

LYRIC-BASED SONG EMOTION DETECTION WITH AFFECTIVE LEXICON AND FUZZY CLUSTERING METHOD

Yajie Hu, Xiaoou Chen and Deshun Yang

Peking University

Institute of Computer Science & Technology

{huyajie,chenxiaoou,yangdeshun}@icst.pku.edu.cn

ABSTRACT

A method is proposed for detecting the emotions of Chinese song lyrics based on an affective lexicon. The lexicon is composed of words translated from ANEW and words selected by other means. For each lyric sentence, emotion units, each based on an emotion word in the lexicon, are found out, and the influences of modifiers and tenses on emotion units are taken into consideration. The emotion of a sentence is calculated from its emotion units. To figure out the prominent emotions of a lyric, a fuzzy clustering method is used to group the lyric's sentences according to their emotions. The emotion of a cluster is worked out from that of its sentences considering the individual weight of each sentence. Clusters are weighted according to the weights and confidences of their sentences and singing speeds of sentences are considered as the adjustment of the weights of clusters. Finally, the emotion of the cluster with the highest weight is selected from the prominent emotions as the main emotion of the lyric. The performance of our approach is evaluated through an experiment of emotion classification of 500 Chinese song lyrics.

1. INTRODUCTION

In order to organize and search large song collections by emotions, we need automatic methods for detecting the emotions of songs. Especially, they should work in small devices such as iPod and PDA. At present, much, if not most, research work on song emotion detection was concentrated on the audio signals of songs. For example, a number of algorithms [2,7,9] that classify songs from their acoustic properties were developed.

The lyric of a song, which will be heard and understood by listeners, plays an important part in determining the emotion of the song. Therefore, detecting the emotions of the lyric effectively contributes to detecting the emotions of the song. However, there is now comparatively less research done on methods for detecting the emotions of songs based on lyrics. There has been indeed a very large

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2009 International Society for Music Information Retrieval.

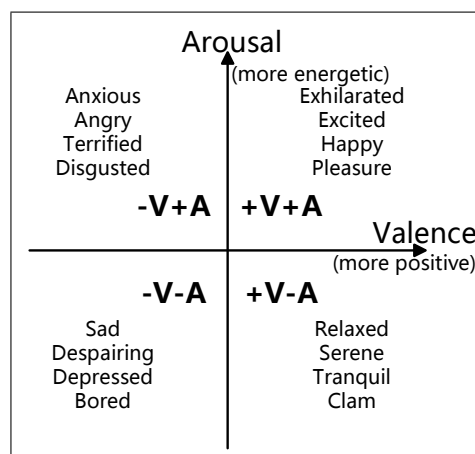


Figure 1. Russell's model of mood

literature already out there on emotion analysis or opinion analysis of text. But, nearly all of them [1, 3, 6] use a one-dimensional model of emotions, such as positive-negative, which is not fine enough to represent lyric emotions which need more dimensions. Lyrics are much smaller in size than other kinds of text, such as Weblogs and reviews, and this makes it hard to detect lyrics' emotions. Being more challenging, lyrics are often abstract and in lyrics, emotions are expressed implicitly.

We propose an approach to detecting the emotions of lyrics based on an affective lexicon. The lexicon is originated from a translated version of ANEW and then extended. According to the lexicon, emotion units (EUs) [13] of a sentence are extracted and the emotion of the sentence is calculated from those EUs.

A lyric generally consists of several sentences and those sentences usually expresses more than one emotions. In order to figure out all the prominent emotions of a lyric, we use a fuzzy clustering method on the sentences of the lyric. The method is robust enough to sustain the noises induced in previous processing steps.

In our approach, Russell's model of mood [11] is adopted, as shown in Figure 1, in which emotions are represented by two dimensions, *valence* and *arousal*. The lyric files we use are in LRC format¹ which have time tags in them and we got the LRC files from the Web. The framework of our approach is illustrated in Figure 2. It consists of three

¹ [http://en.wikipedia.org/wiki/LRC_\(file_format\)](http://en.wikipedia.org/wiki/LRC_(file_format))

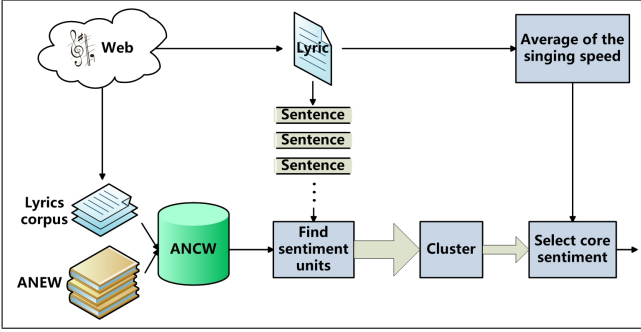


Figure 2. The framework of the proposed approach

main steps: (i) building the affective lexicon (ANCW); (ii) detecting the emotion of a sentence; (iii) integrating the emotions of all sentences.

The rest of this paper is organized as follows. In Section 2, the method for building an affective lexicon is presented. Section 3 describes the method for detecting the emotions of sentences. The approach to integrating the emotions of sentences is described in Section 4. Experiments and discussion are presented in Section 5. Finally, we conclude our work in Section 6.

2. BUILDING THE AFFECTIVE LEXICON

2.1 Translating the Words in ANEW

For analyzing the emotion of Chinese song lyrics, an affective lexicon called ANCW (Affective Norms for Chinese Words) is built from the Bradley’s ANEW [4]. The ANEW list was constructed during psycholinguistic experiments and contains 1,031 words of all four open classes. As described in it, humans assigned scores to each word according to dimensions such as *pleasure*, *arousal*, and *dominance*. The emotional words in ANEW were translated into Chinese and these constitute the basis of ANCW. 10 people took part in the translation work. Each of them was asked to translate all the words in ANEW into Chinese words that he/she thought to be unambiguous and used often in lyrics. The Chinese word that was chosen by the largest number of translators for an ANEW word was picked and added into ANCW. A word may have more than one part of speech (POS), namely performs different functions in different context, and each may have a different emotion. Therefore, the part of speech of an ANCW word must be indicated. The words, the emotions of which in English culture are different from that in Chinese culture, are simply excluded from ANCW. To see if ANCW is consistent with ANEW, we use Meyers’s method [10] to extend ANCW based on a corpus of People’s Daily and the extended ANCW includes 18819 words. Meyers extends ANEW to a word list including 73157 words. The distributions of the emotion classes of the words in the extended ANCW is illustrated in Figure 3. We find that the emotion class distribution of the words in the extended ANCW is similar to the distribution of the words in the extended ANEW. This proves that ANCW is consistent with ANEW and is reasonable.

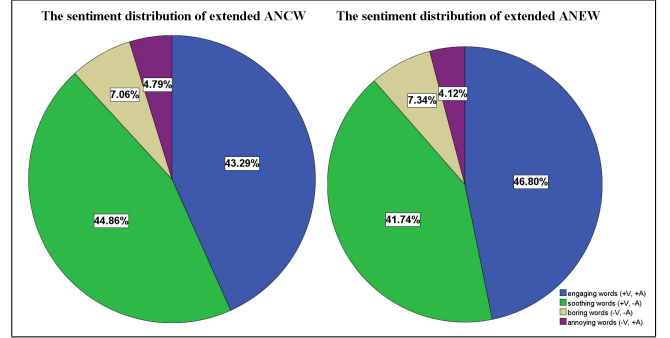


Figure 3. Distributions of the words in the extended ANEW and ANCW

Table 1. The origins of the words in ANCW

Origin	Translated from ANEW	Synonyms	Added by lyrics corpus
# of words	985	2995	71

2.2 Extending ANEW

However, the words translated from ANEW are not sufficient for the purpose of detecting emotions of lyrics so it is necessary to extend ANCW. We extend ANCW in two ways. In one way, with each word in ANCW as a seed, we find out all of its synonyms in TONG YI CI CI LIN². Then, only synonyms with the same part of speech as that of their seed are added to ANCW. In the other way, we extract all constructions of apposition and coordination in a corpus of lyrics (containing 18000 Chinese lyrics) by an off-the-shelf natural language processing tool [8]. If either word in such a construction is in ANCW, its counterpart is added to ANCW. The origins of the words in ANCW is shown in Table 1 and valence-arousal distribution of the words in ANCW is illustrated in Figure 4. To indicate whether a word in ANCW is a translated word from ANEW or a later added word, we attach an *origin* property to each word. Therefore, terms in the affect lexicon have the following form: $\langle word, origin, POS, valence, arousal \rangle$

3. DETECTING THE EMOTION OF A SENTENCE

First, word segmentation, POS annotation and NE recognition are performed for lyrics, with the help of the NLP tool. After stop words removed, the remaining words of a sentence are examined to see if they appear in ANCW, and each of the words that do appear in ANCW constitutes an EU. If there is an adverb that modifies or negates an emotion word, it is included in the corresponding EU as a modifier. We recognize the modifiers of EUs by using the NLP tool. The emotion of an EU is determined as follows:

$$v_u = v_{Word(u)} \cdot m_{Modifier(u),v} \quad (1)$$

$$a_u = a_{Word(u)} \cdot m_{Modifier(u),a} \quad (2)$$

² The lexicon of synonyms is manually built and includes 77,343 terms

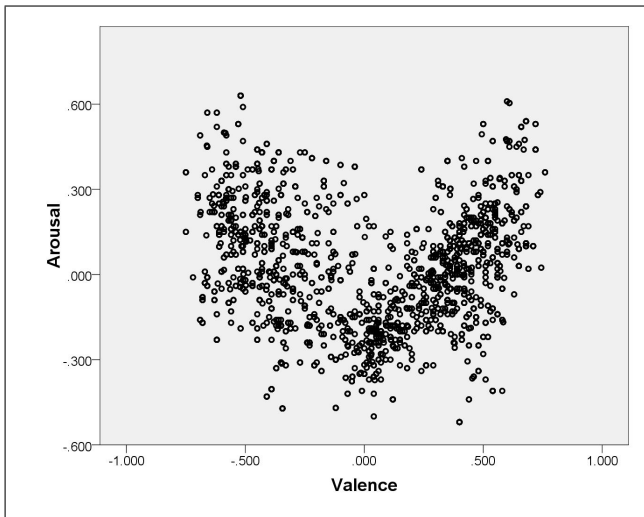


Figure 4. Valence-arousal distribution of the words in ANCW

Where v_u and a_u denote the valence and arousal value of EU u respectively, $v_{Word(u)}$ and $a_{Word(u)}$ denote the valence and arousal value of the EU's emotion word respectively, $m_{Modifier(u),v}$ and $m_{Modifier(u),a}$ denote modifying factors to represent the effect of the EU's modifier on the EU's valence and arousal respectively. $v_{Word(u)}$ and $a_{Word(u)}$, the valence and arousal value of the emotion word are obtained through looking up in ANCW. Sentences that have not any emotion unit are discarded.

We have collected 276 individual modifier words, which cover all the occurrences in the Chinese lyric corpus we use, and a table of modifiers has been set up. According to the polarities and degrees to which modifiers influence the emotions of EUs, we assign each modifier a modifying factor on valence and a modifying factor on arousal. The values of the modifying factors are in the range of $[-1.5, 1.5]$. For a negative modifier adverb, $m_{Modifier(u),v}$ is set to a value in $[-1.5, 0]$ and for a positive modifier adverb, it is set to a value in $[0, 1.5]$.

Tenses influence the emotions of sentences. Some sentences literally depict a happy life or tell romantic stories in one's memory but, actually, the lyric implies the feeling of missing the happiness or romances of past days. Similarly, the sentence with future tense sometimes gives the sense of expectation. Therefore, when we calculate the emotions of sentences, the influence of particular tenses are considered. We use Cheng's method [5] to recognize tenses of sentences and sentences are classified into three categories namely, *past*, *current* and *future*, according to their tenses. A sentence may have more than one EUs. Because the EUs of a sentence always have similar or even identical emotions, they can be unified into one in a simple way, as follows:

$$v_s = \frac{\sum_{u \in U_s} v_u}{|U_s|} \cdot f_{Tense(s),v} \quad (3)$$

Table 2. Adjustment of w_u and r_u of unit u

	Increase when	Decrease when
w_u	u is after adversative words; u is after progressive words; u is in title.	u is before adversative words; u is before progressive words.
r_u	None.	The emotion word's origin is extended; The sentence is adjusted by tense.

$$a_s = \frac{\sum_{u \in U_s} a_u}{|U_s|} \cdot f_{Tense(s),a} \quad (4)$$

where v_s and a_s denote the valence and arousal of sentence s respectively, U_s denotes the set of EUs of the sentence, v_u and a_u denote the valence and arousal of EU $u (u \in U_s)$ respectively, and $f_{Tense(s),v}$ and $f_{Tense(s),a}$ are modifying factors to represent the effect of the tense of the sentence on valence and arousal respectively. The values of the modifying factors representing the effects of tenses on emotions are in the range of $[-1.0, 1.0]$.

There are cases where two sentences (clauses) joined by an adversative or progressive word form an adversative or progressive relation. The following are two examples:

Adversative relation:

You are carefree
But I am at a loss what to do

Progressive relation:

Not only I miss you
But also I love you

Adversative and progressive relations in lyrics tend to remarkably affect the strength of involved EUs in determining the emotions of lyrics. Specifically, an emotion unit following an adversative word in a lyric influences the emotion of the lyric more significantly than a unit before an adversative word does. For example, the EUs in the sentence before *but* is given less weight, while the EUs of the sentence after the adversative word is given more weight. Similarly, in a progressive relation, the emotion unit after the progressive word is thought to be more important. So, a *weight* property is introduced for an EU to represent its strength of influence on lyric emotions. The initial value of weight of an EU is set to 1. A *confidence* property is also attached to an EU. If the emotion word of an EU is a later added word in ANCW, its confidence will be decreased. Also, if the emotion of a sentence is adjusted due to a particular tense, the confidence of its EUs will be decreased. The initial value of confidence of an EU is set to 0. The details of how to adjust the values of the weight and confidence of an EU are shown in Table 2. Accordingly, properties *weight* and *confidence* are also introduced for a sentence, which are calculated from that of its EUs in a simple way as follows:

$$w_s = \sum_{u \in U_s} w_u \quad (5)$$

$$r_s = \sum_{u \in U_s} r_u \quad (6)$$

where w_s and r_s denote the weight and confidence of sentence s respectively, and w_u and r_u denote the weight and confidence of EU u respectively. w_s and r_s are used to determine the main emotion of a lyric in the following processing.

4. INTEGRATING THE EMOTIONS OF ALL SENTENCES

4.1 Challenges

1. Reduce the effect of errors in sentence emotions on the result of the emotions of lyrics.
2. Recognize all the emotions of a lyric on the condition that the lyric has more than one emotion.
3. Select one emotion as the main emotion, if needed, or give a probability to each of the emotions.

4.2 Methodology

In recent years, spectral clustering based on graph partition theories decomposes a document corpus into a number of disjoint clusters which are optimal in terms of some predefined criteria functions. If the sentences of a lyric are considered as documents and the lyric is regarded as the document set, the document clustering technology can conquer the above three challenges. We define an emotion vector space model, where each sentence of a lyric is considered as a node with two dimensions that represent the valence and arousal of an emotion respectively. We choose Wu's fuzzy clustering method [12] because it can cluster the sentences without the need to specify the number of clusters, which meets our demands. Wu's fuzzy clustering method includes three steps: building a fuzzy similarity matrix, generating a maximal tree using Prim algorithm and cutting tree's edges whose weight is lower than a given threshold.

A song usually repeat some sentences. Sometimes the repeated sentences are placed in one line, with each sentence having its own time tag. In other cases, each repeated sentence occupies one line and the line has one time tag. If the repeated sentences are placed in more than one lines, these sentences are bound to form a cluster in the later clustering processing. If the emotions of those repeated sentences were not recognized correctly, subsequent processing will be ruined definitely. Hence, before sentences are clustered, lyrics should be compressed so as to place the iterative sentences in one line, with each sentence having its own time tag.

Having examined hundreds of lyrics, we find that sentences in a lyric always fall into several groups. The sentences of a group have similar emotions which can be uni-

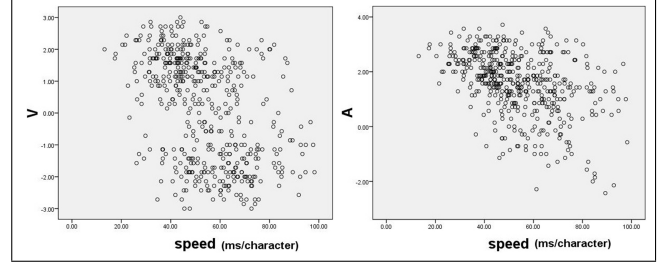


Figure 5. Distribution of speed, V and A

fied to a prominent emotion of the lyric. Therefore, the isolated sentences are mostly noises and will be removed.

There are a dozen of means to measure the similarity between two nodes in vector space. After experiment those means, we select the following means to measure the similarity of the sentences' emotions i, j .

$$Sim_{ij} = 1 - \sigma(|v_i - v_j| + |a_i - a_j|) \quad (7)$$

where v_i, v_j, a_i , and a_j denote the valence and arousal of sentences i and j respectively, and σ is set to 0.3.

The center of a survived cluster is calculated as the weighted mean of emotions of all members of the cluster. The weighted mean is defined as follows:

$$v_c = \frac{\sum_{s \in S_c} v_s \cdot w_s}{|S_c|} \quad (8)$$

$$a_c = \frac{\sum_{s \in S_c} a_s \cdot w_s}{|S_c|} \quad (9)$$

where S_c denotes the set of sentences in cluster c , v_c and a_c denote the valence and arousal respectively of cluster c , and v_s, a_s and w_s denote the valence, arousal and weight respectively of sentence $s (s \in S_c)$.

The weight of cluster c is calculated as follows:

$$w_c = \sum_{s \in S_c} \frac{(\alpha \cdot w_s + \beta \cdot Loop(s))}{-\gamma \cdot r_s + 1} \quad (10)$$

where $Loop(s)$ denotes the number of times sentence $s (s \in S_c)$ repeats, α, β and γ are set to 2, 1, 1, respectively. These constant parameters are adjusted through experimentation and the set of values resulting in the highest F-measure was chosen.

Lyrics we got have time tags and we use these tags to compute the singing speed of sentences in lyrics, which is defined in milliseconds per word. Although, singing speed is not the only determinant of the emotions of lyrics, there is correlation between the singing speed of a song and its emotions, as shown in Figure 5. Hence, we use singing speeds of sentences to re-weight each clustering center. Having analyzed the singing speeds and emotions of the songs in the corpus, we think that Gaussian Model is suitable for expressing the degrees to which different singing speeds influence emotions. The re-weighting is considered as follows:

Table 3. The distribution of the songs corpus

Class	+V,+A	+V,-A	-V,-A	-V,+A
# of lyrics	264	8	174	54

$$w'_c = w_c + \frac{M}{\sqrt{2\pi}\sigma} e^{-\frac{(Speed(c)-\mu_v)^2}{2\sigma^2}} + \frac{M}{\sqrt{2\pi}\sigma} e^{-\frac{(Speed(c)-\mu_a)^2}{2\sigma^2}} \quad (11)$$

$$M = \max(w_c | c \in Lyric) \quad (12)$$

where the μ_v and μ_a are the offset of v and a , respectively. The meaning of σ is the variance of the speed of lyrics. $Lyric$ is the set of emotion clusters of a lyric. $Speed(c)$ is the average speed of sentences in cluster c . Finally, the clustering center with the highest weight is considered the main emotion. If there is a need for the possibility of several emotions, the possibility is computed as follows:

$$p(c) = \frac{w'_c}{\sum_{c \in Lyric} w'_c} \quad (13)$$

5. EXPERIMENTS

Our ultimate goal is to compute the valence and arousal value of lyrics, not to do classification. We do classification for broad classes for the purpose of evaluating our emotion detecting method and comparing the performance of our method with that of other classification methods proposed in the literatures, many of which were for the same broad classes.

5.1 Data Sets

To evaluate the performance of our approach, we collected 981 Chinese songs from the classified catalogue according to emotion in www.kooook.com. These songs are uploaded by netizens and their genres include pop, rock & roll and rap. These songs were labeled by 7 people whose ages are from 23 to 48. Two of them are professors and five are postgraduate students, all native Chinese. Each judge was asked to give only one label to a song. The songs that are labeled by at least 6 judges to the same class are remained. We use these songs' lyrics as the corpus. The distribution of the corpus in four classes is shown in Table 3. Although the number of songs in +V-A class is small, it is not surprising. This phenomenon conforms to the distribution in reality.

5.2 Results

To demonstrate how our approach improves the emotion classification of lyrics in comparison to existing methods, we implemented a emotion classification method based on lyrics with emotion lexicon: Lyricator [10]. Lyricator uses ANEW to extend the emotion lexicon by natural language corpus with a co-occurrence method. Using the extended emotion lexicon, Lyricator computes the emotion of each

Table 4. Evaluation results of Lyricator and our work

Class		Lyricator	Our work
+V+A	Precision	0.5707	0.7098
	Recall	0.7956	0.6856
	F-measure	0.6646	0.6975
+V-A	Precision	0.0089	0.0545
	Recall	0.1250	0.7500
	F-measure	0.0167	0.1017
-V+A	Precision	0.6875	0.6552
	Recall	0.0632	0.3276
	F-measure	0.1158	0.4368
-V-A	Precision	0.0000	0.3125
	Recall	0.0000	0.2778
	F-measure	0.0000	0.2941

sentence of a lyric and the sentence emotion is the mean of emotion values of the emotion words contained in the sentence. The emotion of a lyric is weighted mean of values of the emotions of sentences. The weight is defined as the loop of sentences in the lyric.

To process Chinese lyrics, we translate the lexicon used in Lyricator and implement Lyricator's method. What's more, the parameters are adjusted to gain its best performance. Under the same test corpus that has been mentioned above, we compare Lyricator with our system. Table 4 shows the evaluation results between Lyricator and our work in the same songs corpus. The precision for a class is the number of lyrics correctly labeled the class divided by the total number of lyrics labeled as belonging to the class. The Recall is defined as the number of true positive divided by the total number of lyrics that actually belong to the positive class. The small number of lyrics in +V-A leads to the low precision for this class. Because we have used the wealth of NLP factors and fuzzy clustering method, our method's performance is better than the previous work.

5.3 Discussion

An analysis of the recognition results reveals the following findings:

1. Errors made by the NLP tool are especially salient because lyrics are very different from ordinary texts in word selection and arrangement. It is challenging for the NLP tool to do word segmentation, POS and NE recognition well. For example,

Hope desperation and helpless to fly away

 the NLP tool considered terms "desperation" and "helpless" as verbs while they are actually norms. Without word lemmatization, recognizing POS of words in Chinese is much harder than in English. What's more, it will lead to errors in subsequent processing.
2. Some errors were due to complex and unusual sentence structures, which make it hard for our rather simple method to recognize emotion units correctly.

For example, the subject of a sentence is usually omitted due to the limitation of length of lyrics.

3. It seems that lyrics usually don't express much about *arousal* dimension of emotion. Experimental results show confusion rate between +A and -A is higher than that between +V and -V, suggesting that lyrics don't express much about *arousal* dimension.
4. The emotions of some lyrics were not explicitly expressed, and therefore deduced by human listeners based on his or her knowledge and imagination.

The following sentences come from a typical lyric, the emotions of which are not recognized correctly:

Do you love me? Maybe you love me.

Hanging your head, you are in silence.

Those sentences form the chorus of CherryBoom's *Do You Love Me* and they express intensive emotions. Although it is easy for human listeners to tell the emotions, it is quite difficult for a computer to detect the emotions only literally from the words of the lyric.

6. CONCLUSION

In this paper, we propose an approach to detecting emotions of songs based on lyrics. The approach analyzes the emotion of lyrics with an emotion lexicon, called ANCW. In order to obtain the emotion of a lyric from that of its sentences, we applied a fuzzy clustering technique which can reduce the effect of errors introduced in the process of analyzing emotions of sentences. Finally, we use the mean singing speed of sentences to re-weight the emotion results of clusters. The experimental result is encouraging.

Although this paper handles Chinese lyrics, we also implement an English version of emotion analysis system using English lexicon because our method is not specifically designed for Chinese environment. What's more, the method is unsupervised and training is not needed. Consequently, it takes about two seconds³ to process a lyric and is apt to apply in small devices.

7. REFERENCES

- [1] N. Archak, A. Ghose, and P. Ipeirotis. Show me the money! deriving the pricing power of product features by mining consumer reviews. *In Proceedings of The ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2007.
- [2] D. Bainbridge, S. J. Cunningham, and J. S. Downie. Analysis of queries to a wizard-of-oz mir system: Challenging assumptions about what people really want. *In Proceedings of The 4th International Conference on Music Information Retrieval*, pages 221–222, 2003.
- [3] M. Bansal, C. Cardie, and L. Lee. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. *In Proceedings of The International Conference on Computational Linguistics*, 2008.
- [4] M. M. Bradley and P. J. Lang. Affective norms for english words (anew): Stimuli, instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida, 1999.
- [5] J. Cheng, X. Dai, J. Chen, and Q. Wang. Processing of tense and aspect in chinese-english machine translation. *Application Research of Computer*, Vol 3:79–80, 2004.
- [6] P. Chesley, B. Vincent, L. Xu, and R. Srihari. Using verbs and adjectives to automatically classify blog sentiment. *In AAAI Symposium on Computational Approaches to Analysing Weblogs*, page 27C29, 2006.
- [7] P. Knees, T. Pohle, M. Schedl, and G. Widmer. A music search engine built upon audio-based and web-based similarity measures. *In Proceedings of The 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 23–27, 2007.
- [8] J. Lang, T. Liu, H. Zhang, and S. Li. Ltp: Language technology platform. *The 3rd Student Workshop of Computational Linguistic*, pages 64–68, 2006.
- [9] B. Logan, D. P.W. Ellis, and A. Berenzweig. Toward evaluation techniques for music similarity. *In Proceedings of The 4th International Conference on Music Information Retrieval*, pages 81–85, 2003.
- [10] Owen C. Meyers. A mood-based music classification and exploration system. Master's thesis, Massachusetts Institute of Technology, 2007.
- [11] Russell and James A. A circumplex model of affect. *Journal of Personality and Social Psychology*, Vol 39(6):1161–1178, 1980.
- [12] Z. Shi. *Knowledge Discovery*. Tsinghua University Press, 2002.
- [13] Y. Xia, L. Wang, K. Wong, and M. Xu. Sentiment vector space model for lyric-based song sentiment classification. *In Proceedings of ACL-08: HLT, Short Papers*, pages 133–136, 2008.

³ CPU: 400MHz; Memory: 128MB; OS: Windows Mobile 6.0