Construction of Emotional Lexicon Using Potts Model

Braja Gopal Patra[*], **Hiroya Takamura**[+], **Dipankar Das**[#],
Manabu Okumura[+] **and Sivaji Bandyopadhyay**[*]
[*]Dept. of Computer Science & Engineering, Jadavpur University, Kolkata, India
[#]Dept. of Computer Science & Engineering, NIT Meghalaya, India
[+]Precision and Intelligence Laboratory, Tokyo Institute of Technology, Japan

```
brajagopal.cse@gmail.com, takamura@pi.titech.ac.jp,
dipankar.dipnil2005@gmail.com, oku@pi.titech.ac.jp,
           sivaji_cse_ju@yahoo.com
```

Abstract

Emotion is an instinctive state of mind aroused by some specific objects or situation. Exchange of textual information is an important medium for communication and contains a rich set of emotional expressions. The computational approaches to emotion analysis in textual data require annotated lexicons with polarity tags. In this paper we propose a novel method for constructing emotion lexicon annotated with Ekman's six basic emotion classes (anger, disgust, fear, happy, sad and surprise). We adopt the Potts model for the probability modeling of the lexical network. The lexical network has been constructed by connecting each pair of words in which one of the two words appears in the gloss of the other. Starting with a small number of emotional seed words, the emotional categories of other words have been determined. With manual checking of top 200 words from each class an average precision of 85.41% has been achieved.

1 Introduction

Sentiment analysis and classification from electronic text is a hard semantic disambiguation problem (Das and Bandyopadhyay, 2010). Many recent researches have been conducted in the fields of sentiment extraction (Kim et al., 2012; Taboada et al., 2011), opinion mining, summarization (Aman and Szpakowicz, 2007; Das et al., 2012; Yang et al., 2007; Quan and Ren, 2010) and each of which has a variety of potentially valuable applications. For example, we can efficiently collect people's opinion on a new rule enforced by the Government from Blog sites and at the same time be able to grasp their emotion without having to read their comments. An imperative resource for such kind of emotional analysis is an emotion lexicon annotated with several emotional classes like *happy, sad, fear, anger, surprise* and *disgust*. In the previous example, frequent appearance of words from the *happy* class in a blog document would imply that the writer of the comment is quite happy with the new rule proposed by the Government.

Several works have been conducted on building emotional corpora in different languages such as in English (Aman and Szpakowicz, 2007), Chinese (Yang et al., 2007; Quan and Ren, 2010), and Bengali (Das and Bandyopadhyay, 2010) etc. All these works have focused on developing sentiment lexicon with three sentiment classes. For instance, Takamura et al. (2005) have developed a lexicon of emotion words tagged with the classes *desirable* and *undesirable* using Spin model. A number of other polarity sentiment lexicons are available in English such as SentiWordNet 3.0 (Esuli et al., 2010), Subjectivity Word List (Wilson et al., 2005), WordNet-Affect list (Strapparava et al., 2004), Taboada's adjective list (Taboada et al., 2006). On the other hand, several polarity sentiment lexicons have been developed in different languages like Hindi, Bengali and Telegu (Das and Bandyopadhyay, 2010), Japanese (Torii et al., 2012) etc.

Among all these publicly available sentiment lexicons, SentiWordNet is one of the well-known and widely used ones (number of citations is higher than other resources[1]), having been uti-

---

[1] http://citeseerx.ist.psu.edu/index

lized in several applications such as sentiment analysis, opinion mining and emotion analysis.

Undoubtedly, manual compilation is the best way to create such an emotion lexicon but is much expensive in terms of time and human effort. Thus, the objective of the present paper is to develop a method for automatically creating such a list of words from the glosses of a dictionary, as well as from a thesaurus and a corpus. For this purpose, we have used the *Potts model*, a probabilistic model for lexical network. In the lexical network, each node has one of the three orientation values and the neighboring nodes tend to have the same value. For each of the emotion classes, we estimate the states of the nodes indicating the semantic orientation of each class. However, the proposed method does not deal with words that do not appear in the lexical network.

We have classified the words into six emotion classes using Potts model. First, the manual evaluation has been done to get the accuracies. Then we have automatically calculated accuracies comparing with the WordNet Affect list. We have also classified the words into two classes (positive and negative) and the accuracy is evaluated using the SentiWordNet. The generated emotion lexicon in English also contains the parts of speech (adjective, adverb, noun and verb) information of the emotion words as well as their emotional classes.

The rest of the paper is organized in the following manner. Section 2 discusses briefly the resources available till date. Section 3 provides an overview on Potts model. Section 4 describes the implementation of Potts Model for the construction of our emotion lexicon. Section 5 presents the experiments with detail analysis. Finally, conclusions are drawn and future directions are presented in Section 6.

2   Related Work

Takamura et al. (2005) extracted semantic orientation of words according to the spin model, where the semantic orientation of words propagates in two possible directions like electrons. Electrons propagate their spin direction to neighboring electrons until the system reaches a stable configuration. They have constructed a lexical network by connecting pairs of words. In each pair either word appears in the gloss of the other. They have applied spin model iteratively till energy of the system is minimized.

Esuli and Sebastiani's (2006) approach to develop the SentiWordNet is an adaptation to synset classification based on the training of ternary classifiers for deciding positive and negative (PN) polarity. Each of the ternary classifiers is generated using the Semi-supervised rules.

Strapparava and Valitutti (2004) developed the WORDNET-AFFECT, a lexical resource that assigns to a number of WORDNET synsets one or more affective labels such as emotion, mood, trait, cognitive state, physical state, behavior, attitude and sensation etc. They have prepared a preliminary resource named as AFFECT, then projected part of the affective information from the AFFECT database onto the corresponding senses of WORDNET-AFFECT.

Das and Bandyopadhyay, (2010) created the SentiWordNet for Indian Languages like Hindi, Bengali and Telegu by multiple computational approaches like WordNet based, dictionary based, corpus based or generative approaches. They have used the Bilingual corpus and generated the SentiWordNet(s) for the Indian languages from the English sentiment lexicon merged from the English SentiWordNet and the Subjectivity Word List.

Das et al., (2012) presented a task of developing an emotion lexicon. A lexical network has been developed on the freely available ISEAR dataset using the co-occurrence threshold. They classified words into seven categories, i.e., *anger, disgust, fear, guilt, joy, sadness* and *shame*. SVM and Fuzzy C-mean classifier have been used for the classification. They also computed the precision of top 100 words and reported 95% precision for seven emotion classes.

3   Potts Model

We have employed the Potts model which is a generalization of Ising model (Nishimori, 2001). If a variable has more than two values and there is no ordering relation between the values, such network is called a Potts Model (Wu, 1982). Potts model has been a subject of increasing research interest in the recent years. In this section we present the mathematics of Potts model. Potts model has been used in several applications such as extraction of semantic orientations of phrases from dictionary (Takamura et al., 2007).

It has been observed that the types of similarity or prior polarity scores do not completely solve the problem of classifying emotional words. In fact, finer details are revealed by so-called contextual polarity classification, because

the same textual content can be presented with different emotional slants (Grefenstette et al., 2005). For example, the word 'succumb' can trigger a mix of multiple emotions: 'fear' as well as 'sad'. Considering word-wise emotion identification as a multi-label text classification problem, we deploy a Potts model based classification technique.

3.1 Introduction to Potts Model

Suppose a network of nodes and weighted edges is given. The states of the nodes are collectively represented by n. The weight between nodes i and j is represented by w_{ij}.

The energy function is represented as $H(n)$, which indicates the state of the whole network:

$$H(n) = -\beta \sum_{ij} w_{ij} \delta(n_i, n_j) + \alpha \sum_{i \in L} -\delta(n_i, a_i)$$

where β is a constant called *the inverse-temperature*, L is the set of the indices for the observed variables, a_i is the state of each observed variable indexed by i, and α is a positive constant representing a weight on labeled data. Function δ returns 1 if two arguments are equal to each other, 0 otherwise. The state is penalized if n_i ($i \in L$) is different from a_i. Using $H(n)$, the probability distribution of the network can be represented as $P(n) = \exp\{-H(n)\}/Z$, where Z is a normalization factor.

However, it is computationally difficult to exactly estimate the state of this network. We resort to a mean-field approximation method. In the method, $P(n)$ is replaced by factorized function $\rho(n) = \prod_i \rho_i n(i)$. Then we can obtain the function with the smallest value of the variational free energy:

$$F(n) = \sum_n P(n)H(n) - \sum_n -P(n)\log P(n)$$
$$= -\alpha \sum_i \sum_{n_i} \rho_i(n_i)\delta(n_i, a_i)$$
$$-\beta \sum_{ij} \sum_{n_i, n_j} \rho_i(n_i)\,\rho_j(n_j)w_{ij}\delta(n_i, n_j)$$
$$-\sum_i \sum_{n_i} -\rho_i(n_i)\log \rho_i(n_i)$$

By minimizing $F(n)$ under the condition that $\forall_i, \sum_{n_i} \rho_i(n_i) = 1$, we obtain the following fixed point equation for $i \in L$:

$$\rho_i(n) = \frac{\exp(\alpha\delta(n,a_i)+\beta\sum_i w_{ij}\rho_j(n))}{\sum_m \exp(\alpha\delta(m,a_i)+\beta\sum_i w_{ij}\rho_j(m))}$$

The fixed point equation for $i \notin L$ can be obtained by removing $\alpha\delta(n, a_i)$ from above.

This fixed point equation is solved by an iterative computation. After the computation, we obtain the function $\prod_i \rho_i(n_i)$. When the number of classes is two, the Potts model in this formulation is equivalent to the mean-field Ising model (Nishimori, 2001).

4 Potts Model for Construction of Emotional Lexicon

In this section we describe the methodologies adopted to develop the emotional lexicon wherein words are classified into six emotional classes.

4.1 Constructing Lexical Networks

We have constructed a lexical network which has been termed as gloss network (Takamura et al., 2005). This network is developed by linking two words if one appears in the gloss of other word. Each link belongs to one of the two groups: same orientation links (SL) and different orientation links (DL). If at least one word precedes a negation word (e.g., not) in the gloss of the other word, the said link is considered as a different-orientation link. Otherwise the link is a same-orientation link. Lexical Network contains 88015 words collected from the dictionary. Statistics of the lexical network is shown in Table 1. Next, we assign weights $W = (w_{ij})$ to links as follows:

$$w_{ij} = \begin{cases} \dfrac{1}{\sqrt{d(i)d(j)}} & (e_{ij} \in SL) \\ -\dfrac{1}{\sqrt{d(i)d(j)}} & (e_{ij} \in DL) \\ 0 & otherwise \end{cases}$$

where e_{ij} denotes the link between word i and word j, and $d(i)$ denotes the degree of word i, which is actually the number of words linked with word i. Two words without a connection are regarded as connected by a link of weight 0.

Class	No. of words
Adjective	20497
Adverb	3751
Noun	55285
Verb	8482

Table 1. Statistics of Lexical network

We have also constructed another network, *the gloss thesaurus network* (*GT*), by linking syno-

nyms, antonyms and hypernyms, in addition to the above linked words. Only antonym links are in DL.

We enhanced the gloss-thesaurus network with co-occurrence information extracted from corpus. Hatzivassiloglou and McKeown (1997) focused on conjunctive expressions such as "simple and well-received" and "simplistic but well-received", where the former pair of words tend to have the same semantic orientation, and the latter tend to have the opposite orientation. Following their method, we connect two adjectives if the adjectives appear in a conjunctive form in the corpus. If the adjectives are connected by "and", the link belongs to SL. If they are connected by "but", the link belongs to DL. We call this network the *gloss-thesaurus-corpus network* (*GTC*). We have used *gloss-thesaurus-corpus network* in our experiments.

4.2 Classification of Words

Takamura et al., (2007) used the Potts model for extracting semantic orientation of phrases (pair of adjective and a noun): positive, negative or neutral. In contrast to that, we have used the Potts model for identifying the emotional class (es) of a word.

We have used one seed word from each class to start with the experiment. Each seed word is assigned a class manually. We therefore estimate the state of nodes in the lexical network for each class of emotions. The only drawback is that, it could not assign any emotional class to a word which is not present in the lexical network. These words may be referred to as *unseen words*.

The reason of choosing Potts model over Ising model is that Ising model is helpful for modeling a system involving two classes only (i.e. positive and negative), whereas Potts model can be modeled for more than two classes.

5 Evaluation

We performed our experiments with different values of β, ranging from 0.5 to 1 with an interval of 0.1 and achieved best result for $\beta = 0.9$. We also performed experiments with different set of seed words. Fixed seed words are used with different β values. We prepared three lists of seed words containing 6, 12 and 18 words respectively. They were prepared by picking 1, 2 or 3 words from each emotional class respectively.

We have classified the words into six emotional classes with different seed words and dif-

ferent values of β. Then the accuracies were computed manually as well as using the Word-Net Affect lexicon. We also classified the words into two classes, i.e. positive and negative. The accuracies of two classes were calculated using the SentiWordNet.

Classes (Manually checked)	Precision (in %age)
Happy (200)	80.0
Sad (200)	80.5
Surprise (200)	82.0
Angry (200)	92.5
Fear (200)	92.0
Disgust (200)	85.5
Average	85.41

Table 2. Precision (under manual checking of each class)

5.1 Classification Results

Before discussing the accuracy, we would like to make some interesting observations. There are some words that cannot be classified by the classifier, i.e., for these words the probabilities of each class is the same. The number of these words varies with the change of β values and the number of seed words. We also observed similar change in the number of words put into each class.

We have manually checked top 200 words from each class having highest probability and reported the precision in Table 2. We have achieved maximum precision of 92.5% and 92.0% *angry* and *fear* classes respectively. It has been observed that the *happy* class has lowest precision and is about 80%. The precisions of *sad, surprise* and *disgust* classes are 80.5%, 82% and 85.5% respectively. We also performed several experiments by changing the values of β and the varying the number of seed words. The highest precision is achieved with $\beta = 0.9$ and number of seed word kept at 18, i.e., three words from each group.

We observed that the accuracy of *happy* class is low. The reason may be that many words in this class do not have any relation to happy class and such words are basically neutral words or tough words, i.e. these words do not contain any emotions. For example, "*handel*" and "*olivier*" are identified as *happy* words, whereas they do not have any relation with the *happy* class. We also observed that *happy* emotion class contains

677

	Happy	Sad	Surprise	Angry	Fear	Disgust	Neutral
Happy	160	0	4	0	0	0	36
Sad	0	161	1	13	8	15	2
Surprise	5	6	164	2	8	12	3
Anger	0	7	1	185	1	5	1
Fear	0	9	1	1	184	3	0
Disgust	0	11	2	12	3	171	1
Neutral	0	0	0	0	0	0	0

Table 3. Confusion matrix for manual precision checking.

Classes	Precision (in %age)
Happy	50.8
Sad	52.3
Surprise	46.8
Angry	56.0
Fear	51.4
Disgust	35.5
Average	48.8

Table 4. Precision of each class collected by WordNet-Affect.

Classes	Precision (in %age)
Happy	58.9
Sad	65.4

Table 5. Accuracy of each class based on SentiWordNet.

some words from the *surprised* class. For example, the word "*fortuitous*" means happening by a lucky chance. Another example is "*stunning*", which is classified as happy class, but it belongs to surprise class. In case of *sad* word class, we found some words from *fear, angry* and *disgust* classes. *Angry* class comprises some words from *sad*, *fear, disgust* and *neutral* classes. For example, the word "*stink*" is classified as *sad* class where as it belongs to *disgust* class. It does not contain any word from *happy* class. *Fear* and *disgust* classes contain word from all other classes except *happy* class. The details can be found from confusion matrix given in the Table 3.

There are some words which could belong to more than two classes depending on the context/situation. For example, "*shiver*" falls under the class *sad* and *fear*. We have removed these words while calculating the accuracy of the system.

We have also cross checked the accuracies of our system using the WordNet Affect. Here we have classified the words in six emotion classes and the precision is computed by comparing with WordNet-Affect. As WordNet-Affect contains less numbers of emotion words, so we just checked top 100 words from each emotion classes and the precision is given in Table 4. The average precision calculated is 48.8%. This is due to the fact that WordNet-Affect contains less number of words.

We have also classified the words into two classes, i.e. positive and negative. Then we have computed the accuracy of the output using the SentiWordNet 3.0 (Esuli et al., 2010). Approximately 25000 words occur with same probability and those words are removed at the time of testing. The accuracy is given in the Table 5. We have achieved 58.9 % accuracy in positive class or happy class, whereas 65.4% in negative class or sad class.

A shortcoming of this system is that it cannot handle those words which are not present in the lexical network. Error occurs when a non-emotional word is assigned a class.

6   Conclusion and Future Work

A method has been proposed for extracting emotional orientations of words with high accuracy using Potts model. The major contribution in the task is to prepare the emotional lexicon.

There are several directions for future works. One of them is to incorporate the syntactic information. Since importance of each word in a gloss depends on its syntactic role, syntactic information in glosses should be useful for classification.

A single word could belong to multiple classes. So, the identification of those words and representing them in fuzzy classes is one of the crucial research goals to be achieved in future.

Reducing the number of words having same probability in each emotion classes may be another research work. New words that are not listed in the Lexical Network can be updated in later works.

Finally, we are of the opinion that the proposed model is applicable to other tasks in computational linguistics.

Acknowledgement

References

Amitava Das and Sivaji Bandyopadhyay. 2010. SentiWordNet for Indian Languages. In *Proceedings of the 8th Workshop on Asian Language Resources (ALR)*, August, pages 56-63.

Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In *Text, Speech and Dialogue*. Springer Berlin Heidelberg, pages 196-205.

Andrea Esuli and Fabrizio Sebastiani. 2006. Senti-Word-Net: A publicly available lexical resource for opinion mining. In *Proceedings of the Language Resources and Evaluation* (LREC), pages 417-422.

Carlo Strapparava and Alessandro Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon, pages 1083-1086.

Changhua Yang, Kevin Hsin-Yih Lin and Hsin-Hsi Chen. 2007. Building emotion lexicon from weblog corpora. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*. Association for Computational Linguistics, pages 133-136.

Dipankar Das, Soujanya Poria and Sivaji Bandyopadhyay. 2012. A classifier based approach to emotion lexicon construction. In *Proceedings of the Natural Language Processing and Information Systems.* Springer Berlin Heidelberg, *pages* 320-326.

Fa-Yueh Wu. 1982. The Potts Model. *Reviews of Modern Physics*, 54(1):235–268.

Gregory Grefenstette, Yan Qu, James G. Shanahan, David A. Evans. 2004. Coupling niche browsers and affect analysis for an opinion mining application. In *Proceedings of RIAO* (Vol. 4, pp. 186-194).

Hidetoshi Nishimori. 2001. *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 133–140.

Hiroya Takamura, Takashi Inui and Manabu Okumura. 2007. Extracting Semantic Orientations of Phrases from Dictionary. In *Proceedings of the Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT,2007)*, pages 292-299

Maite Taboada, Anthony Caroline and Voll Kimberly. 2006. Creating semantic orientation dictionaries. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genoa, pp. 427–432.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267-307.

Mitra Mohtarami, Hadi Amiri, Man Lan, Thanh P. Tran, and Chew L. Tan. 2012. Sense Sentiment Similarity: An Analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pages 1706-1712.

Paul Ekman 1993. Facial expression and emotion. *American Psychologist*. 48(4):384–392.

Seungyeon Kim, Fuxin Li, Guy Lebanon and Irfan Essa. 2012. Beyond Sentiment: The Manifold of Human Emotions. *arXiv preprint arXiv:1202.1568*.

Stefano Baccianella, Andrea Esuli and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10),* Valletta, Malta, May.

Theresa Wilson, Janyce Wiebe and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the HLT/EMNLP* 2005, Vancouver, Canada.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th ACL and 8th Conference of the EACL*, pages 174–181.

Yoshimitsu Torii, Dipankar Das, Sivaji Bandyopadhyay and Manabu Okumura. 2011. Developing Japanese WordNet affect for analyzing emotions. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis,* pages 80-86.