Sentiment Vector Space Model for Lyric-based Song Sentiment Classification

Yunqing Xia

Center for Speech and language Tech. RIIT, Tsinghua University Beijing 100084, China yqxia@tsinghua.edu.cn

Kam-Fai Wong

Dept. of SE&EM The Chinese University of Hong Kong Shatin, Hong Kong kfwong@se.cuhk.edu.hk

Abstract

Lyric-based song sentiment classification seeks to assign songs appropriate sentiment labels such as *light-hearted* and *heavy-hearted*. Four problems render vector space model (VSM)-based text classification approach ineffective: 1) Many words within song lyrics actually contribute little to sentiment; 2) Nouns and verbs used to express sentiment are ambiguous; 3) Negations and modifiers around the sentiment keywords make particular contributions to sentiment; 4) Song lyric is usually very short. To address these problems, the sentiment vector space model (s-VSM) is proposed to represent song lyric document. The preliminary experiments prove that the s-VSM model outperforms the VSM model in the lyric-based song sentiment classification task.

1 Introduction

Song sentiment classification nowadays becomes a hot research topic due largely to the increasing demand of ubiquitous song access, especially via mobile phone. In their music phone *W910i*, Sony and Ericsson provide *Sense Me* component to catch owner's mood and play songs accordingly. Song sentiment classification is the key technology for song recommendation. Many research works have been reported to achieve this goal using audio sig-

Linlin Wang

State Key Lab of Intelligent Tech. and Sys. Dept. of CST, Tsinghua University Beijing 100084, China wangll07@mails.tsinghua.edu.cn

> Mingxing Xu Dept. of CST Tsinghua University Beijing 100084, China xumx@tsinghua.edu.cn

nal (Knees et al., 2007). But research efforts on lyric-based song classification are very few.

Preliminary experiments show that VSM-based text classification method (Joachims, 2002) is ineffective in song sentiment classification (see Section 5) due to the following four reasons. Firstly, the VSM model considers all content words within song lyric as features in text classification. But in fact many words in song lyric actually make little contribution to sentiment expressing. Using all content words as features, the VSM-based classification methods perform poorly in song sentiment classification. Secondly, observation on lyrics of thousands of Chinese pop songs reveals that sentiment-related nouns and verbs usually carry multiple senses. Unfortunately, the ambiguity is not appropriately handled in the VSM model. Thirdly, negations and modifiers are constantly found around the sentiment words in song lyric to inverse, to strengthen or to weaken the sentiments that the sentences carry. But the VSM model is not capable of reflecting these functions. Lastly, song lyric is usually very short, namely 50 words on average in length, rendering serious sparse data problem in VSM-based classification.

To address the aforementioned problems of the VSM model, the sentiment vector space model (s-VSM) is proposed in this work. We adopt the s-VSM model to extract sentiment features from song lyrics and implement the *SVM-light* (Joachims, 2002) classification algorithm to assign sentiment labels to given songs.

2 Related Works

Song sentiment classification has been investigated since 1990s in audio signal processing community and research works are mostly found relying on audio signal to make a decision using machine learning algorithms (Li and Ogihara, 2006; Lu et al., 2006). Typically, the sentiment classes are defined based on the Thayer's arousal-valence emotion plane (Thayer, 1989). Instead of assigning songs one of the four typical sentiment labels, Lu et al. (2006) propose the hierarchical framework to perform song sentiment classification with two steps. In the first step the energy level is detected with intensity features and the stress level is determined in the second step with timbre and rhythm features. It is proved difficult to detect stress level using audio as classification proof.

Song sentiment classification using lyric as proof is recently investigated by Chen et al. (2006). They adopt the hierarchical framework and make use of song lyric to detect stress level in the second step. In fact, many literatures have been produced to address the sentiment analysis problem in natural language processing research. Three approaches are dominating, i.e. knowledge-based approach (Kim and Hovy, 2004), information retrieval-based approach (Turney and Littman, 2003) and machine learning approach (Pang et al., 2002), in which the last approach is found very popular. Pang et al. (2002) adopt the VSM model to represent product reviews and apply text classification algorithms such as Naïve Bayes, maximum entropy and support vector machines to predict sentiment polarity of given product review.

Chen et al. (2006) also apply the VSM model in lyric-based song sentiment classification. However, our experiments show that song sentiment classification with the VSM model delivers disappointing quality (see Section 5). Error analysis reveals that the VSM model is problematic in representing song lyric. It is necessary to design a new lyric representation model for song sentiment classification.

3 Sentiment Vector Space Model

We propose the sentiment vector space model (s-VSM) for song sentiment classification. Principles of the s-VSM model are listed as follows.

(1) Only sentiment-related words are used to produce sentiment features for the s-VSM model.

- (2) The sentiment words are appropriately disambiguated with the neighboring negations and modifiers.
- (3) Negations and modifiers are included in the s-VSM model to reflect the functions of inversing, strengthening and weakening.

Sentiment unit is found the appropriate element complying with the above principles.

To be general, we first present the notation for sentiment lexicon as follows.

$$L = \{C, N, M\}; C = \{c_i\}, i = 1, ..., I$$
$$N = \{n_j\}, j = 1, ..., J$$
$$M = \{m_i\}, l = 1, ..., L$$

in which L represents sentiment lexicon, C sentiment word set, N negation set and M modifier set. These words can be automatically extracted from a semantic dictionary and each sentiment word is assigned a sentiment label, namely *light-hearted* or *heavy-hearted* according to its lexical definition.

Given a piece of song lyric, denoted as follows,

$$W = \{w_h\}, h = 1, ..., H$$

in which *W* denotes a set of words that appear in the song lyric, the semantic lexicon is in turn used to locate sentiment units denoted as follows.

$$U = \{u_{v}\} = \{c_{i,v}, n_{j,v}, m_{l,v}\}$$

$$c_{i,v} \in W \cap C; \ n_{i,v} \in W \cap N; \ m_{l,v} \in W \cap M$$

Note that sentiment units are unambiguous sentiment expressions, each of which contains one sentiment word and possibly one modifier and one negation. Negations and modifiers are helpful to determine the unique meaning of the sentiment words within certain context window, e.g. 3 preceding words and 3 succeeding words in our case. Then, the s-VSM model is presented as follows.

$$V_{S} = (f_{1}(U), f_{2}(U), ..., f_{T}(U))$$

in which V_s represents the sentiment vector for the given song lyric and $f_i(U)$ sentiment features which are usually certain statistics on sentiment units that appear in lyric.

We classify the sentiment units according to occurrence of sentiment words, negations and modifiers. If the sentiment word is mandatory for any sentiment unit, eight kinds of sentiment units are obtained. Let f_{PSW} denote count of positive sentiment words (PSW), f_{NSW} count of negative sentiment words (NSW), f_{NEG} count of negations (NEG) and f_{MOD} count of modifiers (MOD). Eight sentiment features are defined in Table 1.

f_i	Number of sentiment units satisfying
f_1	$f_{PSW} > 0, f_{NSW} = f_{NEG} = f_{MOD} = 0$
f_2	$f_{PSW} = 0, f_{NSW} > 0, f_{NEG} = f_{MOD} = 0$
f_3	$f_{PSW} > 0, f_{NSW} = 0, f_{NEG} > 0, f_{MOD} = 0$
f_4	$f_{PSW}=0, f_{NSW}>0, f_{NEG}>0, f_{MOD}=0$
f_5	$f_{PSW} > 0, f_{NSW} = 0, f_{NEG} = 0, f_{MOD} > 0$
f_6	$f_{PSW}=0, f_{NSW}>0, f_{NEG}=0, f_{MOD}>0$
f_7	$f_{PSW} > 0, f_{NSW} = 0, f_{NEG} > 0, f_{MOD} > 0$
f_8	$f_{PSW} = 0, f_{NSW} > 0, f_{NEG} > 0, f_{MOD} > 0$

Table 1. Definition of sentiment features. Note that one sentiment unit contains only one sentiment word. Thus it is not possible that f_{PSW} and f_{NSW} are both bigger than zero.

Obviously, sparse data problem can be well addressed using statistics on sentiment units rather than on individual words or sentiment units.

4 Lyric-based Song Sentiment Classification

Song sentiment classification based on lyric can be viewed as a text classification task thus can be handled by some standard classification algorithms. In this work, the *SVM-light* algorithm is implemented to accomplish this task due to its excellence in text classification.

Note that song sentiment classification differs from the traditional text classification in feature extraction. In our case, sentiment units are first detected and the sentiment features are then generated based on sentiment units. As the sentiment units carry unambiguous sentiments, it is deemed that the s-VSM is model is promising to carry out the song sentiment classification task effectively.

5 Evaluation

To evaluate the s-VSM model, a song corpus, i.e. 5SONGS, is created manually. It covers 2,653 Chinese pop songs, in which 1,632 are assigned label of *light-hearted* (positive class) and 1,021 assigned *heavy-hearted* (negative class). We randomly select 2,001 songs (around 75%) for training and the rest for testing. We adopt the standard evaluation criteria in text classification, namely precision (p), recall (r), f-1 measure (f) and accuracy (a) (Yang and Liu, 1999).

In our experiments, three approaches are implemented in song sentiment classification, i.e. audiobased (AB) approach, knowledge-based (KB) approach and machine learning (ML) approach, in which the latter two approaches are also referred to as text-based (TB) approach. The intentions are 1) to compare AB approach against the two TB approaches, 2) to compare the ML approach against the KB approach, and 3) to compare the VSMbased ML approach against the s-VSM-based one.

Audio-based (AB) Approach

We extract 10 timbre features and 2 rhythm features (Lu et al., 2006) from audio data of each song. Thus each song is represented by a 12-dimension vector. We run *SVM-light* algorithm to learn on the training samples and classify test ones.

Knowledge-based (KB) Approach

We make use of HowNet (Dong and dong, 2006), to detect sentiment words, to recognize the neighboring negations and modifiers, and finally to locate sentiment units within song lyric. Sentiment (SM) of the sentiment unit (SU) is determined considering sentiment words (SW), negation (NEG) and modifiers (MOD) using the following rule.

(1)
$$SM(SU) = label(SW);$$

(2) SM(SU) = -SM(SU) iff SU contains NEG;

(3) *SM*(SU) = *degree*(MOD)**SM*(SU) iff SU contains MOD.

In the above rule, label(x) is the function to read sentiment $label(\in \{1, -1\})$ of given word in the sentiment lexicon and degree(x) to read its modification degree($\in \{1/2, 2\}$). As the sentiment labels are integer numbers, the following formula is adopted to obtain label of the given song lyric.

$$label = sign\left(\sum_{i} SM(SU_{i})\right)$$

Machine Learning (ML) Approach

The ML approach adopts text classification algorithms to predict sentiment label of given song lyric. The *SVM-light* algorithm is implemented based on VSM model and s-VSM model, respectively. For the VSM model, we apply (CHI) algorithm (Yang and Pedersen, 1997) to select effective sentiment word features. For the s-VSM model, we adopt HowNet as the sentiment lexicon to create sentiment vectors.

Experimental results are presented Table 2.

	р	R	<i>f-1</i>	а
Audio-based	0.504	0.701	0.586	0.504
Knowledge-based	0.726	0.584	0.647	0.714
VSM-based	0.587	1.000	0.740	0.587
s-VSM-based	0.783	0.750	0.766	0.732

Table 2. Experimental results

Table 2 shows that the text-based methods outperform the audio-based method. This justifies our claim that lyric is better than audio in song sentiment detection. The second observation is that machine learning approach outperforms the knowledge-based approach. The third observation is that s-VSM-based method outperforms VSMbased method on f-1 score. Besides, we surprisingly find that VSM-based method assigns all test samples light-hearted label thus recall reaches 100%. This makes results of VSM-based method unreliable. We look into the model file created by the SVM-light algorithm and find that 1,868 of 2,001 VSM training vectors are selected as support vectors while 1,222 s-VSM support vectors are selected. This indicates that the VSM model indeed suffers the problems mentioned in Section 1 in lyric-based song sentiment classification. As a comparison, the s-VSM model produces more discriminative support vectors for the SVM classifier thus yields reliable predictions.

6 Conclusions and Future Works

The s-VSM model is presented in this paper as a document representation model to address the problems encountered in song sentiment classification. This model considers sentiment units in feature definition and produces more discriminative support vectors for song sentiment classification. Some conclusions can be drawn from the preliminary experiments on song sentiment classification. Firstly, text-based methods are more effective than the audio-based method. Secondly, the machine learning approach outperforms the knowledgebased approach. Thirdly, s-VSM model is more reliable and more accurate than the VSM model. We are thus encouraged to carry out more research to further refine the s-VSM model in sentiment classification. In the future, we will incorporate some linguistic rules to improve performance of sentiment unit detection. Meanwhile, sentiment features in the s-VSM model are currently equally

weighted. We will adopt some estimation techniques to assess their contributions for the s-VSM model. Finally, we will also explore how the s-VSM model improves quality of polarity classification in opinion mining.

Acknowledgement

Research work in this paper is partially supported by NSFC (No. 60703051) and Tsinghua University under the Basic Research Foundation (No. JC2007049).

References

- R.H. Chen, Z.L. Xu, Z.X. Zhang and F.Z. Luo. Content Based Music Emotion Analysis and Recognition. Proc. of 2006 International Workshop on Computer Music and Audio Technology, pp.68-75. 2006.
- Z. Dong and Q. Dong. *HowNet and the Computation of Meaning*. World Scientific Publishing. 2006.
- T. Joachims. Learning to Classify Text Using Support Vector Machines, Methods, Theory, and Algorithms. Kluwer (2002).
- S.-M. Kim and E. Hovy. *Determining the Sentiment of Opinions*. Proc. COLING'04, pp. 1367-1373. 2004.
- P. Knees, T. Pohle, M. Schedl and G. Widmer. A Music Search Engine Built upon Audio-based and Webbased Similarity Measures. Proc. of SIGIR'07, pp.47-454. 2007
- T. Li and M. Ogihara. *Content-based music similarity* search and emotion detection. Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing, pp. 17–21. 2006.
- L. Lu, D. Liu and H. Zhang. Automatic mood detection and tracking of music audio signals. IEEE Transactions on Audio, Speech & Language Processing 14(1): 5-18 (2006).
- B. Pang, L. Lee and S. Vaithyanathan. *Thumbs up? Sentiment Classification using Machine Learning Techniques*. Proc. of EMNLP-02, pp.79-86. 2002.
- R. E. Thayer, *The Biopsychology of Mood and Arousal*, New York, Oxford University Press. 1989.
- P. D. Turney and M. L. Littman. *Measuring praise and criticism: Inference of semantic orientation from association*. ACM Trans. on Information Systems, 21(4):315–346. 2003.
- Y. Yang and X. Liu. A Re-Examination of Text Categorization Methods. Proc. of SIGIR'99, pp. 42-49. 1999.
- Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization. Proc. ICML'97, pp.412-420. 1997.