

Coarse-to-Fine Sentence-level Emotion Classification based on the Intra-sentence Features and Sentential Context

Jun Xu¹, Ruifeng Xu^{1*}, Qin Lu², Xiaolong Wang¹

¹Key Laboratory of Network Oriented Intelligent Computation, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
{xujun,xuruifeng}@hitsz.edu.cn, wangxl@insun.hit.edu.cn

²Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
csluqin@comp.polyu.edu.hk

ABSTRACT

This paper proposes a novel approach using a coarse-to-fine analysis strategy for sentence-level emotion classification which takes into consideration of similarities to sentences in training set as well as adjacent sentences in the context. First, we use intra-sentence based features to determine the emotion label set of a target sentence coarsely through the statistical information gained from the label sets of the k most similar sentences in the training data. Then, we use the emotion transfer probabilities between neighboring sentences to refine the emotion labels of the target sentences. Such iterative refinements terminate when the emotion classification converges. The proposed algorithm is evaluated on Ren-CECps, a Chinese blog emotion corpus. Experimental results show that the coarse-to-fine emotion classification algorithm improves the sentence-level emotion classification by 19.11% on the average precision metric, which outperforms the baseline methods.

Categories and Subject Descriptors

I.2.7 [Computing Methodologies]: Artificial Intelligence—natural language processing, text analysis

Keywords

Emotion classification, Multi-label classification, Machine learning

1. INTRODUCTION

Emotion is a strong human feeling such as love, hate, or anger. It plays an essential role in decision making, percep-

tion, learning etc., and it influences the very mechanisms of rational thinking [5]. In the past few years, with the popularity of Web 2.0, text emotion analysis has attracted growing attentions. Emotion analysis investigates how to classify a text fragment into a set of emotion labels such as anger, joy etc. To measure and recognize the emotional changes of population in large scale is one of the most important areas in social sciences and economics studies [2, 1].

Text emotion classification focuses on the writer or speech holder's emotional state conveyed through the text. Previous emotion classification researches mainly focused on blog postings [4, 8, 3] from author's perspective. Evaluation for most of them use the LiveJournal¹ blog corpus which has 132 emotion categories and the blog text is classified according to the mood reported by its author during composition. Mishne [4] used the top forty frequent moods as classes and implemented a binary SVM classifiers for mood classification. Instead of using standard machine learning approach, Keshkar and Inkpen [3] proposed a hierarchical approach which uses the hierarchy of possible moods and achieved a better result. Wu et al. [8] represented the emotional state of each sentence as a sequence of semantic labels and attributes. They first used the Apriori algorithm to derive the emotion association rules (EARs) for each emotion, and then developed a separable mixture model to estimate the similarity between the sentence and the EARs of each emotion.

In previous studies, text emotion classification (TEC) is always treated as a single-label classification problem. However, a word or phrase may express more than one emotion. Examples include the idiom “喜忧参半(mixed hope and fear)”, “痛并快乐着(take the pleasure with pain)”, etc.. Statistics on Ren-CECps [6], a Chinese blog emotion annotated corpus, shows that about 15.1% Chinese emotion words are multi-emotion words which express complex feelings in its usage [7]. Furthermore, most previous works classified sentence emotion using intra-sentence features without considering their neighboring sentences. It is common sense that the emotions of neighboring sentences provide relevant information.

Base on the above two observations, a novel approach using a coarse-to-fine analysis strategy is proposed for sentence-

*Corresponding author.

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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$15.00.

¹www.livejournal.com

level emotion classification which incorporates features from training data using both Multi-Label k Nearest Neighbor (ML- k NN) approach [9] as well as contextual neighbors. ML- k NN is first used to obtain coarse multi-emotions of a target sentence. Then, emotion transfer probabilities of its neighboring sentences are incorporated to refine the emotion analysis of a target sentence. The emotions of the neighboring sentences are also updated and used for refinement iteratively. Such a coarse-to-fine analysis iterations terminate until the analysis results up to specified iteration number. The evaluation on the Ren-CECps shows that the proposed approach effectively improves the sentence-level TEC.

The rest of this paper is organized as follows. Section 2 describes the ML- k NN based coarse-to-fine approach for TEC. Section 3 presents the performance evaluation. Section 4 is the conclusion.

2. ALGORITHM DESIGN

In our proposed algorithm, the emotion label set of a sentence s is determined by features from two sources. Firstly, the emotion classification of s considers the emotions of its most k similar sentences in the training corpus. These sentences, referred to as the nearest neighbors in the k NN based algorithm, are compared to through intra-sentence features such as word unigram and character bigram. Secondly, the emotion classification of s considers the emotion of its contextual sentences since the author's emotion normally has its inertia. This is the motivation for using the inter-sentence features in emotion classification. In this work, only the immediate left and immediate right sentences of s are used as contextual information. In other words, the context window size is $[-1, +1]$.

To make use of both the intra-sentence features from the training data and inter-sentence features from contextual sentences, a coarse-to-fine emotion classification strategy is used. ML- k NN is first applied to find out the most similar sentences, from the training data [6]. This step coarsely determines the emotion class of s . Then the emotion class of s is fine tuned by incorporating the emotion transfer probabilities between its neighboring sentences iteratively.

2.1 Coarse Emotion Classification using the ML- k NN Algorithm

Multi Label k NN (ML- k NN) is used to coarsely determine the emotion label set of each sentence. ML- k NN is a variant of the single label k Nearest Neighbor lazy algorithm. For each test sentence s , ML- k NN first identifies its k nearest neighbors $\mathbb{N}(s)$ in the training set. Let Y be the finite set of emotion labels. For an emotion label $\epsilon \in Y$, let H_ϵ^1 denote the event that s has the emotion label ϵ , H_ϵ^0 denote the event that s does not have the emotion label ϵ , and $E_\epsilon^j (j \in \{0, 1, \dots, k\})$ denote the event that, among $\mathbb{N}(s)$, there are exactly j sentences which has label ϵ . Following the maximum a posteriori principle, the category vector $\vec{y}_s(\epsilon)$ can be determined as:

$$\vec{y}_s(\epsilon) = \arg \max_{b \in \{0,1\}} p(H_\epsilon^b | E_\epsilon^{\vec{C}_s(\epsilon)}) \quad (1)$$

Here, $\vec{C}_s(\epsilon)$ is a membership counting vector and $\vec{C}_s(\epsilon) = \sum_{a \in \mathbb{N}(s)} \vec{y}_a(\epsilon)$. It counts the number of neighbors of s belonging to the ϵ -th emotion. By Bayes's theorem, Eq. (1)

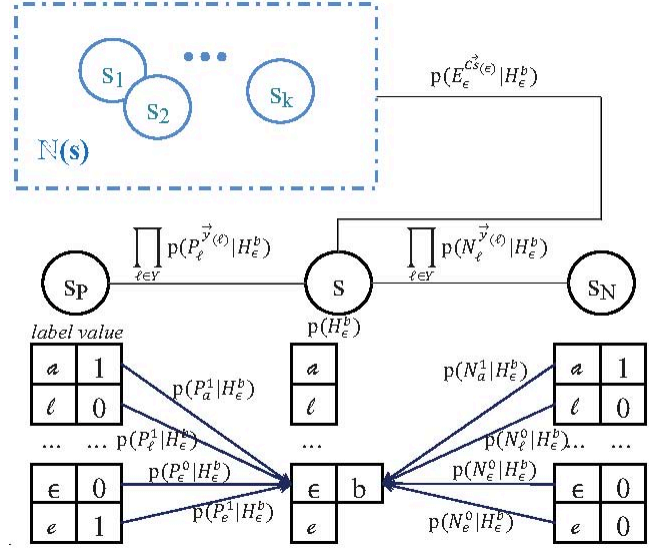


Figure 1: An illustration of incorporating the emotion transfer probabilities between neighboring sentences, which accounts for all emotion transfer cases.

can be rewritten as:

$$\begin{aligned} \vec{y}_s(\epsilon) &= \arg \max_{b \in \{0,1\}} \frac{p(H_\epsilon^b) p(E_\epsilon^{\vec{C}_s(\epsilon)} | H_\epsilon^b)}{p(E_\epsilon^{\vec{C}_s(\epsilon)})} \\ &= \arg \max_{b \in \{0,1\}} p(H_\epsilon^b) p(E_\epsilon^{\vec{C}_s(\epsilon)} | H_\epsilon^b) \end{aligned} \quad (2)$$

The prior probabilities $p(H_\epsilon^b)$ and probabilities $p(E_\epsilon^j | H_\epsilon^b) (j \in \{0, 1, \dots, k\})$ can be estimated from the training set based on frequency counting.

2.2 Coarse-to-Fine Emotion Classification using the Contextual Neighbors

Based on the coarse emotion classification, the emotion transfers between neighboring sentences are used to refine emotion classification. Let s_P denotes the previous sentence of s , and s_N be the next sentence following s , and s_P, s, s_N must all be in the same paragraph. Let P_ℓ^1 be the event that s_P has the emotion label ℓ , and P_ℓ^0 be the event that s_P does not have the label ℓ . Similarly, N_ℓ^1 is the event that s_N has the label ℓ , while N_ℓ^0 is the event that s_N does not have the label ℓ . If s is the first sentence of a paragraph, event P_ℓ^1 and P_ℓ^0 will not happen, while the last sentence of a paragraph s is, event N_ℓ^1 and N_ℓ^0 will not happen.

Assuming that the emotion transfer be a sequence of events and each transfer event is independent, as the joint event $E_\epsilon^{\vec{C}_s(\epsilon)}$ for all emotional transfer, $\vec{y}_s(\epsilon)$ can be determined as:

$$\begin{aligned} \vec{y}_s(\epsilon) &= \arg \max_{b \in \{0,1\}} p(H_\epsilon^b | E_\epsilon^{\vec{C}_s(\epsilon)}, P_\ell^{\vec{y}(\ell)}, \dots, N_\ell^{\vec{y}(\ell)} \dots) \\ &= \arg \max_{b \in \{0,1\}} p(H_\epsilon^b) p(E_\epsilon^{\vec{C}_s(\epsilon)} | H_\epsilon^b) \\ &\quad \times \prod_{\ell \in Y} p(P_\ell^{\vec{y}(\ell)} | H_\epsilon^b) p(N_\ell^{\vec{y}(\ell)} | H_\epsilon^b) \end{aligned} \quad (3)$$

Eq. (3) is also referred to as the AET (All Emotion Transfer) as it accounts for all emotion transfer cases between neigh-

boring sentences. Figure.1 illustrates the different emotion transfers that determines the label of s .

If only the same emotion transfer is considered, $\vec{y}_s(\epsilon)$ is then determined by:

$$\vec{y}_s(\epsilon) = \arg \max_{b \in \{0,1\}} p(H_\epsilon^b) p(E_\epsilon^{\vec{C}_s(\epsilon)} | H_\epsilon^b) p(P_\epsilon^{\vec{y}_s(\epsilon)} | H_\epsilon^b) p(N_\epsilon^{\vec{y}_s(\epsilon)} | H_\epsilon^b) \quad (4)$$

Eq. (4) is also called SET (Same Emotion Transfer) as it counts for the same emotion transfer. In Eq. (3), four kinds of emotion transfer probabilities need to be estimated. They are $p(\ell \rightarrow \epsilon)$, $p(\bar{\ell} \rightarrow \epsilon)$, $p(\ell \rightarrow \bar{\epsilon})$ and $p(\bar{\ell} \rightarrow \bar{\epsilon})$, respectively. $p(\ell \rightarrow \epsilon)$ can be computed by:

$$p(\ell \rightarrow \epsilon) = \frac{\text{counts}[\vec{y}_{s_p}(\ell) = 1, \vec{y}_s(\epsilon) = 1]}{\text{counts}[\vec{y}_s(\epsilon) = 1]} \quad (5)$$

The other three can be estimated similarly. The conditional probabilities $p(P_\ell^i | H_\epsilon^b)$ and $p(N_\ell^j | H_\epsilon^b)$ in Eq. (3) and Eq. (4) can be obtained from the above emotion transfer probabilities. For example, $p(P_\ell^1 | H_\epsilon^1) = p(\ell \rightarrow \epsilon)$ and $p(N_\ell^0 | H_\epsilon^0) = p(\bar{\epsilon} \rightarrow \bar{\ell})$ etc.

Algorithm 1 Coarse-to-fine emotion classification algorithm.

Input: T , test corpus; I , iteration number.

```

1: for each sentence  $s \in T$  do
2:   Identify the  $k$  nearest neighbors  $\mathbb{N}(s)$ ;
3:   for each label  $\epsilon \in Y$  do
4:     Count  $\vec{C}_s(\epsilon) = \sum_{a \in \mathbb{N}(s)} \vec{y}_a(\epsilon)$ ;
5:     Compute  $\vec{y}_s(\epsilon, 0)$  using Eq.2 ;
6:   end for
7: end for
8: for each  $t \in [1, I]$  do
9:   for each sentence  $s \in T$  do
10:    for each label  $\epsilon \in Y$  do
11:      Iteratively refine  $\vec{y}_s(\epsilon, t)$  using Eq.3 (AET) or Eq.4 (SET);
12:    end for
13:   end for
14: end for

```

The pseudo code of the proposed Coarse-to-Fine Emotion Classification algorithm (C2FEC) is given above. ML- k NN is applied to classify the emotions of all sentences in the testing corpus (line 1-7). The emotions obtained for all the sentences by ML- k NN (result at line 7) are refined by using a complex classifier which incorporates the emotion transfer probabilities in iterations (line 8-14). The iteration terminates when the iteration number reaches the specified iteration number I .

3. EVALUATION AND DISCUSSION

3.1 Experimental Setup

To evaluate the sentence-level emotion classification approaches, the Ren-CECps data [6] is used. Ren-CECps is an annotated Chinese blog emotion corpus with either eight emotional categories labeled at the sentence level (emotional labels: *expect*, *joy*, *love*, *surprise*, *anxiety*, *sorrow*, *angry* and *hate*) or *neutral* if the sentence is factual. The corpus contains 1,487 documents, 11,255 paragraphs, 35,096 sentences, and 878,164 Chinese words.

To evaluate the proposed C2FEC algorithm, it is also compared to three other algorithms, namely the Binary Relevance method using Naive Bayes (BR_NB), the Binary Relevance method using linear SVM (BR_SVM) and ML- k NN. The BR approach is used because it is a different kind of approach in classification and it can perform as well or even superior than ML- k NN in some aspects. Intra-sentence features used are unigram of Chinese words after filtering stop words, and term frequency weighting for document representation. The cosine similarity measure is used to compute the similarity of two sentences.

The commonly used metrics for multi-label classification are different from those used in single-label classification due to the inherent differences of the classification problem. In this study, five popular evaluation metrics are adopted in the multi-label classification experiment include hamming loss, one-error, coverage, ranking loss, and average precision.

In the experiments, the Ren-CECps corpus is randomly split into 3 equal parts based on documents. 3 fold cross validation is performed. All the performance given below are the average of 3 folds.

3.2 Performance Evaluation and Analysis

The performance of the emotion classification by different approaches are given in Table 1. The result shown in the table used $k=8$ by ML- k NN, since $k=8$ gives the best performance. Items (4) and (5) are the results achieved by AET, while the iteration number is set to 1 and 100, respectively. Items (6) and (7) are the results obtained by SET, which only considers the same emotion transfer (the iteration number is set to 1 and 100). The contribution of emotion transfer in emotion classification can be observed first by comparing ML- k NN with C2FEC with only one iteration ($I=1$) as shown in Item (4) and (6) compared to Item (3). Results show that with only one iteration, C2FEC can improve the one-error by 11.5% and 12.1%, respectively. As shown in Item (7), the best performance is achieved by SET with 100 iterations. In fact, when iteration reaches 100, compared to ML- k NN, the one-error of SET decreases by 24.02% and average precision increases by 19.11%, the most significant in the two proposed algorithms. All things considered, the proposed algorithm outperforms all the baseline methods due to the use of sentential context information.

As shown in Table 1, BR_SVM outperforms BR_NB on most evaluation metrics except coverage and ranking loss, and slightly outperforms ML- k NN on one-error metrics. It is shown that ML- k NN, a multi-label algorithm adaptation method, is not far superior to problem transfer method BR_SVM. However, when sentential context is introduced, the proposed algorithms are obviously better because ML- k NN has the ability to incorporate the contextual features.

To estimate the performance of both AET and SET with different iterations, comparison is made between the two for iteration from 1 to 100 as shown in Fig. 2 on average precision and one-error, respectively. There are a number of observations. Firstly, both algorithms exhibit similar trend and converge well. Secondly, the iterations from 1 to 10 give the largest improvement and the rest of the improvement is monotonic. It is also important to know that SET always performs better than AET. One reason is that the transfer emotions between neighboring sentences with the highest probabilities are always the same. Another reason is that the assumption of AET that each emotion transfer event

Table 1: Experimental Results of each Multi-label Learning Algorithm(mean±std).

	Method	Hamming Loss	One-error	Coverage	Ranking Loss	Average Precision
(1)	BR_NB	0.2612±0.0138	0.6743±0.0170	2.9562±0.2749	0.2946±0.0131	0.5215±0.0095
(2)	BR_SVM	0.1703±0.0104	0.5976±0.0107	3.2823±0.4475	0.3231±0.0313	0.5434±0.0129
(3)	ML- <i>k</i> NN	0.1590±0.0157	0.6107±0.0245	2.7708±0.4382	0.2655±0.0326	0.5563±0.0254
(4)	C2FEC-AET(<i>I</i> =1)	0.1550±0.0125	0.5404±0.0226	2.4800±0.3383	0.2335±0.0217	0.6081±0.0179
(5)	C2FEC-AET(<i>I</i> =100)	0.1512±0.0095	0.4715±0.0233	2.1949±0.2404	0.2021±0.0117	0.6588±0.0139
(6)	C2FEC-SET(<i>I</i> =1)	0.1507±0.0141	0.5366±0.0226	2.4762±0.3638	0.2293±0.0245	0.6100±0.0191
(7)	C2FEC-SET(<i>I</i> =100)	0.1426±0.0126	0.4640±0.0233	2.1874±0.2910	0.1939±0.0173	0.6626±0.0155

is independent and identically distributed, may not be as suitable as that of SET.

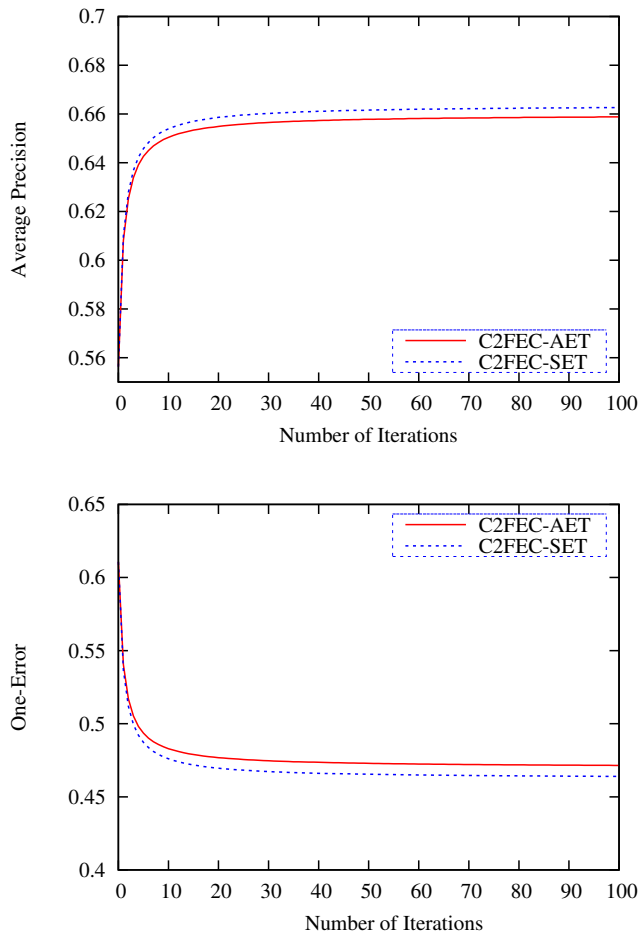


Figure 2: Iteration curves by average precision and one-error, respectively.

4. CONCLUSIONS

In this paper, a novel approach which uses a coarse-to-fine analysis strategy for sentence-level emotion analysis is proposed by incorporating the features from the traditional similar measures and contextual sentences. The emotion label set of the target sentence is determined by ML-*k*NN coarsely and then refined with the emotion transfer probabilities following the coarse-to-fine analysis strategy iteratively.

The ability of the ML-*k*NN to incorporate contextual information makes it possible to obtain further performance gains. The evaluation on a large Chinese blog emotion corpus shows that the proposed algorithms improve sentence-level emotion analysis by significant 24.02% on one-error metric, and 19.11% on average precision, outperform most of the existing methods.

5. ACKNOWLEDGMENTS

This investigation is supported in part by the China Postdoctoral Science Foundation (No. 2011M500670), National Natural Science Foundation of China (No.61203378 and No.60973076), Hong Kong Polytechnic University(ZOEP) and HIT.NSFIR.2010124.

6. REFERENCES

- [1] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science.*, 2(1):1–8, 2011.
- [2] P. S. Dodds and C. M. Danforth. Measuring the happiness of large scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456, 2009.
- [3] F. Keshtkar and D. Inkpen. A hierarchical approach to mood classification in blogs. *Natural Language Engineering*, 18(1):61–81, 2011.
- [4] G. Mishne. Experiments with mood classification in blog posts. In *Proceedings of the 1st Workshop on Stylistic Analysis of Text For Information Access*, 2005.
- [5] R. W. Picard. *Affective Computing*. MIT Press, 1997.
- [6] C. Quan and F. Ren. Construction of a blog emotion corpus for chinese emotional expression analysis. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1446–1454, 2009.
- [7] C. Quan and F. Ren. An exploration of features for recognizing word emotion. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 922–930, 2010.
- [8] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin. Emotion recognition from text using semantic labels and separable mixture models. *ACM Transactions on Asian Language Information Processing*, 5(2):165–183, 2006.
- [9] M.-L. Zhang and Z.-H. Zhou. ML-*k*NN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7):2038–2048, 2007.