A Classifier based approach to Emotion Lexicon Construction

Dipankar Das, Soujanya Poria and Sivaji Bandyopadhyay

Department of Computer Science and Engineering, Jadavpur University,
188, Raja S.C. Mullick Road, Kolkata 700 032, India
dipankar.dipnil2005@gmail.com, soujanya.poria@gmail.com, sivaji_cse_ju@yahoo.com

Abstract. The present task of developing an emotion lexicon shows the differences from the existing solutions by considering the definite as well as fuzzy connotation of the emotional words into account. A weighted lexical network has been developed on the freely available ISEAR dataset using the co-occurrence threshold. Two methods were applied on the network, a supervised method that predicts the definite emotion orientations of the words which received close or equal membership values from the first method, Fuzzy c-means clustering. The kernel functions of the two methods were modified based on the similarity based edge weights, Point wise Mutual Information (PMI) and universal Law of Gravitation (LG_r) between the word pairs. The system achieves the accuracy of 85.92% in identifying emotion orientations of the words from the WordNet Affect based lexical network.

Keywords: Emotion orientations, ISEAR, Fuzzy Clustering, SVM, PMI, Law of Gravitation, WordNet Affect.

1 Introduction

Several researchers have contributed their efforts regarding the semantic orientation of words [2] but most of the works are in English and are used in coarse grained sentiment analysis (e.g., *positive*, *negative* or *neutral*). There are some examples like the word, "*succumb*" which triggers mix of multiple emotions (*fear* as well as *sad*) to a reader. Thus, considering the problem of word emotion identification as a multi-label text classification problem, the present attempt of generating emotion lexicon by incorporating the fuzzy nature of emotion is different from the existing solutions in this field. Therefore, the Fuzzy c-means clustering [4] has been employed followed by Support Vector Machine (*SVM*) [5] based classification to accomplish the goals. The objective function of Fuzzy c-means clustering and kernel function of SVM have been modified by adding some more functions like Point wise Mutual Information (*PMI*), universal Law of Gravitation (LG_r) and similarity based edge weights between the word pairs. It has been observed that SVM helps to predict the definite emotion orientations of the words that receive close or equal membership values from the Fuzzy c-means clustering method. The proposed method extracts emotion orientations with high accuracy on the weighted lexical networks.

2 Construction of Lexical Network

A generalized lexical network (*G*) has been constructed from the psychological statements of the ISEAR (International Survey of Emotion Antecedents and Reactions) dataset [6]. The psychological statements contain about 3~4 sentences pre-classified into seven categories of emotion (*anger, disgust, fear, guilt, joy, sadness* and *shame*). A set of standard preprocessing techniques was carried out, *viz.*, *tokenizing, stemming* and *stop word removal* using *Rapidminer's text plugin* [1]. The Stanford Part-of-Speech (POS) Tagger[2] and the *WordNet* stemmer [7] have been used for (*POS*) tagging and lemmatization of the statements. A total of 320 stop words (quite frequent words such as "*any~thing/body*" and "*Wh~en/ile*" etc.) are removed. Negation words include the 6 words ("*not*", "*no*", "*never*" etc.). In addition to the usual negation words, the words and phrases which mean negations in a general sense have also been included (e.g., "*free from*" and "*lack of*"). Finally, to construct the lexical network, two words are linked if one word appears with the other word in a single statement. A total of 449060 words were found.

Co-occurrence based Network (G_{Co}): Co-occurrence identifies the chances of frequent occurrence of two terms in a text corpus alongside each other in a certain order as an indicator of semantic proximity. In contrast to collocation, co-occurrence assumes interdependency of the two terms. A co-occurrence network (G_{Co}) has been developed from the generalized network (***G***) based on the co-occurrence *threshold* > 1 in the whole corpus. The co-occurrence network (G_{Co}) consists of 82,457 words.

WordNet Affect based Network (G_{WA}): The generalized lexical network (***G***) is transformed into another lexicalized network based on the words present in the *Wordnet Affect* [1]. Each word of a statement and its stem form is searched in any of the six *WordNet Affect* lists. If any match is found, the word is tagged as the *Emotion Word* <*EW*>. Two words are linked if one word is an *Emotion Word* <*EW*> and appears in the context of the other word in a statement. This type of *WordNet Affect* based network (G_{WA}) consists of 63,280 words.

Similarity based Weighted Networks: The weights (ρ) are assigned to the edges of the word pairs in the networks based on two types of similarity measure, WordNet distance based similarity (W_{Sim}) and Corpus distance based similarity (C_{Sim}).

WordNet Distance based Similarity (W_{Sim}). English WordNet 2.1[3] has been used to measure the semantic distance between two words. WordNet::Similarity is an open-source package for calculating the lexical similarity between word (or sense) pairs based on variety of other similarity measures.

Corpus Distance based Similarity (C_{Sim}). In case of Corpus distance based similarity (C_{Sim}), we averaged the lexical distances (L_d) of the co-occurred word pairs in the statements without considering their ordering. We consider the inverse values of the W_d and L_d by assuming that the larger distance between two words implies lower similarity and vice versa. The weights are calculated based on the summation of two similarity scores as shown below. The corresponding weighted networks are denoted as G' G_{Co}' and G_{WA}'.

[1] http://rapid-i.com/content/blogcategory/38/69/

[2] http://nlp.stanford.edu/software/tagger.shtml

[3] http://www.d.umn.edu/tpederse/similarity.html

3 Fuzzy Clustering cum SVM based Classification

The Fuzzy c-means clustering algorithm [4] followed by a SVM based supervised classification [5] has been employed by modifying the objective function and kernel functions using three more functions like Point of Mutual Information (*PMI*) (α), universal Law of Gravitation (LG_r) (β) and similarity based edge weights (ρ). The PMI identifies the association between two words (w_1 and w_2) while the universal Law of gravitation (LG_r) incorporates the lexical affinity between two words for belonging to an emotion cluster. Considering the lexical network as the universe of words, the word pairs as the two masses and the root mean square of their lexical distances (L_d) as the distance between the masses, the affinity or attraction between the word masses has been calculated using the notion of Gravitational Force (F). The Gravitational constant G has been calculated based on the average degree of the nodes present in that network.

$$PMI, \alpha = \log \frac{p(w_1, w_2)}{p(w_1).p(w_2)}, \quad F, \beta = G.\frac{m_1 m_2}{r^2},$$

where $m_1 = f_{count}$ (w_1, w_2), number of times the word w_1 appears before the word w_2 in the corpus and $m_2 = f_{count}$ (w_2, w_1), number of times the word w_2 appears before the word w_1 in the corpus and the distance r, root mean square of the lexical distances (L_d) between an word pair is,

$$r = \sqrt{\frac{\sum_{i=1}^{n} L^2_{d_i}}{n_{co}}}$$

, where n_{co} is the number of Co-occurrences.

The modified objective functions (J) of the Fuzzy c-means clustering is as follows,

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} \mu_{ik}{}^{p} (\| x_k - v_i \|)^2 + (\alpha + \beta + \rho) \sum_{i=1}^{c}\sum_{k=1}^{N} \mu_{ik}{}^{p} (\sum_{x_r \in N_k} (\| x_k - v_i \|)^2), \; p>1$$

where, p, the exponential weight influences the degree of fuzziness of the membership function. x_k is the co-ordinate in an n-dimensional space of k^{th} point where the values of n features are mapped into n-dimensional space such as $x_k = (p_1, p_2, p_3, \ldots, p_n)$ where p_i is the value of the i^{th} feature. The other functions are as follows,
To acquire the optimality, the modified objective function is differentiated with respect to $\mathbf{v}_i$ (for fixed u_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n$) and to u_{ij} (for fixed $\mathbf{v}_i$, $i = 1, \ldots, k$), $\frac{\partial J}{\partial v_i} = 0$. If we assume $\| x_k - v_i \|^2 = D_{ik}$ and $(\sum_{x_r \in N_k} (\| x_r - v_i \|)^2 = y_i$, the modified objective function is turned into the following equation where the Lagrange multiplier (λ) is introduced for solving it.

$$F_m = \sum_{i=1}^{c}\sum_{k=1}^{N} (\mu_{ik}{}^{p} D_{ik} + (\alpha + \beta + \rho) \mu_{ik}{}^{p} y_i) + \lambda (1 - \sum_{i=1}^{c} \mu_{ik})$$

If the modified objective function is differentiated with respect to the membership variable μ_{ik}, the membership value is obtained as follows,

$$\frac{\partial F_m}{\partial \mu_{ik}} = p \mu_{ik}{}^{p-1}. D_{ik} + (\alpha + \beta + \rho) \mu_{ik}{}^{p-1} y_i - \lambda = 0 \text{ and } \frac{\partial F_m}{\partial \mu_{ik}} = 0$$

$$\mu_{ik} = \frac{\lambda}{p(D_{ik} + (\alpha+\beta+\rho)y_i)}, \text{ where } \sum_{i=1}^{c} \mu_{ik} = 1,$$

$$\lambda = \frac{p}{\left(\sum_{j=1}^{c} \left(\frac{1}{(D_{jk} + (\alpha+\beta+\rho)y_j)}\right)^{1/p-1}\right)^{p-1}}, \mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{ik} + (\alpha+\beta+\rho)y_i}{D_{jk} + (\alpha+\beta+\rho)y_j}\right)^{1/p-1}}$$

The centre of the cluster is identified using the following equation.

$$\frac{\partial F_m}{\partial v_i} = 0, \quad v_i = \frac{\sum_{k=1}^{N} \mu_{ik}{}^{p} (x_k + (\alpha+\beta+\rho) \sum_{x_r \in N_k} x_r)}{(1+\alpha+\beta+\rho) \sum_{k=1}^{N} \mu_{ik}{}^{p}}$$

The hyper parameter was selected by using the universal law of gravitation and the Co-occurrence *threshold* >1. The previous results show that the clustering technique can help to decrease the complexity of SVM training [8]. Thus, the *SVM* based classifier has been employed to identify the definite class for an emotion word by modifying its kernel function (using α, β and ρ, the similar functions used in Fuzzy clustering), dividing the training data into sections and excluding the set of clusters with minor probability for support vectors. For the clusters of mixed and uniform category, the support vectors by SVM are extracted and formed into reduced clusters. The best feature set for the classifier has been identified based on the performance of the classifier in terms of accuracy of the classifier. Information Gain Based Pruning (*IGBP*) was carried out to remove the words (e.g. *game*, *gather*, *seem* etc.) that do not play any contributory role in the classification. Features with high Information Gain reduce the uncertainty about the class to the maximum. The input vectors have been generated from the concept template of the statements. Each of the concept template contains the *lexical* pattern *p* using a context window that contains the words around the left and the right of the *Emotion Word*, *<EW>*, e.g.,

p = [*l*-i…*l-3l-2 l-1 <EW> ...</EW> l+1 l+2 l+3…l+i*],

where, *l±i* are considered as the *context* of *p* as the contexts are good predictors for emotion in a corpus. Equivalent emotional seed words are used to learn the context rules that contain information about the equality of the words in context.

4 Experiments

The proposed method is evaluated in terms of precision for the words that are classified with high confidence. In case of SVM, it was found that the top 100 words of each emotion classes achieved 82.78% accuracy. Therefore, the membership value (μ_{ik}) of each cluster as identified during fuzzy clustering was regarded as a confidence measure and incorporated into the SVM. Then, the words with the highest membership values are evaluated with respect to a cluster and the top 100 words for seven emotion classes achieved more than 95% accuracy. The results (in Table 1) show that the membership value of each cluster can work as a confidence measure of the classification as well as it enhances the precision on several lexical networks. It has been observed that the system performs better on the weighted lexical networks in comparison with their ordinary versions. It implies that the similarity based scoring influences the fuzzy clustering as well as the SVM to predict their emotion orientation. Additionally, further experiments conducted using the SVM alone on these networks reveal that the fuzzy clustering helps the SVM to clarify the lexical ambiguities. In most cases, the co-occurrence and the similarity based scoring information from the corpus improve accuracy.

Table 1. Precision of different Classifiers (in %) for top 100 words on various networks.

Approaches	**G**	**G'**	G_{WA}	G_{WA}'	G_{Co}	G_{Co}'
SVM	83.22	85.19	88.23	91.67	84.10	86.77
Fuzzy c-means +SVM	88.01	90.78	92.56	95.02	87.44	91.60

Impact of Psychological Features: Feature plays a crucial role in any machine-learning framework. By reviewing the ISEAR dataset, the following features related to emotion and psychology have been selected to accomplish the classification task. Each of the features is assigned with different numeric values as supplied by the ISEAR dataset. Each word associated with the following features is represented as the feature vector. The detail impact of the psychological features is described in Table 2.

A. Background Variables (*Age, Gender, Religion, Occupation, Country*)
B. General Variables (*Intensity, Timing, Longevity*)
C. Physiological Variables (*Ergotropic and Trophotropic Arousals and Felt temperature*)
D. Expressive Behavior (*Movement, Non-verbal and Paralinguistic activity*)

The individual results based on the general and physiological variables show various interesting insights of the variables from the perspective of emotion (e.g., low *intensity* for emotion classes of *shame* and *guilt* and high for *joy*, *fear* and *sadness*) [3] in addition with the less contributing features, background variables and expressive behaviour. Finally, the system achieves 85.92% accuracy on the *WordNet Affect* based weighted lexical network.

Error Analysis: Mainly four types of errors have been investigated. The identification of emotion words present in a statement but in two separate sentences does not always carry the similarity between their lexical patterns and therefore performs poor on the weighted networks. The second error type is that some of the ambiguous words are still present ("*faint*", "*sick*", "*humble*" etc.) even after including the membership value of the Fuzzy clustering algorithm into SVM. The third error type is the presence of implicit emotions ("*Dreams are in the eyes of the children*" do not contain any direct emotion word but contain an emotional sense). The fourth error type is related to the idiomatic expressions. Idiomatic expressions often do not detect the emotion orientation of the words even after including the present text based psychological knowledge. The current model cannot deal with these types of errors. We leave their solutions as future work.

Table 2. Classification accuracies (%) for different feature combinations.

Features	Fuzzy c-means	SVM	Fuzzy c-means +SVM
PMI	47.45	49.59	51.02
LGr	33.27	46.07	50.23
PMI + LGr	53.55	57.82	59.12
PMI + LGr + A	69.78	70.44	75.72
PMI + LGr + B	61.23	64.88	67.21
PMI + LGr + C	56.87	58.22	60.78
PMI + LGr + D	54.90	57.98	59.14
PMI + LGr + A+B	72.09	74.80	76.06
PMI + LGr + A+B+C	76.22	79.78	81.77
PMI + LGr + A+B+D	73.34	75.80	78.25
PMI + LGr + A+B+C+D	79.77	82.10	84.55

5 Conclusion and Future Work

A method has been proposed for extracting the emotion orientations of words with high accuracy on different types of lexical networks and their weighted versions. There are a number of directions for future work. One is the incorporation of syntactic information and other is to identify solutions for the above mentioned error cases. The textual clues related to psychology may be included to improve the performance.

Acknowledgments. The work reported in this paper was supported by a grant from the India-Japan Cooperative Programme (DSTJST) 2009 Research project entitled “Sentiment Analysis where AI meets Psychology” funded by Department of Science and Technology (DST), Government of India.

References

1. Strapparava, C., Valitutti, A.: Wordnet affect: an affective extension of wordnet. Language Resource and Evaluation (2004)
2. Turney, P. D., Littman, Michael L.: Measuring praise and criticism: Inference of semantic orientation from association. ACM TIS, 21(4):315--346 (2003)
3. Das, D., Bandyopadhyay, S.: Analyzing Emotional Statements – Roles of General and Physiological Variables. In the SAAIP Workshop, 5th IJCNLP, 59--67, (2011)
4. Bezdek, J. C.: Pattern Recognition with Fuzzy Objective Function Algoritms, Plenum Press, New York (1981)
5. Joachims, T.: Text Categorization with Support Machines: Learning with Many Relevant Features. In European Conference on Machine Learning (ECML), 137--142 (1998)
6. Scherer K. R.: What are emotions? And how can they be measured? Social Science Information, 44(4):693–727 (2005)
7. Miller, A. G.: WordNet: a lexical database for English. In Communications of the ACM, vol. 38 (11), 39--41 (1995)
8. Cervantes, J., Xiaoou Li, Wen Yu.: Support Vector Machine Classification Based on Fuzzy Clustering for Large Data Sets. MICAI, LNAI 4293, 572--582 (2006)