Lexicon Generation for Emotion Detection from Text

Article *in* Intelligent Systems, IEEE · January 2017

DOI: 10.1109/MIS.2017.22

4 authors:

Anil Sriharsha Bandhakavi
Robert Gordon University

Nirmalie Wiratunga
Robert Gordon University

Stewart Massie
Robert Gordon University

Deepak P
Queen's University Belfast

Lexicon Generation for Emotion Detection from Text

Anil Bandhakavi*, Nirmalie Wiratunga*, Stewart Massie*, Deepak P.†
*Robert Gordon University, Aberdeen, UK †Queen's University, Belfast, UK
{a.s.bandhakavi, n.wiratunga, s.massie}@rgu.ac.uk
{deepaksp}@acm.org

***Abstract*—General Purpose Emotion Lexicons (GPELs) that associate words with emotion categories remain a valuable resource for emotion detection. However the static and formal nature of their vocabularies, make them an inadequate resource for detecting emotions in domains that are inherently dynamic in nature. This calls for lexicons that are not only adaptive to the lexical variations in a domain but also provide finer-grained quantitative estimates to accurately capture word-emotion associations. In this paper we demonstrate how labelled (blogs, news headlines) and weakly-labelled (tweets) emotion text can be harnessed to learn a word-emotion association lexicon by jointly modelling emotionality and neutrality of words using a generative unigram mixture model (UMM). Empirical evaluation confirms that UMM generated emotion language models (topics) have significantly lower perplexity compared to those from state-of-the-art generative models like supervised Latent Dirichlet Allocation (sLDA). Further emotion detection tasks involving word-emotion classification and document-emotion ranking confirm that the UMM lexicon significantly out performs GPELs and also state-of-the-art domain specific lexicons.**

***Index Terms*—Emotion Detection, Domain specific Lexicon, Mixture Model, Word-classification, Emotion-ranking.**

I. Introduction

Textual emotion detection is the computational study of natural language expressed in text, in order to identify its association with emotions such as *anger, fear, joy, sadness* etc. It has potential in many different applications for industry, media and government organisations. However, its uptake has arguably been slow, mainly due to the challenges involved in modelling fine-grained subjectivity and the subtlety of emotive expressions in text.

Until recently popular resources like sentiment lexicons [1] and general purpose emotion lexicons (GPELs) (e.g. WordNet-Affect [2]) have been used for emotion detection from text. However sentiment lexicons, due to lack of granular emotion information, and GPELs, due to the static and formal nature are inadequate for emotion detection in domains such as social media, which are inherently dynamic. For instance, on Twitter, informal vocabulary (e.g. #romeisawesome, #loveisbliss!!! etc) and emoticons (e.g. :-), :-(etc) are used to convey emotions, instead of formal vocabulary as in GPELs. Further, the association between words and emotions vary from one domain to another and calls for contextual disambiguation. For example *Glee* may normally indicate *joy*, but, would need to be interpreted as neutral in a corpus of documents talking about the television series with the same name[1]. Further, *unfair* may be associated with *anger* despite being more dominant in documents expressing *sadness*; the crisp binary memberships of words in GPELs cannot capture such fuzzy associations between words and emotion classes. Therefore, it is necessary to build domain emotion specific lexicons (DSELs) which offer quantitative fine-grained estimates for word-emotion associations within a domain. Accordingly, recent efforts in emotion detection focused on learning emotion lexicons from labelled emotion corpora as well as weakly-labelled social media content [3–6].

Social media (e.g. tweets) offers access to weakly-labelled emotional data by users with emoticons and emotion hashtags, which can be leveraged to learn DSELs for a variety of tasks concerning emotion detection. In particular, DSELs offer useful knowledge to design a range of document representations from simple binary, to frequency counts, to more sophisticated emotion concepts. Further, DSELs can be deployed to search and index vast amounts of emotional content (e.g. song lyrics, video descriptions etc) on the social web, in order to infer emotions of social groups/communities. Our contributions in this paper are as follows:

- We propose a generative unigram mixture model in order to learn a word-emotion association lexicon from an input document corpus.
- We empirically evaluate the quality of the emotion language models (topics) generated by: the proposed method; and by sLDA using standard metrics, such as Perplexity.
- We evaluate the quality of the emotion lexicons generated by: the proposed method; and by state-of-the-art baseline methods on two emotion detection tasks: *word-emotion classification; and document-emotion ranking*

In the rest of the paper we review related literature in Section 2. In section 3 we outline the problem. Section 4 formulates the mixture model, followed by parameter estimation and lexicon generation is sections 5 and 6. In section 7 we outline the evaluation tasks. Section 8 presents results from the empirical evaluation and the comparative study, followed by conclusions and future directions in Section 9.

[1]http://en.wikipedia.org/wiki/Glee_(TV_series)

II. RELATED WORK

Emotion lexicons unlike sentiment lexicons [1, 7] offer granular emotion information. WordNet synsets were manually labelled with Ekman [8] basic emotions to generate WordNet-Affect [2]. The NRC word-emotion lexicon [9] was obtained by crowd sourcing emotion annotations to 14182 words obtained from Google n-gram corpus[2]. Unlike the earlier lexicons, recently semantically-rich lexicons such as SenticNet [10, 11] are proposed to model sentiment of multi-word expressions using common-sense knowledge derived from ConceptNet [12]. Further fuzzy clustering and machine learning techniques are applied to assign WordNet-Affect emotion labels to concepts in SenticNet to obtain EmoSenticNet [13]. A common limitation for the aforementioned emotion lexicons is that their vocabulary is static and formal, which makes it challenging to deploy them in dynamic and informal domains (e.g. social media) for emotion detection. Addressing the above limitation, methods for building lexicons which capture the domain level associations between words and emotions have been proposed [14–16].

Existing methods for building domain specific lexicons are mostly supervised, since they rely either on labelled or weakly-labelled emotive content in a domain. For instance, Pointwise Mutual Information (PMI) was applied to learn a word-emotion lexicon, from tweets weakly-labelled with emotion hashtags [3]. Staiano and Guerini [14] proposed to leverage crowd-annotated emotional news articles[3] for lexicon generation, by combining the document-frequency distributions of words and the emotion distributions over documents.

Further in literature, generative models like Latent Dirichlet Allocation (LDA) are also applied to lexicon generation. Yanghui Rao et al. [4], combined user emotion ratings on documents[4], document-frequency distributions and document-topic distributions from LDA to learn word-emotion, topic-emotion lexicons. Min Yang et al. [6] proposed a semi-supervised LDA approach, which uses a minimal set of domain-independent emotion seed words to guide the LDA process to learn emotion-relevant topics. However the topics learnt from this approach are not consistently accurate, since the coverage of seed words varies from one domain to another. Nevertheless, supervised LDA (sLDA) [17] offers a more accurate means to learn emotion-topic models for lexicon generation, from labelled or weakly-labelled emotion corpora.

In this work we propose a mixture model to learn domain specific word-emotion lexicon. Our model assumes documents to be a mixture of emotional and neutral words, which is different from the generative model of sLDA that assumes documents to be a mixture of multiple emotion (topic) words. We expect the joint modelling of emotionality and neutrality at word-level to be more effective on real-world emotion corpora, since not every word in them connotes emotions.

[2]https://catalog.ldc.upenn.edu/LDC2006T13
[3]http://www.rappler.com/
[4]http://news.sina.com.cn/society/

Notation	Description
X	Corpus of emotion labelled documents
E	Set of emotion labels
D_{e_t}	Documents labelled with emotion e_t
N	Neutral (background) language model
θ_{e_t}	Language model for emotion e_t
V	Set of unique words from documents in X
w_i	i^{th} word in the vocabulary V
Z_{w_i}	Hidden (unobserved) variable corresponding to w_i
λ_{e_t}	Mixture parameter (empirically estimated)
n	EM iteration number
$Q(\theta_{e_t}^{(n+1)}; \theta_{e_t}^{(n)})$	Q-function
$c(w, d_i)$	#times word w occurs in document d_i
$Lex(i, j)$	Emotional valence between word w_i and emotion e_j
$Lex(i, k+1)$	Neutral valence for the word w_i

TABLE I: Notations

III. PROBLEM DEFINITION

The problem essentially is to learn a word-emotion lexicon from an input corpus of emotion labelled documents and is formulated as follows. Given a corpus of documents X, with emotion labels from $E = \{e_1, \ldots, e_k\}$, we learn a word-emotion lexicon Lex, where $Lex(i, j)$ is the emotional valence of the i^{th} word in vocabulary V to the j^{th} emotion in E and $Lex(i, k+1)$ corresponds to its neutral valence. The word-emotion lexicon is learnt using a set of k unigram mixture models (UMMs), where the t^{th} UMM assumes that documents in X labelled with emotion e_t are a mixture of words bearing emotion e_t and some background (neutral) words. Therefore each UMM is a linear combination of two unigram language models, θ and N along with a mixing parameter λ. The conceptual diagram of the proposed UMM is shown in Figure 1. Initial models $\theta_{e_t}^{(0)}$ and N are learnt from the training data. Mixture parameter λ_{e_t} is set empirically. The estimation of the hidden variable, Z_w happens in the expectation step (E-step). In the maximization step (M-step) parameter, θ_{e_t} is updated. This process repeats until the values of θ_{e_t} do not change significantly. The important mathematical notations are summarized in Table I.

IV. GENERATIVE MODEL FOR DOCUMENTS

We now outline our generative model for emotion bearing documents. We use an example from real world data to motivate this discussion.

A. Unigram Mixture Model

Real-world emotion data typically is a mixture of emotion rich words as well as background (emotion-neutral) words. For example consider the tweet *Sunday in Lasvegas #excited #joyous* which explicitly connotes emotion *joy*. However the word *Sunday* is evidently not indicative of *joy*. Further *Lasvegas* could connote emotions such as *Love*. Therefore it is important to have a model which accounts for such word mixtures in the documents. The mixture model in our case is as follows. Let D_{e_t} be the documents labelled with emotion e_t, then according to the unigram mixture model, documents in D_{e_t} are generated

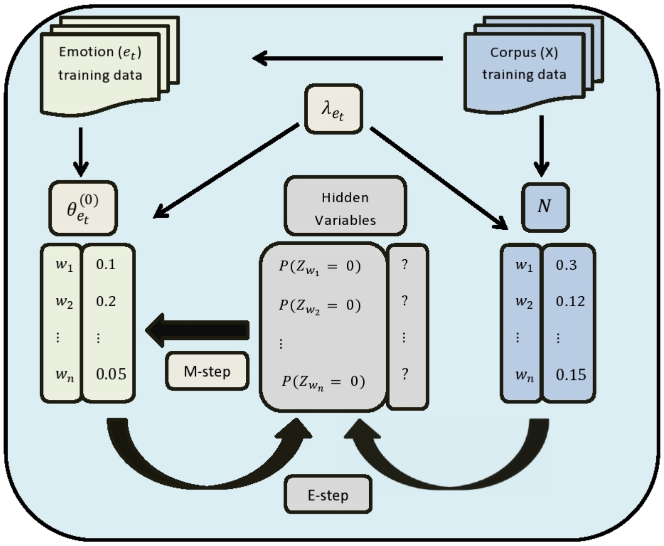


Fig. 1: Visualization of the UMM generation and the Expectation Maximization (EM) iterative process for emotion e_t.

independently from a linear mixture of an emotion language model θ_{e_t} and a background language model N as follows:

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

Note that the above mixture model reduces to a single language model when λ_{e_t} is 1. Thus λ_{e_t} in our case indicates the noisy (neutral and other emotion) words which occur in documents connoting emotion e_t. Finally Z_w is the hidden (latent) binary variable corresponding to word w, which indicates the mixture component (language model) which generated the word w. For each word $w \in V$ its corresponding hidden variable is defined as follows:

$$Z_w = \begin{cases} 1 & \text{if word } w \text{ is from the neutral model} \\ 0 & \text{otherwise} \end{cases}$$

In the following sections we illustrate the estimation of parameters (θ_{e_t}, λ_{e_t} and Z) of the mixture model, followed by lexicon generation.

V. Parameter Estimation of the Mixture Model

The objective is to find the parameters (θ_{e_t}, λ_{e_t} and Z) that maximize the probability of generating documents D_{e_t}. λ_{e_t} can be estimated using maximum likelihood estimation (MLE) as follows:

$$\hat{\lambda}_{e_t} = \underset{\lambda_{e_t}}{\operatorname{argmax}} \sum_{i=1}^{|D_{e_t}|} \sum_{w \in d_i} c(w, d_i) log[\lambda_{e_t} P(w|\theta_{e_t}) + (1 - \lambda_{e_t})P(w|N)] \quad (2)$$

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. In the E-step, the value of the hidden variable (Z_w) is estimated. Observe that

$$P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) + P(Z_w = 1|D_{e_t}, \theta_{e_t}^{(n)}) = 1 \quad (3)$$

Further from Bayes' theorem it follows that:

$$P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) = C \times \lambda_{e_t} \times P(w|\theta_{e_t}^{(n)}) \quad (4)$$

combining 3 and 4 gives:

E-step:

$$P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) = \frac{\lambda_{e_t} P(w|\theta_{e_t}^{(n)})}{\lambda_{e_t} P(w|\theta_{e_t}^{(n)}) + (1 - \lambda_{e_t})P(w|N)} \quad (5)$$

The M-step involves maximizing the following function:

$$Q(\theta_{e_t}^{(n+1)}; \theta_{e_t}^{(n)}) = \sum_{i=1}^{|D_{e_t}|} \sum_{w \in d_i} c(w, d_i)[P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) log(\lambda_{e_t} P(w|\theta_{e_t}^{(n+1)})) + P(Z_w = 1|D_{e_t}, \theta_{e_t}^{(n)}) log((1 - \lambda_{e_t})P(w|N))] \quad (6)$$

We thus consider the auxiliary function

$$g(\theta_{e_t}^{(n+1)}) = Q(\theta_{e_t}^{(n+1)}; \theta_{e_t}^{(n)}) + \mu(1 - \sum_{w \in V} P(w|\theta_{e_t}^{(n+1)})) \quad (7)$$

where μ is the Lagrange multiplier. Computing the first-order partial derivative of $g(\theta_{e_t}^{(n+1)})$ with respect to the parameter variable $P(w|\theta_{e_t}^{(n+1)})$ and equating to zero we get:

M-step:

$$P(w|\theta_{\theta_{e_t}}^{(n+1)}) = \frac{\sum_{i=1}^{|D_{e_t}|} P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) c(w, d_i)}{\sum_{w \in V} \sum_{i=1}^{|D_{e_t}|} P(Z_w = 0|D_{e_t}, \theta_{e_t}^{(n)}) c(w, d_i)} \quad (8)$$

A. EM Initialization

The initial language model $\theta_{e_t}^{(0)}$ for emotion e_t is defined as follows:

$$P(w_i|\theta_{e_t}^{(0)}) = \frac{f(w_i, D_{e_t})}{\sum_{w \in V} f(w, D_{e_t})} \quad (9)$$

where $f(w_i, D_{e_t})$ is the frequency of the i^{th} word in V in the training documents for emotion e_t. The background (neutral) language model is defined as follows:

$$P(w_i|N) = \frac{f(w_i, X)}{\sum_{w \in V} f(w, X)} \quad (10)$$

where $f(w_i, X)$ is the training corpus frequency for word w_i.

VI. Lexicon Generation

The word-emotion lexicon is learnt using the k emotion language models and the background model N as follows:

$$Lex^{(n)}(w_i, \theta_{e_j}) = \frac{P(w_i|\theta_{e_j}^{(n)})}{\sum_{t=1}^{k} [P(w_i|\theta_{e_t}^{(n)})] + P(w_i|N)} \quad (11)$$

$$Lex^{(n)}(w_i, N) = \frac{P(w_i|N)}{\sum_{t=1}^{k} [P(w_i|\theta_{e_t}^{(n)})] + P(w_i|N)} \quad (12)$$

where k is the number of emotions in the corpus, and $Lex^{(n)}$ is a $|V| \times (k+1)$ matrix generated after the n^{th} EM iteration.

VII. Lexicon Evaluation Tasks

In this section we formulate the different evaluation tasks for assessing the quality of the lexicons.

A. Word-Emotion Classification

In this task we evaluate the ability of a lexicon to classify a collection of target words hand labelled with emotions. More formally given an arbitrary word w the task is to predict an emotion label $e \in E$ for w using the word-emotion lexicon. Because a DSEL quantifies the associations between words in a vocabulary V and a range of emotions in E, for any given arbitrary word w the dominant emotion e being expressed is calculated using the lexicon as follows:

$$e = \arg\max_{j} Lex(w, j) \tag{13}$$

In contrast in a GPEL, $Lex(w, j)$, is modelled as a list of words per class:

$$Lex(w, j) = \begin{cases} 1 & \text{if } w \in List(E_j), \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $List(E_j)$ is the word-list for the j^{th} emotion.

B. Document-Emotion Ranking

The objective of this task is to assess the quality of the lexicon in predicting the association between a document and multiple emotions. More formally given a document d, expressing emotions $(e_1, \ldots, e_m)$ in decreasing order of magnitude, the task is to predict the order of emotions for d using a lexicon. For any given document d, an emotion ranking could be formed using an ordered list of emotions expressed by d, $(e_1, \ldots, e_m) \mid$ for $i, j \in (1, m)$, if $i < j$, then $d[e_i] > d[e_j]$, where $d[e]$ is calculated using the lexicon as follows:

$$d[e] = \sum_{w \in d} Lex(w, e) \times c(w, d) \tag{15}$$

VIII. Evaluation

In this section we begin with the details of the benchmark data sets used in the evaluation, followed by results and discussion for perplexity analysis and lexicon quality assessment. Significance is reported using a paired one-tailed t-test using 95% confidence.

A. Datasets

Our comparative study of lexicons is carried out on four benchmark data sets. In our evaluation, we use 90% of the training data in each data set for learning the lexicons and use the remaining 10% as development data for parameter tuning (e.g. MLE estimation of λ[5] for the UMMs). The test data in each data set is used for perplexity analysis and lexicon quality assessment

1) News Dataset (SemEval-2007): Consists of 1250 emotional news headlines[6], where each headline was provided with emotion ratings in the range [-100, 100] for the Ekman basic emotions. We used this data set for emotion ranking, as it provides an ordered list of emotions on each news item.

2) Twitter Dataset: A collection of 2.6 million emotional tweets[7] crawled from the Twitter search API using tweet identification numbers. We used the training data set for learning domain specific lexicons in our comparative study. We deployed the learnt lexicons in the emotion ranking task on a tweet event data set discussed later.

3) Blog Dataset: Consists of 5500 blog sentences[8] annotated with Ekman basic emotions. Also words which reflect the emotion of the sentence are provided as part of the data set. Hence we used this data set to evaluate the quality of lexicons in predicting word-level emotions. We performed 5 fold cross validation for our experiments (and not 10 fold due to the smaller size of the data set).

4) Emotion event Dataset: Collection of 200 tweets describing emotional events[9]. Each event is annotated with a ranked list of emotions by two annotators with agreement (kappa of 0.68). We used this data set to test the quality of the lexicons on the emotion ranking task. Since this data set is very small, lexicon learnt on the Twitter data was used here as both data sets are crawled from Twitter.

B. Baselines and Metrics

Our comparative study includes: baseline GPELs such as WordNet-Affect (WNA), NRC emotion lexicon and EmoSenticNet (ESN); baseline DSELs generated using PMI [3], Word-Emotion dictionary (WED) [4] and supervised LDA (sLDA) [17]; and a DSEL generated using the proposed method (see section III).

Performance assessment of DSELs is done on both the evaluation tasks (see section VII), whereas GPELs can only be used for comparison in word-emotion classification task, since they do not offer word-emotion quantifications needed for emotion ranking. In the word-emotion classification task performance is reported using the standard metric F-score. For document-emotion ranking Mean Reciprocal Rank (MRR) is used to measure the lexicon quality in predicting the dominant emotion present in the document, whereas the ability of a lexicon to order the multiple emotions connoted by a document is measured by Mean Average Precision (MAP).

C. Perplexity Analysis

Perplexity is a measure of how well an emotion language model θ_{e_k}, learnt using the training data $D^{train}_{e_k}$, predicts the test (unseen) data $D^{test}_{e_k}$ and is calculated as follows:

$$Perp(D^{test}_{e_k}) = 2^{-\frac{\sum_{i=1}^{|D^{test}_{e_k}|} \sum_{j=1}^{|d_i|} logP(d_{ij}|\theta_{e_k})}{|V_{e_k}|}} \tag{16}$$

[5] we experimented with 10 values (0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0) of λ on each data set for MLE

[6] http://nlp.cs.swarthmore.edu/semeval/tasks/task14/summary.shtml

[7] http://knoesis.org/?q=projects/emotion

[8] http://saimacs.github.io/

[9] http://ahclab.naist.jp/resource/eped/

Method	Avg Overall F-score
Baseline GPELs	
WNA	29.96%
NRC	39.05%
ESN	28.30%
Baseline DSELs	
PMI	42.12%
WED	24.51%
sLDA	38.72%
Proposed DSEL	
UMM	**52.84**%

TABLE II: Word-Emotion Classification Results on Blogs.

where V_{e_k} is the total number of words in the test data $D_{e_k}^{test}$. Therefore smaller the perplexity score, the better is the language model in predicting unseen data.

Perplexity analysis is applied to UMM language models by considering values from the final EM iteration. Figures 2a, 2b and 2c show the results for perplexity analysis on blogs, news (SemEval-07) and tweets respectively. UMM emotion topics were found to have significantly lower perplexity than those of sLDA on all the three data sets, suggesting that UMM is more effective than sLDA in capturing the emotional characteristics of the documents.

D. Word-Emotion Classification Results

Word classification results on Blog data appear in Table II. Here the results are the average overall F-scores obtained over the five folds (five fold cross validation). The proposed UMM lexicon performed significantly better than GPELs (WNA, NRC and ESN) and the baseline DSELs (PMI, slDA and WED). This evaluation clearly suggests that GPELs in general are inadequate for emotion detection, due to poor coverage of domain vocabulary. The assumption of DSELs such as WED and sLDA- that is that documents exhibit multiple emotions- proved to be less effective for predicting the emotion of a word in a context.PMI performed the best among the baselines by far; however the proposed UMM's ability to penalize emotionally neutral words resulted in the best performance in predicting emotions at word-level.

E. Document-Emotion Ranking Results

The document-emotion ranking results for DSELs on news headlines and events captured by tweets are shown in Tables IIIa and IIIb. Comparing the results for sLDA and WED lexicons on both the corpora suggest that they are more effective when the training documents exhibit multiple emotion characteristics as in SemEval-07. On the other hand PMI gives better performance, when documents exhibit single emotion characteristics such as in tweets. However the ability of UMM to quantify emotionality and neutrality of words resulted in effective discrimination and ordering of document level emotion associations across both the corpora.

Method	MAP	MRR
Baseline DSELs		
PMI	64.66%	30.53%
WED	78.10%	53.08%
sLDA	67.44%	35.42%
Proposed DSEL		
UMM	**80.33**%	**56.05**%

(a) News (SemEval-07)

Method	MAP	MRR
Baseline DSELs		
PMI	57.96%	52.56%
WED	50.07%	46.30%
sLDA	56.75%	48.73%
Proposed DSEL		
UMM	**61.7**%	**57.27**%

(b) Events

TABLE III: Document-Emotion Ranking Results

IX. Conclusions and Future work

In this paper we comparatively evaluate both general-purpose lexicons (GPELs) and domain-specific lexicons (DSELs) for emotion detection from text. We introduced a generative unigram mixture model to learn a lexicon which can jointly model both emotionality and neutrality of documents at word level. Results from a comprehensive study of existing and proposed lexicons on emotion detection tasks on benchmark data sets confirm that DSELs have significant performance gains over GPELs. Closer examination of DSEL results show that the proposed lexicon outperformed those generated by state-of-the-art techniques like PMI and supervised LDA in all emotion detection tasks. A deeper empirical analysis suggest that the proposed method generates emotion language models (topics) that have significantly lower perplexity compared to those from supervised LDA. In future we plan to extend the proposed lexicon generation method to learn multi-word-emotion lexicons (i.e. bigram and trigram) following the recent trend in multi-word sentiment and emotion detection [11]. We also plan to utilize the knowledge of the proposed DSEL in conjunction with knowledge bases such as SenticNet and EmoSenticNet to extract effective features to represent documents for emotion classification.

References

[1] A. Esuli, S. Baccianella, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proc of LREC*, pp 18 -38, 2010.

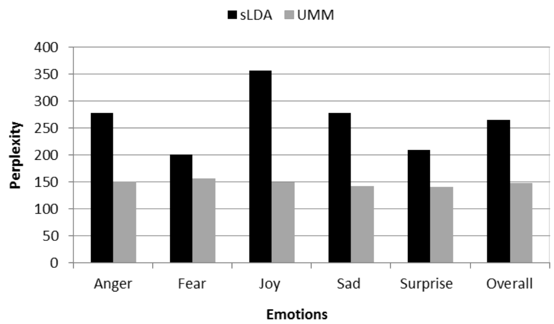
[2] C. Strapparava and A. Valitutti, "Wordnet-affect: An affective extension of wordnet," in *Proc of LREC, pp. 1083-1086*, 2004.

[3] S. M. Mohammad, "#emotional tweets," in *Proc of The First Joint Conference on Lexical and Computational Semantics, pp. 246-255*, 2012.

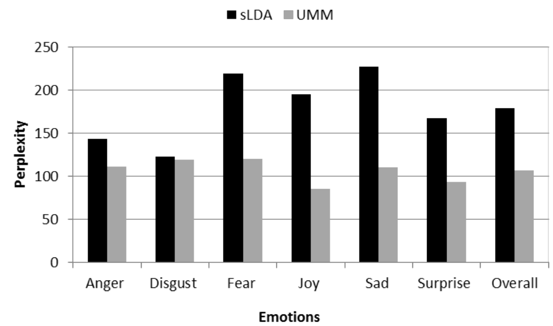
[4] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," *World Wide Web, Vol 17, pp. 723-742*, 2014.

[5] C. Yang, K. H. Y. Lin, and H. H. Chen, "Emotion classification using web blog corpora," in *Proc of the IEEE/WIC/ACM International Conference on Web Intelligence*, ser. WI '07, 2007.

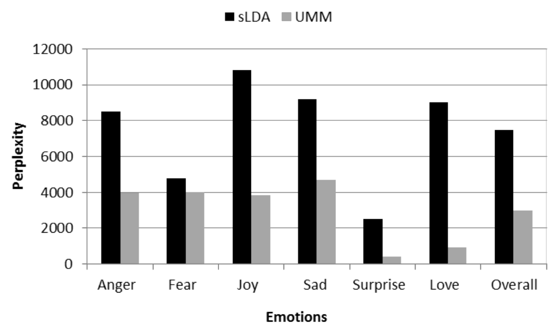
[6] M. Yang, B. Peng, Z. Chen, and a. K. C. Dingju Zhu, "A topic model for building fine-grained domain-specific emotion lexicon," in *Proc of the 52nd Annual Meeting of the Assoc for Computational Linguistics, pp. 421-426*, 2014.



(a) Blogs



(b) News (SemEval-07)



(c) Twitter

Fig. 2: Perplexity scores for emotion topics. UMM generated emotion topics obtained significantly lower perplexity compared to sLDA generated emotion topics.

[7] Y. Chen and S. Skiena, "Building sentiment lexicon for all major loanguages." in *Proc of the 52nd Annual Meeting of the Assoc for Computational Linguistics, pp. 383-389*, 2014.

[8] P. Ekman, "An argument for basic emotions," *Cognition and Emotion, 6(3), pp. 169-200*, 1992.

[9] S. M. Mohammad and P. Turney, "Crowdsourcing a word-emotion association lexicon," *Computational Intelligence, 29(3), pp. 436-465*, 2013.

[10] E. Cambria, S. Poria, R. Bajpai, and B. Schuller, "Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives," in *COLING, Osaka (2016)*.

[11] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems 31(2), pp. 102-107*, 2016.

[12] H. Liu and P. Singh., "Conceptnet- a practical commonsense reasoning tool-kit," *BT Technology Journal, 22(4), pp. 211-226*, 2004.

[13] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G.-B. Huang, "Emosenticspace: A novel framework for affective common-sense reasoning," *Knowledge-Based Systems 69, pp. 108-123*, 2014.

[14] J. Staiano and M. Guerini, "Depechemood: a lexicon for emotion analysis from crowd-annotated news," in *Proc of the 52nd Annual Meeting of the Assoc for Computational Linguistics, pp. 427-433*, 2014.

[15] K. Song, S. Feng, W. Gao, D. Wang, L. Chen, and C. Zhang, "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.

[16] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, and Y. Yu, "Mining social emotions from affective text," in *IEEE Trans. Knowledge and Data Eng. 24(9), pp. 1658-1670*, 2012.

[17] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Advances in Neural Information Processing Systems, pp. 121-128*, 2007.