aberdeen, UK, e-mail: a.s.bandhakavi, n.wiratunga, s.massie @rgu.ac.uk

Deepak P.

Queen's University, Belfast, UK, e-mail: deepaksp@acm.org

lection of words. These lexicons are either hand-crafted (e.g. opinion lexicon [15], General Inquirer [36] and MPQA subjectivity lexicon [38]) or generated (e.g. Senti-WordNet [9] and SenticNet [7]) using linguistic resources such as WordNet [10] and ConceptNet [20]. However, on social media (e.g. Twitter), text contains special symbols resulting in non-standard spellings, punctuations and capitalization; sequence of repeating characters and emoticons for which the aforementioned lexicons have limited or no coverage.

As a result domain-specific sentiment lexicons were developed to capture the informal and creative expressions used on social media to convey sentiment [24, 11]. The extraction of such lexicons is possible with limited effort, due to the abundance of weakly-labelled sentiment data on social media, obtained using emoticons [13, 14]. However, sentiment on social media is not limited to conveying positivity and negativity. Socio-linguistics suggest that on social media, people express a wide range of emotions such as *anger, fear, joy, sadness* etc [6]. Following the trends in lexicon based sentiment analysis, research in the textual emotion detection also developed lexicons that can not only capture the emotional orientation of words [25, 31], but also quantify their emotional intensity [35, 33].

Though research in psychology defines sentiment and emotion differently [26], it also provides a relationship between them [4]. Further research in emotion classification [37, 12] demonstrated the usefulness of sentiment features extracted using a lexicon for document representation. Similarly emoticons used as features to represent documents improved sentiment classification [14, 24]. However, the exploration of emotion knowledge for sentiment analysis is limited to emoticons [14, 16, 17], leaving a host of creative expressions such as emotional hashtags (e.g. #loveisbliss), elongated words (e.g. haaaappyy!!!) and their concatenated variants unexplored. An emotion-corpus crawled on Twitter using seed words for different emotions as in [23, 37] can potentially serve as a knowledge resource for sentiment analysis. Adopting such corpora for sentiment analysis, e.g. sentiment lexicon extraction is particularly interesting, given the challenges involved in developing effective models which can cope with the lexical variations on social media.

Therefore, in this work we explore the role of a Twitter emotion corpus for extracting a sentiment lexicon, which can be used to analyse the sentiment of tweets. We do a qualitative comparison between standard sentiment lexicons that are agnostic to the emotion-knowledge, emotion-aware sentiment lexicons generated using techniques such as supervised latent dirichlet allocation (slDA) and the proposed sentiment lexicons. Our contributions in this paper are as follows:

1. We propose two different methods to generate sentiment lexicons from a corpus of emotion-labelled tweets by combining our prior work on domain-specific emotion lexicon generation [1, 2], with the emotion-sentiment mapping presented in Psychology (see figure 1) [4]; and
2. We comparatively evaluate the quality of the proposed sentiment lexicons, emotion-aware sentiment lexicons learnt using sLDA and the standard sentiment lexicons found in literature through different sentiment analysis tasks: *sentiment intensity prediction and sentiment classification* on benchmark Twitter data sets.

In the rest of the paper we review related literature in Section 2. In Sections 3 and 4 we formulate the methods to extract sentiment lexicons from an emotion corpus of tweets. Section 5 presents the baseline methods to extract sentiment lexicons from an emotion corpus of tweets. In Section 6 we describe our experimental set up and analyse the results. Section 7 presents our conclusions.

2 Related Work

In this section we review the literature concerning sentiment lexicons, followed by a review of different emotion theories and their relationship with sentiments proposed in Psychology.

2.1 *Lexicons for Sentiment Analysis*

Broadly sentiment lexicons are of two types: *hand-crafted and automatic*. Hand-crafted lexicons such as opinion lexicon [15], General Inquirer [36] and MPQA subjectivity lexicon [38] have human assigned sentiment scores. On the other hand automatic lexicons are of two types: *corpus-based and resource-based*. Lexicons such as SentiWordNet [9] and SenticNet [7] are resource-based, since they are extracted using linguistic resources such as WordNet [10] and ConceptNet [20].

A common limitation of resource-based and hand-crafted lexicons is that, they have static vocabulary, making them limitedly effective to mine sentiment on social media, which is inherently dynamic. Corpus-based lexicons such as in [24, 11], gauge the corpus level variations in sentiment using statistical models, and are found to be very effective on social media. Further with the abundance of weakly-labelled sentiment data on social media, these lexicons can be updated with very low costs. Similarly research in emotion analysis lead to the development of resource-based [25, 31] and corpus-based emotion lexicons [35, 33].

Prior research in sentiment analysis developed models that exploit emotion knowledge, such as emoticons to gain performance improvements [14, 16, 17]. However, other forms of emotion knowledge such as an emotion corpus and the lexicons learnt from it, could potentially have richer sentiment-relevant information, compared to that of emoticons. Therefore it is interesting to study role of such emotion knowledge for sentiment analysis, in particular for sentiment lexicon generation and validate its usefulness. Our work focusses on this aspect, by exploiting an emotion-labelled corpus of tweets to learn sentiment lexicons. We achieve this by combining our prior work on generative mixture models for lexicon extraction and the emotion-sentiment mapping provided in psychology.

2.2 *Emotion Theories*

Research in psychology proposed many emotion theories, wherein each theory organizes a set of emotions into some structural form (e.g. taxonomy). In the following sections we detail the most popular emotion theories studied in psychology.

2.2.1 Ekman Emotion Theory

Paul Ekman, an American psychologist focused on identifying the most basic set of emotions that can be expressed distinctly in the form of a facial expression. The emotions identified as basic by Ekman are *anger, fear, joy, sadness, surprise and disgust* [8].

2.2.2 Plutchik's Emotion Theory

Unlike the Ekman emotion model Plutchik's emotion model defines eight basic emotions such as *anger, anticipation, disgust, joy, fear, sadness and surprise* [30]. These basic emotions are arranged as bipolar pairs namely: *joy-sadness, trust-disgust, fear-anger, surprise-anticipation*.

2.2.3 Parrot's Emotion Theory

Parrot organised emotions in a three level hierarchical structure [29]. The levels represent primary, secondary and tertiary emotions respectively. Parrot identified emotions such as *love, joy, surprise, anger, sadness and fear*, as the primary emotions. Though Ekman and Plutchik emotion models are popular, research in Twitter emotion detection [37], [32] focussed on emotions that largely overlap with that of Parrot, given their popular expressiveness on social media. We use the Parrot emotion-labelled twitter corpus [37] in this study for generating sentiment lexicons.

2.2.4 Emotion-Sentiment Relationship in Psychology

One of the popular approaches for emotion modelling in Psychology is the dimensional approach, wherein each emotion is considered as a point in the continuous multidimensional space where each aspect or characteristic of an emotion is represented as a dimension. Affect variability is captured by two dimensions namely valence and arousal [18]. Valence (pleasure - displeasure) depicts the degree of positivity or negativity of an emotion. Arousal (activation- deactivation) depicts the excitement or the strength of an emotion. The dimensional approach depicting parrot's primary emotions in the valence arousal 2D space is shown in Figure 1 [3].

3 Emotion-Aware Models for Sentiment Analysis

In this section we formulate two different methods which utilize a corpus of emotion-labelled documents for sentiment analysis of text. The first method learns an emotion lexicon and further transforms it into a sentiment lexicon using the emotion-sentiment mapping (refer section 2.2) proposed in Psychology. The second method on the other hand learns the sentiment labels for the documents in the emotion corpus using the emotion-sentiment mapping, followed by a sentiment lexicon extraction. The two proposed methods are illustrated visually in figures 2a and 2b.
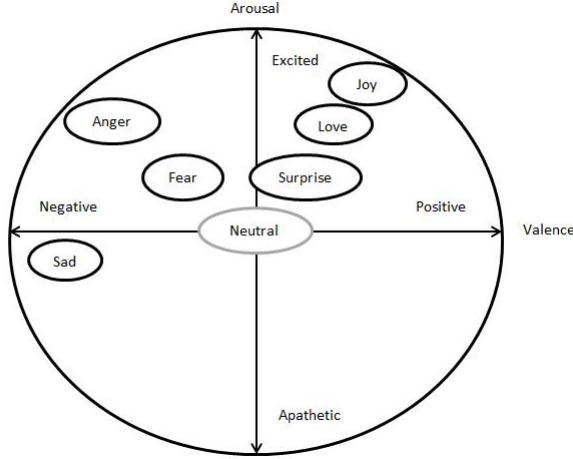
Fig. 1: Parrot's emotions in the valence-arousal plane of the dimensional model

3.1 *Emotion Corpus-EmoSentiLex*

A simple way to utilize a corpus of emotion-labelled documents, X_E for sentiment analysis is to first learn an emotion lexicon, and further transform it into a sentiment lexicon. An emotion lexicon UMM_EmoLex in our case is a $|V| \times (k+1)$ matrix, where $UMM\_EmoLex(i, j)$ is the emotional valence of the ith word in vocabulary V to the jth emotion in E (set of k emotions) and $UMM\_EmoLex(i, k+1)$ corresponds to its neutral valence (refer section 4). Observe that k emotions are considered and the $k+1$ is the neutral class. Further using the emotion-sentiment mapping proposed in Psychology we transform the emotion lexicon UMM_EmoLex into a sentiment lexicon $UMM\_EmoSentiLex$, which is a $|V| \times 1$ matrix as follows:

$$UMM\_EmoSentiLex(i) = Log \left(\frac{\sum_{m \in E^+} UMM\_EmoLex(i, m)}{\sum_{n \in E^-} UMM\_Emolex(i, n)} \right) \qquad (1)$$

where $E^+ \subset E$ and $E^- \subset E$ are the set of positive and negative emotions according to the emotion-sentiment mapping. Further m and n are iterators over the set of positive and negative emotions. Note that the log scoring assigns a positive value for words having stronger associations with emotions such as *Joy, Surprise and Love* and negative values for words having stronger associations with emotions such as *Anger, Sadness and Fear*. Therefore we expect that sentiment knowledge for words is implicitly captured in an emotion lexicon, which can be easily extracted using this simple transformation.

Using the above method, any automatically generated emotion lexicon can be converted into a sentiment lexicon. For example automatic emotion lexicons learnt from a corpus of emotion labelled tweets using methods such as latent dirichlet allocation (LDA) can be used to learn emotion-aware sentiment lexicons. We refer

(a) Emotion Corpus-EmoSentiLex
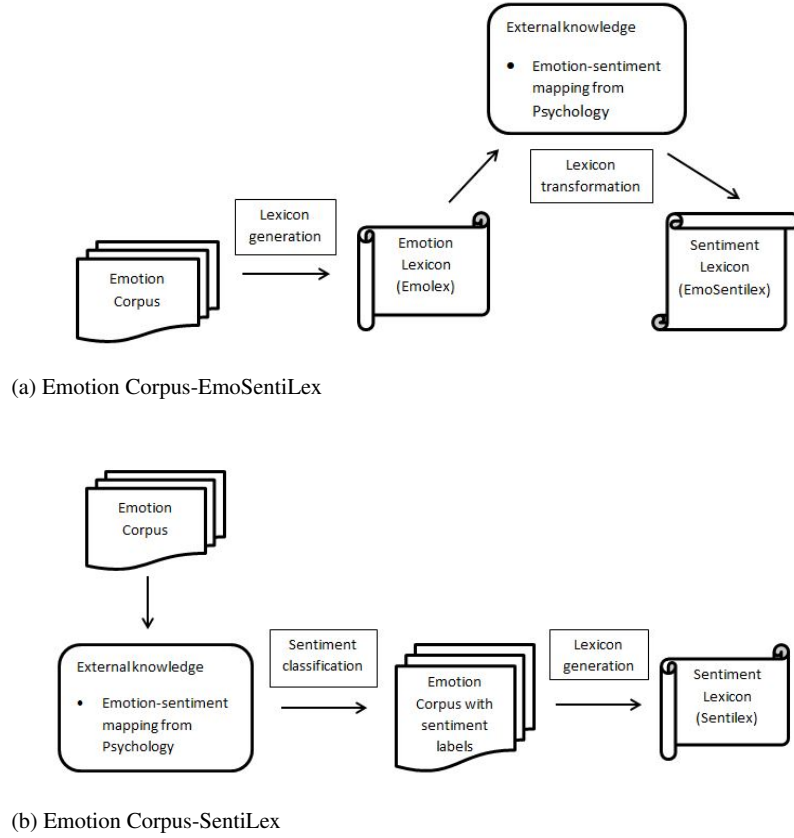

(b) Emotion Corpus-SentiLex

Fig. 2: Emotion-Aware Models for Sentiment Analysis

the readers to section 5 for the lexicon generation process using LDA and word-frequency statistics. Though the above method induces a sentiment lexicon it does not model the document-sentiment relationships to learn the lexicon, which is important to quantify word-sentiment associations. Therefore we introduce an alternate method which overcomes this limitation while utilizing an emotion corpus for sentiment lexicon generation.

3.2 *Emotion Corpus-SentiLex*

An alternate way to utilize the emotion corpus, X_E for sentiment analysis is to transform it into a sentiment corpus, X_S by learning the sentiment label for each document $d \in X_E$. This is done by using the emotion-sentiment mapping as follows:

$$Sentiment(d) = \begin{cases} positive & \text{if emotion(d)} \in E^+ \\ negative & \text{if emotion(d)} \in E^- \end{cases} \tag{2}$$

The sentiment lexicon *UMM_SentiLex* learnt from the corpus X_S is a $|V| \times 3$ matrix, where *UMM_SentiLex*$(i,1)$, *UMM_SentiLex*$(i,2)$ and *UMM_SentiLex*$(i,3)$ are the positive, negative and neutral valences corresponding to the ith word in vocabulary V. Observe that unlike the method which learns *UMM_EmoSentiLex*, by aggregating word-level emotion scores into sentiment scores, this method learns the sentiment-class knowledge corresponding to the documents, before learning a word-sentiment lexicon. We expect this additional layer of supervision, to benefit performance, following the findings of earlier research in supervised and unsupervised sentiment analysis. In the following section we briefly explain our proposed method to generate *UMM_SentiLex* and *UMM_EmoLex*. Further details about our proposed method can be found in [1, 2]

4 Mixture Model for Lexicon Generation

In this section we describe our proposed unigram mixture model (UMM) applied to the task of emotion lexicon (*UMM_EmoLex*) generation. Sentiment lexicon (*UMM_SentiLex*) generation is a special case of emotion lexicon generation, where the k emotion classes are reduced to positive and negative classes. Therefore we continue the presentation for the general case, i.e. *UMM_EmoLex* generation.

We model real-world emotion data to be a mixture of emotion bearing words and emotion-neutral (background) words. For example consider the tweet *going to Paris this Saturday #elated #joyous*, which explicitly connotes emotion *joy*. However, the word *Saturday* is evidently not indicative of *joy*. Further *Paris* could be associated with emotions such as *love*. Therefore we propose a generative model which assumes a mixture of two unigram language models to account for such word mixtures in documents. More formally our generative model is as follows to describe the generation of documents connoting emotion e_t:

$$P(D_{e_t}, Z | \theta_{e_t}) = \prod_{i=1}^{|D_{e_t}|} \prod_{w \in d_i} [(1 - Z_w)\lambda_{e_t} P(w|\theta_{e_t})$$
$$+ (Z_w)(1 - \lambda_{e_t})P(w|N)]^{c(w,d_i)} \quad (3)$$

where θ_{e_t} is the emotion language model and N is the background language model. λ_{e_t} is the mixture parameter and Z_w is a binary hidden variable which indicates the language model that generated the word w.

The estimation of parameters θ_{e_t} and Z can be done using expectation maximization (EM), which iteratively maximizes the complete data (D_{e_t}, Z) by alternating between E-step and M-step. The E and M steps in our case are as follows:
E-step:

$$P(Z_w = 0 | D_{e_t}, \theta_{e_t}^{(n)}) = \frac{\lambda_{e_t} P(w|\theta_{e_t}^{(n)})}{\lambda_{e_t} P(w|\theta_{e_t}^{(n)}) + (1 - \lambda_{e_t})P(w|N)} \quad (4)$$

M-step:

$$P(w|\theta_{\theta_{e_t}}^{(n+1)}) = \frac{\sum_{i=1}^{|D_{e_t}|} P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)})c(w, d_i)}{\sum_{w \in V} \sum_{i=1}^{|D_{e_t}|} P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)})c(w, d_i)} \qquad (5)$$

where n indicates the EM iteration number. The EM iterations are terminated when an optimal estimate for the emotion language model θ_{e_t} is obtained. EM is used to estimate the parameters of the k mixture models corresponding to the emotions in E. The emotion lexicon UMM_EmoLex is learnt by using the k emotion language models and the background model N as follows:

$$UMM\_EmoLex(w_i, \theta_{e_j}) = \frac{P(w_i|\theta_{e_j}^{(n)})}{\sum_{t=1}^{k}[P(w_i|\theta_{e_t}^{(n)})] + P(w_i|N)} \qquad (6)$$

$$UMM\_EmoLex(w_i, N) = \frac{P(w_i|N)}{\sum_{t=1}^{k}[P(w_i|\theta_{e_t}^{(n)})] + P(w_i|N)} \qquad (7)$$

where k is the number of emotions in the corpus, and UMM_EmoLex is a $|V| \times (k + 1)$ matrix.

5 Baseline Domain-specific Lexicon Generation Methods

In this section we formally present the other lexicon generation methods proposed in the literature using latent dirichlet allocation and word-document frequency statistics. These lexicon generation methods can be used to induce emotion and sentiment lexicons from documents labelled for emotions and sentiments respectively.

5.1 Supervised Latent Dirichlet Allocation based Emotion Lexicon

Latent Dirichlet Allocation (LDA) [5] is a popular topic detection algorithm which models documents to exhibit characteristics of multiple topics. In sentiment analysis LDA is applied to capture the relationships between words and sentiment (positivity, negativity) in addition to the topics [22, 19]. Similarly in emotion detection, LDA has been applied in a semi-supervised manner using a minimal set of domain-independent seed emotion words to learn emotion-relevant topics [39]. However supervised LDA (sLDA) [21] offers a more accurate means to learn emotion-relevant topics from labelled/weakly-labelled emotion corpora, because the usage of a minimal set of seed emotion words, does not guarantee the same level of coverage for all domains, thereby affecting the accuracy of the topics generated. Accordingly sLDA can be used to learn topic (emotion) distributions and map these into a word-emotion lexicon. More formally, let $\theta_{e_1}, \theta_{e_2}, \ldots, \theta_{e_n}$ be the topic distributions learnt for emotions $e_1, e_2, \ldots, e_n$, then the emotion lexicon is induced as follows:

$$sLDA\_EmoLex(w_j, e_n) = \frac{P(w_j|\theta_{e_n})}{\sum_{i=1}^{|E|} P(w_j|e_i)} \qquad (8)$$

where θ_{e_n} is the topic distribution for emotion e_n obtained from sLDA, where w_j is the jth word in the vocabulary V.

We learn the emotion-aware sentiment lexicons (*sLDA_EmoSentiLex* and *sLDA_SentiLex*) using *sLDA_EmoLex* following the same process as illustrated in sections 3.1 and 3.2.

5.2 Word-Document Frequency based Emotion Lexicon

Crowd-sourced emotion annotations provided by readers of the documents (e.g news stories) are used to learn word-emotion lexicon. These emotion annotations are in the form of numerical ratings, which can be normalized to define a probability distribution of emotions on each document. [33] proposed a lexicon generation method by combining the document-frequency distributions of words and the emotion distributions over documents. Since this method involves modelling of frequencies of words in emotional documents and emotion ratings, we refer to the method as word-document-frequency (WDF) lexicon. The generation method for the lexicon can be formally described as follows:

$$WDF - EmoLex(w_j, e_n) = \frac{\sum_{i=1}^{|X|} P(w_j|d_i)r_{in}}{\sum_{n=1}^{|E|} \sum_{i=1}^{|X|} P(w_j|d_i)r_{in}} \tag{9}$$

where w_j is the jth word in vocabulary V and r_{in} is the normalized emotion rating of the nth emotion in E on the ith document in the corpus X. Observe that unlike the sLDA and UMM lexicons, the WDF lexicon requires the emotion labels for the documents in the form of numerical ratings. As in the case of sLDA based emotion lexicon we also learn emotion-aware sentiment lexicons (*WDF_EmoSentiLex* and *WDF_SentiLex*) using *WDF_EmoLex* following the process illustrated in sections 3.1 and 3.2.

6 Evaluation

Our evaluation is a comparative study, of the performance of the standard sentiment lexicons, and the proposed emotion corpus based sentiment lexicons through a variety of sentiment analysis tasks on benchmark Twitter data sets. Significance is reported using a paired one-tailed t-test using 95% confidence (i.e. with p-value $\leq$ 0.05). Observe that in all our experimental results, the best performing methods are highlighted in bold.

6.1 *Evaluation Tasks*

Our evaluation includes the following sentiment analysis tasks.

1. *Sentiment intensity prediction*: Given a collection of words/phrases extracted from sentiment bearing tweets, the objective is to predict a sentiment intensity score for each word/phrase and arrange them in decreasing order of intensity.

The predictions are validated against a ranking given by humans. Formally, given a phrase P, the sentiment intensity score for the phrase is calculated as follows:

$$SentimentIntensity(P) = \sum_{w \in P} Log\left(\frac{Lex(w,+)}{Lex(w,-)}\right) \times count(w,P) \qquad (10)$$

where w is a word in the phrase P, $count(w,P)$ is the number of times w appears in P. $Lex(w,+), Lex(w,-)$ are the positive and negative valences for the word w in a lexicon. Some lexicons offer the sentiment intensity scores (e.g. SenticNet, S140 lexicon), in which case we use them directly. The aforementioned computation applies to the UMM based lexicons like *Sentilex* and S140-UMM lexicon.

2. *Sentiment classification*: Given a collection of documents (tweets), the objective is to classify them into positive and negative classes. The predictions are validated against human judgements. Formally, given a document d, the sentiment class is predicted using a lexicon as follows:

$$d[+] = \sum_{w \in d} Lex(w,+) \times count(w,d) \qquad (11)$$

where $d[+]$ is the positive intensity of d. Similarly $d[-]$ indicates the negative intensity of d. Finally the sentiment class of d is determined as follows:

$$Sentiment(d) = \begin{cases} positive \text{ if } d[+] > d[-] \\ negative \text{ if } d[-] > d[+] \end{cases} \qquad (12)$$

6.2 *Datasets*

We use four benchmark data sets in our evaluation. Note that the emotion corpus is used in two different ways to learn sentiment lexicons (refer sections 3 and 4). Further the S140 training data is used to learn a sentiment lexicon using the proposed method (refer section 4). The remaining data sets are used for evaluation. We expect our evaluation to test the transferability of each of the lexicons, given that the training and test data are not always from the same corpus, albeit from similar genre.

6.2.1 Emotion Dataset

A collection of 0.28 million emotional tweets crawled from Twitter streaming API[1] using emotion hashtags provided in [37]. The emotion labels in the data set correspond to Parrot's [29] primary emotions and were obtained through distant-supervision[2]. Parrot's emotion theory identifies an equal number of positive and negative emotions. Therefore we expect the sentiment lexicons learnt on this corpus to be able to mine both positive and negative sentiment in the test corpora.

---

[1] https://dev.twitter.com/streaming/public

[2] http://www.gabormelli.com/RKB/Distant-Supervision-Learning-Algorithm

6.2.2 S140 Dataset

A collection of 1.6 million (0.8 million positive and 0.8 million negative) sentiment bearing tweets harnessed by Go et.al [13] using the Twitter API. Further the data set also contains a collection of 359 (182 positive and 177 negative) manually annotated tweets. We generate a sentiment lexicon using the proposed method in section 4 on the 1.6 million tweets and compare it with the S140 lexicon [24].

6.2.3 SemEval-2013 Dataset

A collection of 3430 (2587 positive and 843 negative) tweets hand-labelled for sentiment using Amazon Mechanical Turk [27]. Note that unlike the S140 test data, there is high skewness in the class distributions. Therefore it would be a greater challenge to transfer the lexicons learnt on the emotion corpus and also those learnt on the S140 training corpus to sentiment classification.

6.2.4 SemEval-2015 Dataset

A collection of 1315 words/phrases hand-labelled for sentiment intensity scores [34]. A higher score indicates greater positivity. Further the words/phrases are arranged in decreasing order of positivity. We used this data set to validate the performance of different lexicons in ranking words/phrases for sentiment.

6.3 *Baselines and Metrics*

The following different models are used in our comparative study:

1. Resource-based sentiment lexicons SentiWordNet and SenticNet;
2. Corpus-based sentiment lexicons S140 lexicon [24] and NRCHashtag lexicon [24];
3. Corpus-based sentiment lexicon (S140-UMM lexicon) learnt using the proposed method on S140 corpus (refer section 4);
4. Corpus-based sentiment lexicons ($WDF\_EmoSentiLex$, $WDF\_SentiLex$, $sLDA\_EmoSentiLex$ and $sLDA\_SentiLex$) learnt on the emotion corpus (refer section 6.2.1) using WDF, sLDA methods (refer sections 5);and
5. Corpus-based sentiment lexicons ($UMM\_EmoSentiLex$ and $UMM\_SentiLex$) learnt on the emotion corpus (refer section 6.2.1) using the proposed method (refer sections 3 and 4)

Performance evaluation is done using using Spearman's rank correlation coefficient and F-score for sentiment ranking and sentiment classification respectively. F-score is chosen for the classification task since it measures the performance of an algorithm in terms of both precision and recall.

6.4 *Results and Analysis*

In this section we first analyse the quality of the different lexicon generation methods visually using word clouds, thereafter we analyse the sentiment ranking results and the sentiment classification results obtained using the different lexicons.

6.5 Emotion word clouds for Lexicons

In this section we analyse the word-emotion associations learnt by the different lexicons, WDF, sLDA and UMM lexicon from the twitter emotion corpus. This analysis is particularly interesting, as we expect it to reveal interesting trends that could effect the performance on the sentiment analysis tasks, whose results are discussed later in this section. Figures 3 and 4show the most expressive words for emotions *fear and joy* (other emotions not shown due to space constraints) identified by WDF, sLDA and UMM lexicons. It is evident from the figures that all these lexicons capture the domain-specific vocabulary that is expressed informally. This is very important in order to be able to effective emotion detection in dynamic domains such as Twitter.

The word clouds presented in the figures are the top 100 words for each emotion, after removing the common words in English language. We observed that WDF lexicon is biased towards the majority class (*joy* here) in the corpus in learning the word-emotion associations. This done by observing the class level performance metrics like accuracy and F-score. For example it identified words connoting *joy* such as *succeed!* and *Ha!* as top *anger* words, similarly for other emotions. This is due to the fact that WDF lexicon is designed for emotion rated documents and it is less effective in capturing word-emotion associations on a corpus that have discrete emotion labels. On the other hand sLDA lexicon, because of the assumption of its underlying generative model that documents are a mixture of multiple topics (emotions) learnt better word-emotion associations compared to WDF lexicon. However sLDA lexicon was not able to discriminate effectively between words that strongly convey a particular emotion and those that are weakly associated with an emotion. For example words such as *scared, worried and nervous* are not well distinguished from other words for emotion *fear* and similarly for other emotions. As a result it was observed in the word clouds for the sLDA lexicon that top words for each emotion have similar size. This is not desirable since the word-emotion association scores form an important knowledge resource for sentiment analysis.

It was observed that the proposed UMM lexicon discriminate between strong and weak words for each emotion effectively. This is very promising, since this knowledge will be very useful for the sentiment intensity prediction task. Further UMM is also observed to capture words that are emotion-relevant but are rare. For example words such as *:)* and *fun!* for the emotion *joy*. We expect this word-level analysis to help infer useful insights about the performance gains of the proposed lexicon over the baselines in different sentiment analysis tasks. In the following section we analyse the performance of the different lexicons in sentiment analysis tasks.

Fig. 3: Top fear words for WDF (top), sLDA (middle) and UMM (bottom) lexicons

6.5.1 Sentiment Ranking

Table 1 summarizes the sentiment ranking results obtained for different lexicons. In general resource-based lexicons SentiWordNet and SenticNet are outperformed by all the corpus-based lexicons. This is expected, because the vocabulary coverage of these lexicons relevant to social media is limited compared to other lexicons. Furthermore, the results also suggest that the sentiment intensity knowledge captured by the corpus-based lexicons is superior to that of resource-based lexicons.

NRCHashtag lexicon performed significantly better than the remaining baselines and the proposed *UMM_EmoSentiLex*. The significant performance differences between NRCHashtag lexicon and S140 lexicon and NRCHashtag lexicon and S140-UMM lexicon clearly suggests the superiority of the NRCHashtag corpus over the S140 corpus in learning transferable lexicons for sentiment intensity prediction. It would be interesting to compare the performance of these lexicons in the sentiment classification tasks. In the case of emotion-aware sentiment lexicons learnt using WDF, sLDA we observed that *WDF_SentiLex* and *sLDA_SentiLex* outperformed their counterparts *WDF_EmoSentiLex* and *sLDA_EmoSentiLex*. This is expected since the former ones models document-sentiment relationships. Further

Fig. 4: Top joy words for WDF (top), sLDA (middle) and UMM (bottom) lexicons

sLDA based lexicons performed better over the WDF ones given that it was better able to model word-emotion associations from the documents as illustrated in section 6.5.

It is extremely promising to see that the proposed lexicons outperform most of the baselines significantly. Amongst the proposed lexicons, *UMM_SentiLex* performed significantly better than *UMM_EmoSentiLex*. This is not surprising, since *UMM_SentiLex* has the ability to incorporate the sentiment-class knowledge of the documents in the learning stage. This exactly follows the findings of earlier research in supervised and unsupervised sentiment analysis.

6.5.2 Sentiment Classification

Sentiment classification results for the S140 data set are shown in Table 2. Here unlike in the sentiment intensity prediction task, SentiWordNet demonstrated comparable performance with that of corpus-based lexicons. However, SenticNet does perform the worst amongst all the lexicons. This suggests that SentiWordNet is better transferable onto social media compared to SenticNet.

Table 1: Sentiment Ranking Results

Method	Spearman's Rank Correlation Coefficient
Emotion-agnostic Sentiment Lexicons (Baselines)	
SentiWordNet	0.479
SenticNet	0.425
S140 lexicon	0.506
NRCHashtag lexicon	0.624
S140-UMM-lexicon	0.517
Emotion-aware Sentiment Lexicons (Baselines)	
WDF-EmoSentiLex	0.489
WDF-Sentilex	0.497
sLDA-EmoSentiLex	0.493
sLDA-SentiLex	0.514
Emotion-aware Sentiment Lexicons (Proposed Methods)	
UMM-EmoSentiLex	0.572
UMM-SentiLex	**0.682**

The S140 corpus based lexicons significantly outperform NRCHashtag lexicon, given their advantage to train on a corpus, that is similar to the test set. In the case of lexicons based on sLDA and WDF we observed similar trends as seen in the sentiment intensity prediction task. Overall *sLDA_SentiLex* performed better than the other sLDA and WDF lexicons given that it has the ability to model word-emotion associations and document-sentiment relationships more effectively.

However, the proposed lexicon $UMM\_SentiLex$ recorded the best performance on this data set. once again the superiority of $UMM\_SentiLex$ over $UMM\_EmoSentiLex$ is evidenced, given its ability to incorporate sentiment-class knowledge of the documents in the learning stage. The performance improvements of emotion corpus based sentiment lexicons over a majority of baseline lexicons, clearly suggests that emotion knowledge when exploited effectively is very useful for sentiment analysis.

Table 3 summarizes the results for different lexicon on the SemEval-2013 data set. Unlike the previous, this data set has a very skewed class distribution. The impact of this is clearly reflected in the results. Majority of the lexicons recorded strong performances in classifying positive class documents. Once again SentiWordNet demonstrated that it is better transferable onto social media compared to SenticNet.

Similar to the previous data set, S140 corpus based lexicons performed better than NRCHashtag corpus based lexicon. Overall comparison across the evaluation tasks suggests that S140 corpus based lexicons record better performance in sentiment classification, whereas NRCHashtag lexicon records better performance in sentiment quantification. This offers interesting directions for future work on composing different corpora for learning sentiment lexicons.

In the case of sLDA and WDF based lexicons we observed that they perform poorly compared to baseline lexicons that are emotion-agnostic. This could be due to the sensitive nature of these methods to class distributions while learning lexi-

Table 2: Sentiment Classification Results on S140 test data set

Method	Positive F-score	Negative F-score	Overall F-score
Emotion-agnostic Sentiment Lexicons (Baselines)			
SentiWordNet	69.42	67.60	68.51
SenticNet	59.88	59.84	59.86
S140-lexicon	71.55	69.42	70.48
NRCHashtag-lexicon	66.66	64.75	65.70
S140-UMM-lexicon	**75.14**	69.36	72.25
Emotion-aware Sentiment Lexicons (Baselines)			
WDF-EmoSentiLex	58.57	48.24	53.40
WDF-SentiLex	59.63	50.12	54.87
sLDA-EmoSentiLex	61.68	57.23	59.45
sLDA-SentiLex	62.93	58.11	60.52
Emotion-aware Sentiment Lexicons (Proposed Methods)			
UMM-EmoSentiLex	67.51	71.14	69.32
UMM-SentiLex	72.93	**74.11**	**73.52**

cons. In general we observed that these two lexicon generation methods were not able to leverage the emotion corpus effectively to learn sentiment polarity lexicons suggesting the usefulness of the proposed UMM method which has consistently recorded best performances in all the sentiment analysis tasks. We highlight our findings for the UMM lexicon on the SemEval-2013 data set below.

The proposed lexicon *UMM_EmoSentiLex* performed significantly below most of the lexicons on this data set. We believe the inability to learn the document-sentiment relationships, coupled with the skewed class distribution characteristics of the data set resulted in such performance degradation. However, our proposed lexicon *UMM_SentiLex* significantly outperformed all the remaining lexicons. The consistent performance of *UMM_SentiLex* in all the evaluation tasks, strongly evidences the correlation between emotions and sentiments. We believe that the emotion-sentiment mapping in psychology effectively clusters the emotion corpus into sentiment classes, thereafter the ability of the UMM model to effectively capture the word-sentiment relationships resulted in the performance improvements for *UMM_SentiLex*.

7 Conclusions

In this paper we study the mapping proposed in psychology between emotions and sentiments, from a computational modelling perspective in order to establish the role of an emotion corpus for sentiment analysis. By combining a generative unigram mixture model (UMM) with the emotion-sentiment mapping, we propose two different methods to extract lexicons for Twitter sentiment analysis from an emotion labelled Twitter corpus. Further we also evaluate how the proposed UMM lexi-

Table 3: Sentiment Classification Results on SemEval-2013 data set

Method	Positive F-score	Negative F-score	Overall F-score
Emotion-agnostic Sentiment Lexicons (Baselines)			
SentiWordNet	80.14	50.38	65.26
SenticNet	54.95	55.94	55.45
S140-lexicon	80.13	57.87	69.00
NRCHashtag-lexicon	80.25	53.98	67.11
S140-UMM-lexicon	78.87	55.85	67.36
Emotion-aware Sentiment Lexicons (Baselines)			
WDF-EmoSentiLex	59.83	47.39	53.61
WDF-SentiLex	62.18	59.88	61.03
sLDA-EmoSentiLex	67.84	58.99	63.41
sLDA-SentiLex	73.45	60.02	66.73
Emotion-aware Sentiment Lexicons (Proposed Methods)			
UMM-EmoSentiLex	64.51	48.37	56.44
UMM-SentiLex	**83.06**	**60.98**	**72.02**

con generation method fares in comparison with other automatic lexicon extraction methods proposed using supervised latent dirichlet allocation (sLDA) and word-document frequency statistics (WDF) in learning emotion-aware sentiment polarity lexicons. We comparatively evaluate the quality of the proposed emotion-aware sentiment lexicons, those generated using sLDA, WDF and standard sentiment lexicons that are agnostic to emotion knowledge through a variety of sentiment analysis tasks on benchmark Twitter data sets. Our experiments confirm that the proposed sentiment lexicons, yield significant improvements over standard lexicons in sentiment classification and sentiment intensity prediction tasks. It is extremely promising to see the potential of an emotion corpus as a useful knowledge resource for sentiment analysis, especially on social media where emotions and sentiments are widely expressed. Further the cost-effectiveness of the emotion-sentiment mapping to cluster the emotion corpus into positive, negative classes (0.28 million tweets in a second) makes it practically possible to adopt large emotion corpora, in order to extract sentiment lexicons with improved coverage.

References

1. Bandhakavi, A., Wiratunga, N., Deepak, P., Massie, S.: Generating a word-emotion lexicon from #emotional tweets. In: Proc of the 3rd Joint Conference on Lexical and Computational Semantics (*SEM 2014) (2014)
2. Bandhakavi, A., Wiratunga, N., Massie, S., Deepak, P.: Lexicon generation for emotion detection from text. IEEE Intelligent Systems, January/February (2017)
3. Binali, H., Potdar, V.: Emotion detection state-of -the-art. In: Proc of the CUBE International Information Technology Conference, pp. 501–507 (2012)
4. Binali H. Potdar, V., Wu, C.: Computational approaches for emotion detection in text. In: 4th IEEE International Conference on Digital Ecosystems and Technologies DEST (2010)

5. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. the Journal of machine Learning research **3**, 993–1022 (2003)
6. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In: Proc of the 43rd Hawaii International Conference on System Sciences. (2010)
7. Cambria, E., Olsher, D., Rajagopal, D.: Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. In: 28th AAAI conf on Artificial Intelligence, pp. 1515-1521 (2014)
8. Ekman, P.: An argument for basic emotions. Cognition and Emotion, 6(3), pp. 169-200 (1992)
9. Esuli, A., Baccianella, S., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proc of LREC (2010)
10. Fellbaum, Christiane: Wordnet and wordnets. In: Encyclopedia of Language and Linguistics pp. 665–670 (2005)
11. Feng, S., K.Song, D.Wang, G.Yu: A word-emotion mutual reinformcement ranking model for building sentiment lexicon from massive collection of microblogs. World Wide Web, 18(4):949-967 (2015)
12. Ghazi, D., Inkpen, D., Szpakowicz, S.: Hierarchical approach to emotion recognition and classification in texts. In: Proc of the 23rd Canadian conference on Advances in Artificial Intelligence (2010)
13. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Processing, pp 1-6 (2009)
14. Hogenboom, A., Bal, D., Frasincar, F., Bal, M.: Exploiting emoticons in polarity classification of text. Journal of Web Engineering (2013)
15. Hu, M., Liu., B.: Mining and summarizing customer reviews. In: Proc of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2004)
16. Hu, X., Tang, J., Gao, H., Liu, H.: Unsupervised sentiment analysis with emotional signals. In: Proc of the International World Wide Web Conference (WWW) (2013)
17. Jiang, F., Liu, Y.Q., Luan, H.B., Sun, J.S., Zhu, X., Zhang, M., Ma, S.P.: Microblog sentiment analysis with emoticon space model. Journal of Computer Science and Technology, vol 30(5), pp 1120-1129 (2015)
18. Jin, X., Wang, Z.: An emotion space model for recognition of emotions in spoken chinese. In: Proc of the First international conference on Affective Computing and Intelligent Interaction (2005)
19. Lin, C., He, Y.: Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on Information and knowledge management, pp. 375–384. ACM (2009)
20. Liu, H., Singh., P.: Conceptnet- a practical commonsense reasoning tool-kit. BT Technology Journal, 22(4), pp. 211-226 (2004)
21. Mcauliffe, J.D., Blei, D.M.: Supervised topic models. In: Advances in Neural Information Processing Systems, pp. 121-128 (2007)
22. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th international conference on World Wide Web, pp. 171–180. ACM (2007)
23. Mohammad, S.M.: #emotional tweets. In: Proc of The First Joint Conference on Lexical and Computational Semantics, pp. 246-255 (2012)
24. Mohammad, S.M., Kiritchenko, S., Zhu, X.: Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. In: 7th International Workshop on Semantic Evaluation (SemEval 2013), pp 321-327 (2013)
25. Mohammad, S.M., Turney, P.: Crowdsourcing a word-emotion association lexicon. Computational Intelligence, 29(3), pp. 436-465 (2013)
26. Munezero, M., Montero, C.S., Sutinen, E., Pajunen, J.: Are they different? affect, feeling, emotion,sentiment, and opinion detection in text. IEEE Transactions on Affective Computing, Vol 5 No 2 (2014)
27. Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., Wilson, T.: Semeval-2013 task2: Sentiment analysis in twitter. In: Proc of the 7th International Workshop on Semantic Evaluation (SemEval-2013) (2013)

28. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval **2(1)**, 1–135 (2008)
29. Parrott, W.: Emotions in social psychology. Psychology Press, Philadelphia (2001)
30. Plutchik., R.: A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), Emotion: Theory, research, and experience: **Vol. 1.**, (pp. 3–33) (1980)
31. Poria, S., Gelbukh, A., Cambria, E., Hussain, A., Huang, G.B.: Emosenticspace: A novel framework for affective common-sense reasoning. Knowledge-Based Systems 69, pp. 108-123 (2014)
32. Qadir, A., Riloff, E.: Bootstrapped learning of emotion hashtags #hashtags4you. In: the 4th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2013) (2013)
33. Rao, Y., Lei, J., Wenyin, L., Li, Q., Chen, M.: Building emotional dictionary for sentiment analysis of online news. World Wide Web, Vol 17, pp. 723-742 (2014)
34. Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S.M., Ritter, A., Stoyanov, V.: Semeval-2015: Sentiment analysis in twitter. In: Proc of the 9th International Workshop on Semantic Evaluation (SemEval-2015) (2015)
35. Song, K., Feng, S., Gao, W., Wang, D., Chen, L., Zhang, C.: Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a hetereogeneous graph. In: Proc of the 26th ACM Conference on Hypertext & Social Media, pp. 283-292 (2015)
36. Stone, P.J., Dexter, D.C., Marshall, S.S., Daniel, O.M.: The general inquirer: A computer approach to content analysis. The MIT Press (1966)
37. Wang, W.: Harnessing twitter "big data" for automatic emotion identification. In: Proc of the ASE/IEEE International Conference on Social Computing and International Conference on Privacy, Security, Risk and Trust (2012)
38. Wilson, T., Wiebe, J., Hoffmann., P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Proc. of HLT-EMNLP-2005 (2005)
39. Yang, M., Peng, B., Chen, Z., Dingju Zhu, a.K.C.: A topic model for building fine-grained domain-specific emotion lexicon. In: Proc of the 52nd Annual Meeting of the Assoc for Computational Linguistics, pp. 421-426 (2014)