A Regression Approach to Affective Rating of Chinese Words from ANEW

Wen-Li Wei, Chung-Hsien Wu, and Jen-Chun Lin

Department of Computer Science and Information Engineering,
National Cheng Kung University, Tainan, Taiwan, R.O.C.
{lilijinjin,chunghsienwu,jenchunlin}@gmail.com

Abstract. Affective norms for the words is an important issue in textual emotion recognition application. One problem with existing research is that several studies were rated with a large number of participants, making it difficult to apply to different languages. Moreover, difference in culture across different ethnic groups makes the language/culture-specific affective norms not directly translatable to the applications using different languages. To overcome these problems, in this paper, a new approach to semi-automatic labeling of Chinese affective norms for the 1,034 words included in the affective norms for English words (ANEW) is proposed which use a rating of small number of Chinese words from ontology concept clusters with a regression-based approach for transforming the 1,034 English words' ratings to the corresponding Chinese words' ratings. The experimental result demonstrated that the proposed approach can be practically implemented and provide adequate results.

Keywords: affective norm, ANEW, ontology, regression.

1 Introduction

The aim of affective computing, introduced by Picard in 1997 [1], is to give computers the ability to recognize, express, and in some cases, "have" emotions. Recently, affective computing has many fields of applications in computer science such as human robot interaction and dialogue systems. In terms of emotion representation, in psychologist's definition, the dimensional model is an important way of representing affect. For instance, Thayer [2] proposed a model for emotion description, as shown in Fig. 1. The two-dimensional emotion space was divided into four quadrants in terms of valence and arousal. "Valence" stands for the degree of pleasantness of the emotion, which is typically characterized as a continuous range of affective responses extending from "unpleasant (negative)" to "pleasant (positive)." "Arousal" stands for the level of activation of the emotion, and it is characterized as a range of affective responses extending from "calm" to "excited." Hence, a growing number of emotion recognition studies are using the two-dimensional plane to be a better emotion representation.

S. D´Mello et al. (Eds.): ACII 2011, Part II, LNCS 6975, pp. 121–131, 2011.

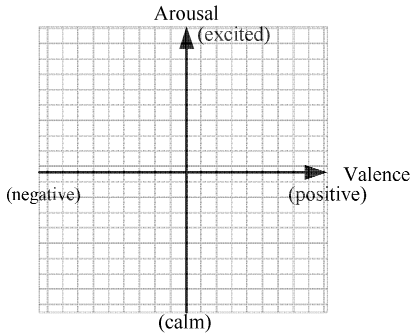

Fig. 1. Thayer's two-dimensional emotion model

Mehrabian [3] and Ambady [4] noted that the affective state can be transmitted by various channels and levels of message including face, voice, and semantic content for communication. In addition to audio or visual emotion recognition [5], [6], text sentimental recognition [7], [8] also plays an important role in many applications of human computer interaction. With the increasing usage of Web blog, the text sentimental analysis has been widely investigated and the word is fundamental for text-based emotion recognition [9]. Hence, this paper will focus on affective norms for the words through the two-dimensional emotion plane.

In text-based emotion recognition, there have been several studies in the literature focusing on how to assess each word in the affective dimensions. For example, the affective norms for English words (ANEW) introduced by Bradley and Lang in 1999 [10] contains a set of normative emotional ratings for 1,034 English words. The ANEW is developed and distributed by national institute of mental health center for the study of emotion and attention in order to provide standardized materials available to researchers in the study of emotion and attention. The goal is to use the self-assessment manikin (SAM), which was originally devised by Lang in 1980 [11] for rating affective words. Based on SAM, subjects rated the words in the ANEW on the dimensions of pleasure, arousal, and dominance. The subjects introduced to participate in the experiment were the students with psychology major. Since the experimental procedure will spend a lot of time and manpower on labeling all words, the rating method and result are difficult to apply to other languages in practice.

Based on the ANEW, Redondo [12] constructed the Spanish version of the ANEW using an adaptation approach. It is similar to the ANEW discussed above, in which the evaluations were also done in the dimensions of valence, arousal and dominance using SAM. In addition, the assessments are based on 720 participants' rating of the 1,034 Spanish words translated from the words included in the ANEW. The purpose of the study is to find if cultural difference exists between the American and the Spanish populations in the ratings of the words included in the ANEW. The findings suggest that the existence of statistical differences between the mean values of the Spanish and American ratings in the three emotional dimensions is remarkable. For example, regarding the arousal dimension, the ANEW words were rated as more activating by Spanish subjects than those by American ones. One explanation for this is that the Spanish subjects interpreted the ratings of the words in terms of a higher emotional reactivity. Although the findings will be useful to analyze the result of subjects' rating in the different languages, the assessments of 1,034 words were also done manually. The experimental procedure on labeling all the words is time-consuming and labor-intensive and the rating results are thus difficult to apply to other languages in practice.

As shown in the above literature review, existing research in affective rating still needs to label all the words. This tedious procedure limits the application on different languages. For this reason, in light of these concerns, the purpose of this paper is to present a new approach to obtain the affective rating results of 1,034 Chinese words which were translated from the words included in the ANEW based on a small set of Chinese words labeling. In detail, the primary methods that we propose are described as follows: (see Fig. 2) (a) divide the 1,034 words included in the ANEW into four quadrants based on the values of arousal and valence to keep the affective dimensional characteristic; (b) cluster the English words in each quadrants through a suggested upper merged ontology (SUMO) concept [13]; (c) randomly select a small set of English words from each SUMO cluster; (d) translate the selected small set of English words into Chinese words according to the mapped WordNet sense; (e) label the valence and arousal of the selected small set of Chinese words (162 Chinese words) through the SAM; (f) generate a regression model for affective norm transformation from each SUMO cluster; and (g) transform the 1,034 English words' ratings into the 1,034 Chinese words' ratings through the trained regression model.

We have organized the rest of this paper as follows: Section 2 describes the purpose of SUMO-based clustering algorithm and provides a thorough description to the proposed affective norms regression. Section 3 presents and discusses the results for a number of analyses. Finally, Section 4 draws the conclusions.

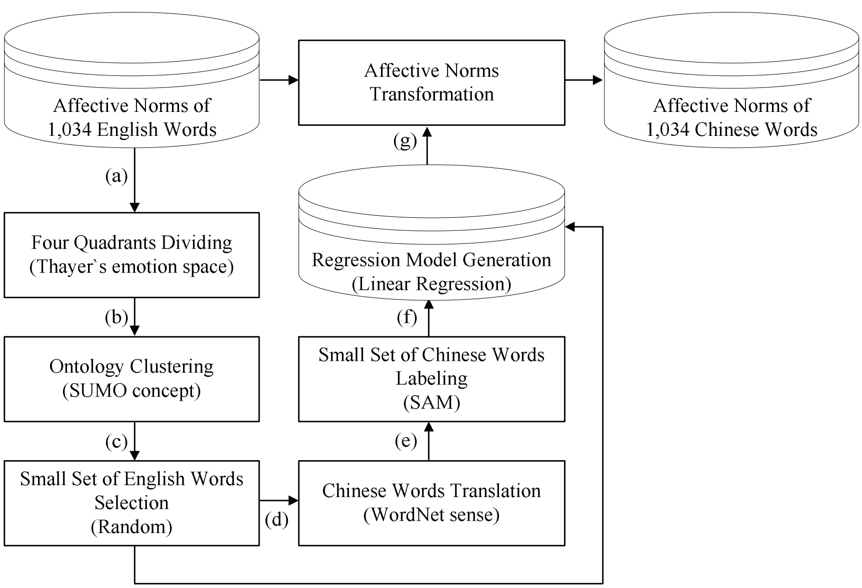

Fig. 2. The flow chart of the proposed method

2 Method

2.1 A SUMO-Based Clustering Algorithm

SUMO, developed by the IEEE standard upper ontology working group and now having a variety of applications in search, linguistics and reasoning, is adopted for word clustering in this work. SUMO is the formal ontology that has been mapped to the entire WordNet lexicons [14]. The concept of SUMO is to link the conceptual hierarchy through the way of an inheritance tree shown in Fig. 3.

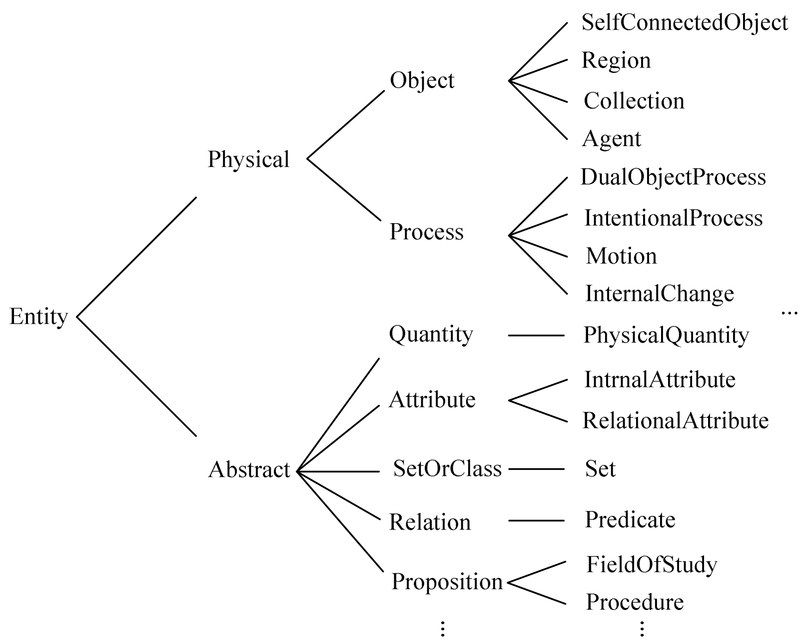

Fig. 3. A hierarchical tree of SUMO concept

The connection structure of the SUMO and WordNet is illustrated in Fig. 4. As shown in Fig. 4, the solid line nodes indicate the SUMO concept; the dotted line nodes represent the WordNet synset; the solid lines denote the SUMO concept relation; and the dotted lines represent the SUMO-WordNet connection. Hence, the relation between SUMO and WordNet enriches WordNet database files by tagging each synset with the corresponding SUMO concept. Using Table 1 as an example, the results reflected in Table 1 indicate that “bouquet” and “chocolate” map to the same “SelfConnectedObject” concept; “delight” and “happy” also map to the same “InternalAttribute” concept. However, in order to keep the unique characteristic for each quadrant, the same SUMO concepts may be used in four quadrants respectively.

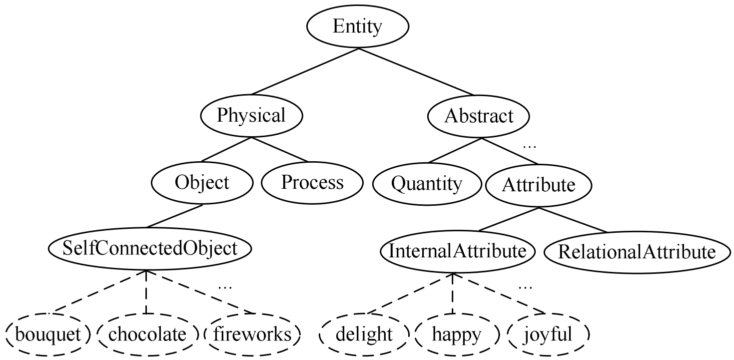

Fig. 4. The relation between SUMO and WordNet

Table 1. The examples illustrate the relation between words and the SUMO concepts

English	Chinese	Quadrant	SUMO Concept	English	Chinese	Quadrant	SUMO concept
bouquet	花束	I	SelfConnectedObject	delight	欣喜	I	InternalAttribute
chocolate	巧克力	I	SelfConnectedObject	happy	高興的	I	InternalAttribute
fireworks	煙火	I	SelfConnectedObject	joyful	快樂的	I	InternalAttribute
bomb	炸彈	II	SelfConnectedObject	angry	生氣的	II	InternalAttribute
gun	槍	II	SelfConnectedObject	insane	發狂的	II	InternalAttribute
poison	毒物	II	SelfConnectedObject	rage	盛怒	II	InternalAttribute
crutch	拐杖	III	SelfConnectedObject	gloom	憂鬱	III	InternalAttribute
pus	膿	III	SelfConnectedObject	sad	悲傷	III	InternalAttribute
scar	傷痕	III	SelfConnectedObject	shy	膽小的	III	InternalAttribute
book	書	IV	SelfConnectedObject	kindness	仁慈	IV	InternalAttribute
clothing	衣服	IV	SelfConnectedObject	relaxed	輕鬆	IV	InternalAttribute
pillow	枕頭	IV	SelfConnectedObject	thankful	感激的	IV	InternalAttribute

In this paper, a SUMO-based clustering algorithm was proposed to help select a small set of words for labeling. We employed the SUMO concept to cluster the 1,034 English words. First, we have divided the 1,034 English words into four quadrants according to the original valence and arousal from ANEW. Next, we assumed that the WordNet synsets with the same SUMO concept has similar valence and arousal values. Hence, we used SUMO to partition the words in each quadrant into several concept groups. In practice, in order to reduce the total number of clusters (i.e. to keep a small set of words for labeling), we partition the words in each quadrant into the number of clusters at the fourth level of the hierarchical tree based on SUMO concept. The fourth level of the hierarchical tree has a total of 39 clusters. However, not all the SUMO concepts have mapping words. For example, in the first quadrant, the “Set” concept has no mapping word. Following SUMO clustering, the first quadrant contains 16 clusters; the second quadrant contains 16 clusters; the third quadrant contains 14 clusters; and the forth quadrant contains 20 clusters. Based on the proposed SUMO-based clustering, we then randomly select at most three words from each SUMO cluster. Hence, we have a total of 162 words selected for labeling.

2.2 Affective Norms Linear Regression

The linear regression has been widely used for regression analysis for its easy performance analysis and reliable prediction performance [15]. There is no need for temporal information or geometric operation. In general, linear regression has been widely used for modeling the relationship between two variables. Hence, in this paper we employ a linear regression function f to transform the source language words' affective ratings x to the target language words' affective ratings y. It can be described mathematically as:

$$y = f(x) \tag{1}$$

where x is the valence or arousal values of the source word and y is the valence or arousal values of the target word. For example, as shown in Fig. 5, we hope to transform each English word's rating to the corresponding Chinese word's rating. That is, the valence and arousal of "vandal" is 2.71 and 6.40 in English, respectively and the transformed valence and arousal of the corresponding Chinese word "摧殘者" is 2.00 and 7.50, respectively.

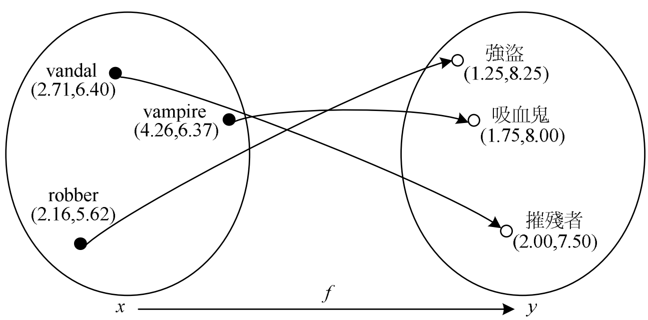

Fig. 5. The idea of regression approach to transform English words' affective ratings to Chinese words' affective ratings

Let f be a linear function, defined by

$$f(x) = \beta_0 + \beta_1 x \tag{2}$$

Equation (2) can be written as a linear regression model

$$y = \beta_0 + \beta_1 x \tag{3}$$

where β_0 and β_1 are the regression coefficients.

Suppose there are n data point pairs (x_i, y_i) in each SUMO cluster, where i=1,…,n. The goal is to calculate the point estimates of β_0 and β_1. The least squares method finds its optimum when the sum of squares, S, is a minimum [16].

$$S = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2 \tag{4}$$

The minimum of the sum of squares is found by setting the gradient to zero.

$$\begin{cases} \dfrac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right] = 0 \\ \dfrac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right] x_i = 0 \end{cases} \tag{5}$$

Thus, we get

$$\begin{cases} \sum_{i=1}^{n} y_i = n \cdot \beta_0 + (\sum_{i=1}^{n} x_i)\beta_1 \\ \sum_{i=1}^{n} x_i y_i = (\sum_{i=1}^{n} x_i)\beta_0 + (\sum_{i=1}^{n} x_i^2)\beta_1 \end{cases} \tag{6}$$

We can then see that β_0 and β_1 are given by

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i / n}{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 / n} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \\ \hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x} \end{aligned} \tag{7}$$

Finally, substitute $\hat{\beta}_0$ and $\hat{\beta}_1$ into equation (3) yields

$$y = \hat{\beta}_0 + \hat{\beta}_1 x \tag{8}$$

Hence, we can transform the source language words' affective ratings x (i.e. valence or arousal of English words) to the target language words' affective ratings y (i.e. valence or arousal of Spanish or Chinese words) through a linear regression model for each SUMO cluster in each quadrant.

3 Experiment

3.1 Experiment Design

Our database consists of two parts. Part I contains valence and arousal rating results of 1,034 words in the original ANEW [10] and Part II contains valence and arousal rating results of 1,034 words in Spanish [12]. Based on the proposed regression model, the experimental results of transformation have two results: English ratings transformed to Spanish ratings (E $\rightarrow$ S) and English ratings transformed to Chinese ratings (E $\rightarrow$ C).

The rating procedure was implemented according to a paper-and-pencil version of the SAM. The participants were four native Chinese speakers: two females and two males, ranging in age from 28 to 30 years old. Following the word selection process, in this experiment, 162 words were selected. Before rating, the instructions provided in [10] were provided to each subject. In order to avoid fatigue, the rating procedure was divided into three sessions during one week. The procedure took approximately half an hour for each subject.

3.2 Experiment Results

To evaluate the proposed SUMO-based clustering algorithm, in this paper, we compare the proposed method with traditional *k*-means clustering algorithm on the $\mathrm{E} \rightarrow \mathrm{S}$. Following the $\mathrm{E} \rightarrow \mathrm{S}$ transformation results, the mean value of Euclidean distance (MED) was used to investigate the difference between the transformed 1,034 Spanish words' ratings and the original 1,034 rated Spanish words' ratings. Suppose there are m clusters in each quadrant and n rating point pairs $(y_{i,j}^{v}, y_{i,j}^{a})$ in each cluster, where i=1,…,m and j=1,…,n. The MED of each quadrant is estimated as

$$MED = \frac{\sum_{i=1}^{m}\left(\sum_{j=1}^{n}\sqrt{(y_{i,j}^{v}-\hat{y}_{i,j}^{v})^2+(y_{i,j}^{a}-\hat{y}_{i,j}^{a})^2}\right)}{m \times n} \tag{9}$$

where $y_{i,j}^{v}$ and $y_{i,j}^{a}$ are the valence and arousal of the original rating for the j-th word in the i-th cluster, respectively. Similarly, $\hat{y}_{i,j}^{v}$ and $\hat{y}_{i,j}^{a}$ are the valence and arousal of the transformed rating for the j-th word in the i-th cluster, respectively.

Table 2 presents the MED between *k*-means and the proposed method in four quadrants and the average of four quadrants. The results indicate that the proposed method outperformed the *k*-means in each quadrant. We also demonstrate the above results in Fig. 6. In Fig. 6 cluster #3 obtained from the *k*-means algorithm and the SUMO concept “InternalChange” in quadrant IV was compared. The *x*-axis and *y*-axis represent valence and arousal of 9-point rating scale, respectively; the black cross indicates the original English words' rating; the red cross represents the original Spanish words' rating; the blue mark means the predicted rating from $\mathrm{E} \rightarrow \mathrm{S}$; and the three circles show the small set of words selected. The result shows the data distribution of the English words in the *k*-means cluster has high compactness based on the distance measure. However, the English words mapped to Spanish words have significant variation. A reasonable explanation is that the *k*-means clustering does not consider the sense of the words. Hence, the result of *k*-means cluster may contain various senses of the words, and will be significantly influenced by the impact of cultural difference. Contrast to the *k*-means method, the English words in the SUMO-based clusters has close sense. Hence the mapping from English words to Spanish words is more consistent and the transformation result MED was smaller than that using *k*-means method. Consequently, the result indicated that the proposed SUMO-based clustering algorithm was useful to affective norms clustering.

Based on the above analysis, we perform E → C with the ratings from a small set of Chinese words according to the SUMO-based clustering algorithm. The purpose of selecting three Chinese words for labeling from each SUMO clustering was to evaluate the performance of the proposed regression approach (i.e. two for regression model training and another one for evaluation). The effect of the proposed regression approach reflected in Table 3 shows that the norm differences between the original Chinese ratings and the transformed Chinese ratings are smaller than that between the original English ratings and the corresponding Chinese ratings. The reason is that the difference in culture across different ethnic groups or different languages is an important factor for affective interpretation. In addition, this finding also confirms that the proposed regression approach can be adopted to transform the ratings in one language to another language.

Table 2. MED between *k*-means and the proposed SUMO-based clustering algorithm

Clustering method \ Quadrant	I	II	III	IV	Avg.
k-means	0.956	0.765	0.805	0.727	0.813
SUMO-based (proposed method)	0.870	0.590	0.735	0.705	0.725

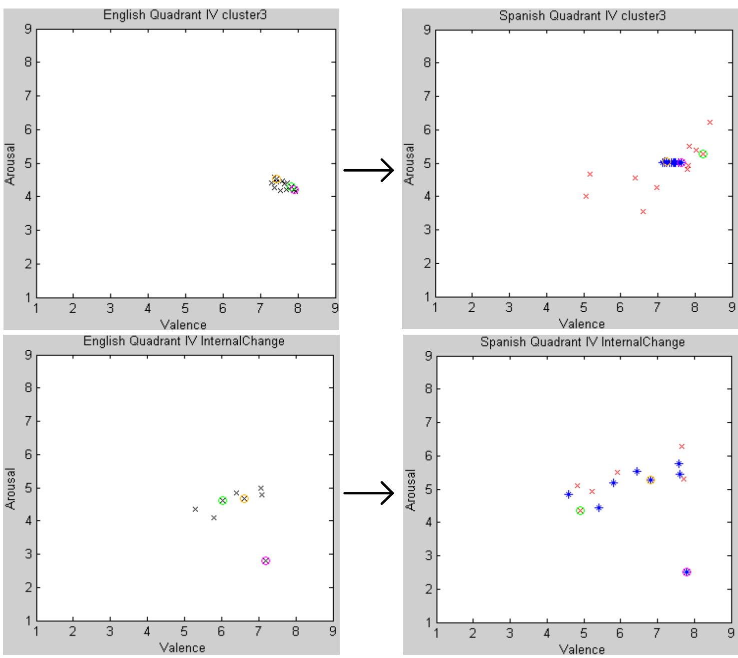

Fig. 6. Examples of results from E → S for *k*-means (top) and SUMO-based approach (bottom)

Table 3. Transformation results for different norms

Norm difference \ Quadrant	I	II	III	IV	Avg.
Rated English ⇔ Rated Chinese	1.260	1.218	1.712	1.605	1.449
Rated Chinese ⇔ Transformed Chinese	0.912	0.893	1.185	0.991	0.995

4 Conclusion

In this paper, we propose a new approach to obtain affective ratings of a large set of Chinese words based on the ratings from a small set of Chinese words using a linear regression model. The experimental result have demonstrated that the proposed approach can be practically implemented and obtain adequate results. Hopefully, it can be applied to affective rating in different languages efficiently and serve as a basis for further study in textual emotion recognition.

References

1. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
2. Thayer, R.E.: The Biopsychology of Mood and Arousal. Oxford University Press, Oxford (1989)
3. Mehrabian, A.: Communication Without Words. Psychol. Today 2(4), 53–56 (1968)
4. Ambady, N., Rosenthal, R.: Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. Psychol. Bull. 111, 256–274 (1992)
5. Wu, C.H., Yeh, J.F., Chuang, Z.J.: Emotion Perception and Recognition from Speech. In: Affective Information Processing, ch. 6, pp. 93–110 (2009)
6. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 39–58 (2009)
7. Wu, C.H., Chuang, Z.J., Lin, Y.C.: Emotion Recognition from Text Using Semantic Labels and Separable Mixture Models. In: ACM Transactions on Asian Language Information Processing, vol. 5, pp. 165–182. ACM, New York (2006)
8. Wu, C.H., Liang, W.B.: Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels. IEEE Transactions on Affective Computing 2(1), 1–12 (2011)
9. Yang, C.H., Lin, K.H.Y., Chen, H.H.: Building Emotion Lexicon from Weblog Corpora. In: Proceedings of 45th Annual Meeting of Association for Computational Linguistics, pp. 133–136. Czech Republic, Prague (2007)
10. Bradley, M.M., Lang, P.J.: Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. Technical Report, The Center for Research in Psychophysiology, University of Florida (1999)
11. Lang, P.J.: Behavioral Treatment and Bio-Behavioral Assessment: Computer Applications: Technology in Mental Health Care Delivery Systems. In: Sidowski, J.B., Johnson, J.H., Williams, T.A. (eds.) pp. 119–137. Ablex Publishing, Norwood (1980)
12. Redondo, J., Fraga, I., Padrón, I., Comesaña, M.: The Spanish Adaptation of ANEW (Affective Norms for English Words). J. Behavior Research Methods 39, 600–605 (2007)

13. The Academia Sinica Bilingual Ontological Wordnet (BOW) - Ontology, http://bow.sinica.edu.tw/ont/
14. Niles, I., Pease, A.: Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Proceedings of the IEEE Conference on Information and Knowledge Engineering, Las Vegas, Nevada, pp. 412–416 (2003)
15. Yang, Y.H., Lin, Y.C., Su, Y.F., Chen, H.H.: A Regression approach to Music Emotion Recognition. IEEE Transactions on Audio, Speech and Language Processing 16(2), 448–457 (2008)
16. Sen, A., Srivastava, M.: Regression Analysis: Theory, Methods, and Applications. Springer, New York (1990)