MUSIC EMOTION CLASSIFICATION OF CHINESE SONGS BASED ON LYRICS USING TF*IDF AND RHYME

Xing Wang, Xiaoou Chen, Deshun Yang, Yuqian Wu
Institute of Computer Science and Technology, Peking University
{wangxing,chenxiaoou,yangdeshun,wuyuqian}@icst.pku.edu.cn

ABSTRACT

This paper presents the outcomes of research into an automatic classification system based on the lingual part of music. Two novel kinds of short features are extracted from lyrics using tf*idf and rhyme. Meta-learning algorithm is adapted to combine these two sets of features. Results show that our features promote the accuracy of classification and meta-learning algorithm is effective in fusing the two features.

1. INTRODUCTION

Music itself is an expression of emotion. Music emotion plays an important role in music information retrieval and recommendation system. Because of the explosive growth of music libraries, traditional emotion annotation carried out only by experts can no longer satisfies the needs. Thus, automatic recognition of emotions becomes the key to the problem. Though having received increasing attention, it is still at the early stage. [5]

Many methods have been applied to automatic classification of songs' emotions. Traditionally, features such as MFCC and chord are extracted from audio content to build emotion classifiers. Natural language texts are the abstraction of the human cognition, emotion included. Endowed with emotion, lyrics are quite effective in predicting music emotion [2]. As the Internet booms, music related web documents and social tags [13] also provide valuable resources. With the complementarities of features extracted from different modalities,more and more work [6] focus on multimodal classification.

Here we focus on the emotion classification of music based on lyrics only. As it is pointed out in [5], lyrics based approaches are particularly difficult because feature extraction and schemes for emotional labeling of lyrics are non-trivial, especially when considering the complexities involved with disambiguating affect from text. In spite of those difficulties, the linguistic aspects of songs contains lots of emotion information. Firstly, some lexical items in lyrics are highly relevant to certain emotion. Secondly, the pronunciation of words must conform with the emotion, just as in spoken language, loudness and pitch play an important role in identifying the speakers' emotion [4].

In this work, we propose two sets of low dimensional features based on lyrics. We extend the work of Zaanen [11] to get the first set of features based on tf*idf while the other is proposed based on rhymes [1]. Then classifier combination approach is adopted to fuse these two sets of features.

The rest of this paper is organized as follows. We first present related work(Section 2). Then we will describe the taxonomy of emotion(Section 3), features devised for emotion classification(Section 4) and classifier combination approach(Section 5). Experimental results and analyses are presented in Section 6. Section 7 concludes the paper.

2. RELATED WORK

Relatively few research focuses on the use of lyrics as the sole feature for emotion classification. Traditional methods such as the Vector Space Model(VSM) [3] are commonly used in text categorization,but shortcomings exist. Vector space often has very high dimensionality and is noisy, resulting in huge computational cost and low accuracy. We have to turn to features selection techniques.

Recently, more information is integrated into the features, as in Semantic Vector Space Model(SVSM) [14] and Fuzzy Clustering Method(FCM) [15]. In SVSM, all kinds of emotion units are extracted from Lyrics. Emotion unit is composed of an emotional word and the qualifier and negative related to it. The count of emotion unit of each type is used as the feature. FCM analyses the emotion of each sentence based on emotion dictionary ANCW. Then a fuzzy clustering method is implemented to choose the main emotion of a song. Both of them use additional dictionaries and depend too much on the syntactic analysis.However these resources are not mature.

Without the use of additional resources, Zaanen proposes

a new approach to tf*idf feature [11]. He uses tf*idf to measure the correlation between a term and an emotion. Lyrics are transformed to a feature vector, and each dimension of the vector represents the correlation between the lyrics and an emotion. Beside, as far as we know, there's no work focusing on the rhyme of lyrics for classification of emotion.

In this study, we focus on simple and low dimensional features. The simple means that syntactic analysis and additional dictionary which are not mature are not needed; low dimensionality means the features can be processed fast enough in practice. Two sets of features are proposed, one based on the work of Zaanen's and the other based on the rhyme of lyrics. Then we go on to find a way to combine those features.

The methods to fuse these two sets of features can be divided into two categories:features level fusion and classifiers level fusion. In the features level fusion, a new set of feature is generated by operations such as concatenating and features selection. A machine learning algorithm is then used to construct a classifier. In the classifiers level fusion, one classifier is built on each set of features. The final result is obtained by fusing the output of each classifier.

Classifier combination is an effective way to improve the performance [10]. The methods to fuse classifiers generated from different sets of features can be categorized into either base-learning or meta-learning. Meta-learning studies how to choose the right bias dynamically, as opposed to base-learning where the bias is fixed priori, or user parameterized [12].

Combinations with fixed structures are base-learning methods. For example, sum of scores holds the assumption that the label with the biggest sum of score is true label. On the other hand, Combinations which are trained using available training samples are meta-learning methods. Boosting and stacked generalization are examples of meta-learning methods. Boosting algorithm is originally designed for improving the accuracy of classifiers based on one set of features, which does not fit our needs. Stack generalization uses the outputs of basic classifiers as the inputs of the meta-classifier to predict the final result.

3. TAXONOMY

We adopt Thayer's arousal-valence emotion plane [9] as our emotion taxonomy. In this taxonomy, emotion is described by two dimensions:arousal(from calm to excited) and valence(from angry to happy). These two dimensions are most important and universal in expressing emotion [8]. As shown in figure 1, four emotion classes happy, angry, sad, and relaxing are defined according to the four quadrants of the emotion plane.

Figure 1. Thayer's AV model

4. FEATURES

Zaanen proposed a new feature space based on tf*idf [11].The feature vector is short and the method is robust. By taking the part of speech(POS) into consideration, we improve the emotion expressive ability of Zaanen's model. Further more, we make use of rhyme related cues of lyrics which are highly related to expression of emotion.

4.1 pos tf*idf

Some abbreviations are clarified here: POS is part of speech, tf is term frequency and idf is inverse document frequency. In this section, I will describe Zaanen's work first ,and then method for incorporating POS information will be shown.

First, Zaanen merges the lyrics in the training set belonging to emotion e_j into a single document doc_j. In this way, document set D has been produced, with each document in the set corresponding to one emotion class. As shown in equation 1, for a term t_i ,$tf_j(t_i)$ represents the importance of a t_i in the expression of emotion e_j. $idf(t_i)$ represents the ability of a word in distinguishing different emotion as shown in equation 2.

$$tf_j(t_i) = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

where $n_{i,j}$ is the count of term t_i in doc_j.

$$idf(t_i) = \frac{|D|}{|\{doc_j : t_i \in doc_j\}|} \tag{2}$$

Then lyrics lrc_l is represented by feature vector fv_l as shown in equation 3. This feature vector is then used for training classifier and making prediction.

$$fv_l = (f_1, ..., f_c)^T \tag{3}$$

where c is the number of categories and each dimension of the vector is calculated by equation 4.

$$f_j = \sum_{\{k|w_k \in lrc_l\}} tf_j(w_k) * idf(w_k) \tag{4}$$

We know that words of different POS are different in the ability to express emotion. For example, verbs and adjectives are more emotional than articles. In the following of this section, I will describe the method for incorporating POS information.

Based on Zaanen's feature model, we introduce a new feature model which incorporates POS information in lyrics. Instead of combining lyrics belonging to an emotion into one document, we combine them into several documents with each document corresponding to one POS. For each POS, we get four documents corresponding to the emotion taxonomy just like Zaanen's. We get a feature vector of four components for each POS as shown in equation 5. Then we concatenate them to form the final feature vector as shown in equation 6.

$$fv_{l,POS} = (f_{1,POS}, ..., f_{c,POS})^T \quad (5)$$

$$fv_l = (f_{1,verb}, ..., f_{c,verb}, ..., f_{1,noun}, ..., f_{c,noun})^T \quad (6)$$

4.2 rhyme

A rhyme is a repetition of similar sounds in two or more words and is most often used in poems and lyrics. Most Chinese poems obey tail rhyme and lyrics of Chinese songs also obey tail rhyme to some extent.

Rhyme is highly relevant to the emotion expression [1]. Broad sounds such as [a] usually express happiness and excitement while fine sounds such as [i] are related to gentle and sorrow. Broad sounds and fine sounds can be distinguished by the level of obstruction in the vocal tract. Besides the difference between the broad and the fine, intonation also weighs a lot for the expression of emotion. Mandarin has four tones:rises, falls, dips and stays.

There is a system of rhyme in old Chinese songs. It consists of 19 main categories in terms of the broadness and fineness, meanwhile, each main category is divided into three sub-categories by the tones. Then there are totally 57 rhyme categories.

We propose a rhyme frequency(rf) feature based on the rhyme system mentioned above as shown in equation 7.

$$rfv(lrc_j) = (rf_{1,j}, ..., rf_{57,j})^T \quad (7)$$

where

$$rf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (8)$$

This metric measures the importance of rhyme r_i in lyrics lrc_j, with $n_{i,j}$ denoting the number of occurrences of the tail rhyme r_i in lrc_j, divided by number of all tail rhyme occurrences in lrc_j.

5. COMBINATION APPROACH

We fusion these two sets of features on the features level and classifiers level. For the classifiers level fusion, both base-learning combination method and meta-learning combination method are tried.

5.1 Feature Level Fusion

For lyrics lrc_l, we simply concatenate POS tf*idf feature vector and rhyme feature vector to create a new feature vector as shown in equation 9. Then a machine learning algorithm such as SMO is applied to train a classifier and make prediction.

$$fv'_l = (f_{1,verb}, ..., f_{c,verb}, ..., rf_{1,l}, ..., rf_{57,l})^T \quad (9)$$

5.2 Classifier Level Fusion

We use the POS tf*idf feature and rhyme feature as described above for song emotion classification. For each of the two kinds of features, a classification learning algorithm is selected based on experimental results. SMO is chosen for the POS tf*idf feature and Naive Bayes for the rhyme feature.

The combination framework is shown in figure 2. For each instance, basic classifiers output the confidence for each class label. Then combination classifier output the final class label based on the outputs of basic classifiers. The base-learning method and the meta-learning method differ in the implementation of combination classifier.

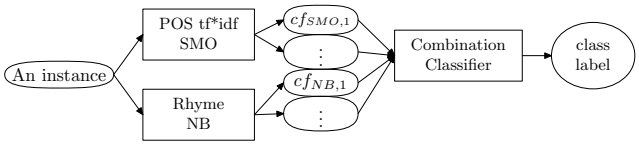

Figure 2. Combination Framework

5.2.1 *Base-learning methods*

For base-learning methods, combination classifier is simple. Combination classifier may choose the class label with the largest confidence value. Besides,a weighted average of confidence value for each class label can be calculated by equation 10, then the class label with the largest cf_i is chosen as the final label. In our study, the latter method is used as the baseline and the parameter setting is $w_1 = w_2 = 0.5$.

$$cf_i = w_1 * cf_{SMO,i} + w_2 * cf_{NB,i} \quad (10)$$

where

$$w_1 + w_2 = 1 \quad (11)$$

Class	+V,+A	+V,-A	-V,-A	-V,+A
# of lyrics	274	5	52	169

Table 1. Distribution of data set

5.2.2 Meta-learning methods

For meta-learning methods, combination classifier is obtained by learning from training set. We use stack generalization as the meta-learning algorithm. The training data of meta-learning is obtained by the following procedure.

Given a training set $T : \{lrc_i, c_i\}_{i=1}^{m}$ for basic classifier, SMO learner and NB learner are applied to training set T_{SMO} and T_{NB} to hypothesis h_{SMO} and h_{NB}.

$$T_{SMO} : \{(F_{POS,i}, c_i)\}_{i=1}^{m} \tag{12}$$

$$T_{NB} : \{(F_{Rhyme,i}, c_i)\}_{i=1}^{m} \tag{13}$$

$F_{POS,i}$ and $F_{Rhyme,i}$ are feature vectors of lyrics lrc_i.

The training data for combination classifier is built on another training set $T' : \{lrc'_i, c'_i\}_{i=1}^{n}$ to prevent over-fitting. The generated training set for combination classifier is shown in equation 14.

$$T_{combiantion} = \{(h_{SMO}(lrc'_i), h_{NB}(lrc'_i), c'_i)\} \tag{14}$$

The generation of training set for combination classifier is done via k-fold cross validation. The whole training set is split into k folds. Each time, k-1 folds are used as training set T for basic learner and the remaining one is used as training set T' to build training data for combination classifier. Results of each fold are merged into the final training set for combination classifier.

C4.5 is chosen as the learning algorithm for the combination classifier as it is similar with the arbitration process of human.

6. EXPERIMENTS AND RESULTS

6.1 Experiment Settings

6.1.1 Data set

The data set we use is the same as that used by Hu [15]. It is made up of 500 Chinese pop songs,and the emotions of the songs are labeled through a subjective test conducted by 8 participants. The lyrics of the songs are downloaded from the web by a web crawler.

The distribution of the songs over the four emotion classes is shown in Table 1. Although the number of songs in class '+V-A' is small, it conforms to the distribution in reality.

Method	Baseline	POS tf*idf	Fuzzy Clustering
F-measure(av.)	0.3886	0.5942	0.547

Table 2. A comparison of word oriented methods

region	Zaanen tf*idf	POS tf*idf	rhyme	# of song
+V,+A	0.7074	**0.762**	0.438	274
+V,-A	0	0	0	5
-V,-A	0	0	**0.22**	52
-V,+A	0	**0.514**	0.353	169
av.	0.3886	**0.594**	0.382	500

Table 3. Results of single classifier

6.1.2 Machine learning algorithm

SMO, Naive Bayes, and J48 classification library in WEKA [7] are used to train classifiers.

6.1.3 Measurement

We choose f-measure as our metric.In each of the experiments, f-measure is computed using 5 fold cross-validation. For the tf*idf feature is computed on the training set, the tf*idf values are recomputed for each experiment.

6.2 POS tf*idf

The result of the POS tf*idf feature is shown in table 2. We choose Zaanen's method as our baseline. In contrast with the baseline, our method which incorporates POS gets a performance increase of 53%. The POS tf*idf model even outbalance Fuzzy Clustering method of Hu [15].

6.3 Combination Approach

In this part, we will describe the results of combination methods.

The results of single classifier are shown in table 3. Though the result using rhyme as feature is much smaller than that of POS tf*idf, it is similar with result of Zaanen's tf*idf. Rhyme frequency is an effective feature.

region	Features Level	Classifiers Level		# of song
	concatenation	base learning	meta learning	
+V,+A	0.728	0.581	**0.774**	274
+V,-A	0	0	0	5
-V,-A	0.09	**0.261**	0.049	52
-V,+A	0.489	0.451	**0.547**	169
av.	0.58	0.509	0.615	500

Table 4. Combination methods Analysis

The combination of the two get a better result, though there is a big difference between the two classifiers. F-measures increases in all regions indicating the effectiveness of the meta-learning algorithm. Rhyme classifier has poor performance on the whole, but it is better at dealing with instances in '-V,-A' region. And those misclassified by the rhyme classifier are corrected by the POS tf*idf classifier.

As mentioned in section 5, fusion on features level and classifiers level are tried. By comparing POS tf*idf column in table 3 and concatenation column in table 4, we find that fusion on features level fails to improve the result. For different features have different meanings, it's not appropriate to concatenate them simply.

For fusion on the classifiers level, we try both base-learning and meta-learning for classifier combination. We use weighted average method for base-learning and stack generalization for meta-learning. From table 4, we find that the meta-learning outperforms the base-learning by 0.1, which proves the effectiveness of meta-learning in the task of classifier combination. Besides, the base-learning even lowers the f-measure compared to single classifier based on POS tf*idf. Simple strategies could not guarantee the effectiveness of combination.

7. CONCLUSION AND FUTURE WORK

In this paper, we present three main contributions. Firstly, we get a great performance improvement in classification of music emotion by extending the work of Zaanen. Secondly, we propose to use rhyme cues in music emotion classification to complement traditional word based features. Finally,a meta-learning algorithm is used to combine classifiers based on different features.

There are more to be explored with lyrics. New features such as the tone changes and the mental images can be extracted from lyrics. Combining audio content,we can turn to the field of multi-modal music emotion classification.

8. ACKNOWLEDGMENT

The work is supported by Beijing Natural Science Foundation(Multimodal Chinese song emotion recognition).

9. REFERENCES

[1] Shen guo hui. The research into the rhyme,emotions and their association. *Journal of City Polytechnic ZhuHai*, 2009.

[2] Xiao Hu and J. Stephen Downie. WHEN LYRICS OUTPERFORM AUDIO FOR MUSIC MOOD CLASSIFICATION: A FEATURE ANALYSIS . In J. Stephen Downie and Remco C. Veltkamp, editors, *11th International Society for Music Information and Retrieval Conference*, August 2010.

[3] Xiao Hu, J. Stephen Downie, and Andreas F. Ehmann. LYRIC TEXT MINING IN MUSIC MOOD CLASSIFICATION. In J. Stephen Downie and Remco C. Veltkamp, editors, *10th International Society for Music Information and Retrieval Conference*, August 2009.

[4] Petri Juslin, Patrik N.;s Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5):770–814, Sep 2003.

[5] Youngmoo E. Kim, Erik M. Schmidt, Raymond Migneco, Brandon G. Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A. Speck, and Douglas Urnbull. Music Emotion Recognition: a State of the Art Review. In J. Stephen Downie and Remco C. Veltkamp, editors, *11th International Society for Music Information and Retrieval Conference*, August 2010.

[6] Qi Lu, Xiaoou Chen, Deshun Yang, and Jun Wang. BOOSTING FOR MULTI-MODAL MUSIC EMOTION . In J. Stephen Downie and Remco C. Veltkamp, editors, *11th International Society for Music Information and Retrieval Conference*, August 2010.

[7] Geoffrey Holmes Bernhard Pfahringer Peter Reutemann Ian H. Witten Mark Hall, Eibe Frank. The weka data mining software: An update. *SIGKDD Explorations*, 11, 2009.

[8] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 1980.

[9] Robert E. Thayer. *The biopsychology of mood and arousal*. Oxford University Press, September 1989.

[10] Sergey Tulyakov, Stefan Jaeger, Venu Govindaraju, and David S. Doermann. Review of classifier combination methods. In *Machine Learning in Document Analysis and Recognition*, pages 361–386. 2008.

[11] Menno van Zaanen and Pieter Kanters. AUTOMATIC MOOD CLASSIFICATION USING TF*IDF BASED ON LYRICS. In J. Stephen Downie and Remco C. Veltkamp, editors, *11th International Society for Music Information and Retrieval Conference*, August 2010.

[12] Ricardo Vilalta and Youssef Drissi. A perspective view and survey of meta-learning. *Artificial Intelligence Review*, 18:77–95, 2002.

[13] Jun Wang, Xiaoou Chen, Yajie Hu, and Tao Feng. Predicting High-level Music Semantics using Social Tags via Ontology-based Reasoning. In J. Stephen Downie and Remco C. Veltkamp, editors, *11th International Society for Music Information and Retrieval Conference*, August 2010.

[14] Yunqing Xia, Linlin Wang, Kam-Fai Wong, and Mingxing Xu. Sentiment vector space model for lyric-based song sentiment classification. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, HLT-Short '08, pages 133–136, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[15] Xiaoou Chen Yajie Hu and Deshun Yang. LYRIC-BASED SONG EMOTION DETECTION WITH AFFECTIVE LEXICON AND FUZZY CLUSTERING METHOD. In J. Stephen Downie and Remco C. Veltkamp, editors, *10th International Society for Music Information and Retrieval Conference*, August 2009.