Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources

Ge Xu, Xinfan Meng, Houfeng Wang
Key Laboratory of Computational Linguistics (Peking University), Ministry of Education
Institute of Computational Linguistics, Peking University
{xuge, mxf, wanghf}@pku.edu.cn

Abstract

For sentiment analysis, lexicons play an important role in many related tasks. In this paper, aiming to build Chinese emotion lexicons for public use, we adopted a graph-based algorithm which ranks words according to a few seed emotion words. The ranking algorithm exploits the similarity between words, and uses multiple similarity metrics which can be derived from dictionaries, unlabeled corpora or heuristic rules. To evaluate the adopted algorithm and resources, two independent judges were asked to label the top words of ranking list.

It is observed that noise is almost unavoidable due to imprecise similarity metrics between words. So, to guarantee the quality of emotion lexicons, we use an iterative feedback to combine manual labeling and the automatic ranking algorithm above. We also compared our newly constructed Chinese emotion lexicons (happiness, anger, sadness, fear and surprise) with existing counterparts, and related analysis is offered.

1 Introduction

Emotion lexicons have a great impact on the results of related tasks. With high-quality emotion lexicons, systems using simple methods can achieve competitive performance. However, to manually build an emotion lexicon is time-consuming. Many research works in building lexicons use automatic methods to assist the building procedure. Such works commonly rank words by the similarities to a set of seed words, then those words with high ranking scores are more likely to be added to the final lexicons or used as additional seed words.

For Chinese, emotion lexicons are scarce resources. We can get a small set of emotion words from semantic dictionary (such as CCD, HowNet, synonym dictionaries) or directly from related papers (Xu and Tao, 2003) (Chen et al. , 2009), but it is often not sufficient for practical systems. Xu et al. (2008) constructed a large-scale emotion ontology dictionary, but it is not publicly available yet.

In this paper, we adopted a graph-based algorithm to automatically rank words according to a few seed words. Similarity between words can be utilized and multiple resources are used to boost performance. Combining manual labeling with automatic ranking through an iterative feedback framework, we can produce high-quality emotion lexicons. Our experiments focused on Chinese, but the method is applicable to any other language as long as suitable resources exist.

The remainder of this paper is organized as follows. In Section 2, related works are introduced. In Section 3, we describe a graph-based algorithm and how to incorporate multiple resources. Section 4 gives the details of applying the algorithm on five emotions and shows how to evaluate the results. Section 5 focuses on how to build and evaluate emotion lexicons, linguistic consideration and instruction for identifying emotions are also included. Finally, conclusion is made in Section 6.

Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 1209–1217,
Beijing, August 2010

2 Related work

Riloff and Shepherd (1997) presented a corpus-based method that can be used to build semantic lexicons for specific categories. The input to the system is a small set of seed words for a category and a representative text corpus. The output is a ranked list of words that are associated with the category. An approach proposed by (Turney, 2002) for the construction of polarity started with a few positive and negative seeds, then used a similarity method (pointwise mutual information) to grow this seed list from web corpus. Our experiments are similar with these works, but we use a different ranking method and incorporate multiple resources. To perform rating inference on reviews, Goldberg and Zhu (2006) created a graph on both labeled and unlabeled reviews, and then solved an optimization problem to obtain a smooth rating function over the whole graph. Rao and Ravichandran (2009) used three semi-supervised methods in polarity lexicon induction based on WordNet, and compared them with corpus-based methods. Encouraging results show methods using similarity between words can improve the performance. Wan and Xiao (2009) presented a method to use two types of similarity between sentences for document summarization, namely similarity within a document and similarity between documents. The ranking method in our paper is similar to the ones used in above three papers, which fully exploit the relationship between any pair of sample points (both labeled and unlabeled). When only limited labeled data are available, such method achieves significantly better predictive accuracy over other methods that ignore the unlabeled examples during training.

Xu et al. (2008) at first formed a taxonomy for emotions, under which an affective lexicon ontology exploiting various resources was constructed. The framework of ontology is filled by the combination of manual classification and automatic methods. To our best knowledge, this affective lexicon ontology is the largest Chinese emotion-oriented dictionary.

3 Our method

3.1 A graph-based algorithm

For our experiments, we chose the graph-based algorithm in (Zhou et al. , 2004) which is transductive learning and formulated as follows:

Given a point set $\chi = \{x_1, ..., x_l, x_{l+1}, ..., x_n\}$, the first l points $x_i (i \leq l)$ are labeled and the remaining points $x_u (l+1 \leq u \leq n)$ unlabeled. The goal is to rank the unlabeled points.

Let F denotes an n-dimensional vector whose elements correspond to ranking scores on the data set χ. Define another n-dimensional vector Y with $Y_i = 1$ if x_i is labeled and $Y_i = 0$ otherwise. Y denotes the initial label assignment.

The iterative algorithm is shown in the following:

Algorithm 1 A graph-based algorithm

1. Construct the weight matrix W and set W_{ii} to zero to avoid self-reinforcement. W is domain-dependent.
2. Construct the similarity matrix $S = D^{1/2} W D^{1/2}$ using symmetric normalization. D is a diagonal matrix with $D_{ii} = \Sigma_j W_{ij}$.
3. Iterate $F(t+1) = \alpha S F(t) + (1-\alpha) Y$ until convergence, where α is a parameter in (0, 1), and $F(0) = Y$. We clamp labeled points to 1 after each iteration.
4. Let F^* denote $F(t)$ when the iteration converges.

In our experiments, labeled points are seed emotion words, S_{ij} denotes the similarity between *i*th word and *j*th word. In an iteration, each word absorbs label information from other words. More similar two words are, more influence they have on each other. The label information (initially from seed emotion words) will propagate along S. The final output F^* contains ranking scores for all words, and a score indicates how similar the corresponding word is to the seed emotion words.

The implementation of the iterative algorithm is theoretically simple, which only involves basic matrix operation. Compared with methods which do not exploit the relationship between samples, experiments showing advantages of graph-based learning methods can be found

in (Rao and Ravichandran, 2009),(Goldberg and Zhu, 2006),(Tong et al. , 2005),(Wan and Xiao, 2009),(Zhu and Ghahramani, 2002) etc. When labeled data are scarce, such graph-based transductive learning methods are especially useful.

3.2 Incorporate multiple resources

For building the emotion lexicons, we are faced with lots of resources, such as semantic dictionaries, labeled or unlabeled corpora, and some linguistic experiences which can be presented as heuristic rules. Naturally we want to use these resources together, thus boosting the final performance. In graph-base setting, such resources can be used to construct the emotion-oriented similarity between words, and similarities will be represented by matrices.

The schemes to fuse similarity matrices are presented in (Sindhwani et al. , 2005), (Zhou and Burges, 2007), (Wan and Xiao, 2009) and (Tong et al. , 2005) etc. In our paper, not aiming at comparing different fusion schemes, we used a linear fusion scheme to fuse different similarities matrices from different resources. The scheme is actually a convex combination of matrices, with weights specified empirically.

The fusion of different similarity matrices falls in the domain of multi-view learning. A well-known multi-view learning method is Co-Training, which uses two views (two resources) to train two interactive classifiers (Blum and Mitchell, 1998). Since we focus on building emotion lexicons using multiple resources (multiple views), those who want to see the advantages of multi-view learning over learning with one view can refer to (Blum and Mitchell, 1998), (Sindhwani et al. , 2005), (Zhou and Burges, 2007), (Wan and Xiao, 2009) and (Tong et al. , 2005) etc.

4 Experiments

We use the method in section 3 to rank for each emotion with a few seed emotion words. Once we implement the ranking algorithm 1, the main work resides in constructing similarity matrices, which are highly domain-dependent.

4.1 Construct similarity matrices

Here, we introduce how to construct four similarity matrices used in building emotion lexicons. Three of them are based on cooccurrence of words; the fourth matrix is from a heuristic rule.

We use ictclas3.0[1] to perform word segmentation and POS tagging.

In our experiments, the number of words involved in ranking is 93506[2], so theoretically, the matrices are 93506×93506. If the similarity between any pair of words is considered, the computation becomes impractical in both time and space cost. So we require that each word has at most 500 nearest neighbors.

Four matrices are constructed as follows:

4.1.1 Similarity based on a unlabeled corpus

The unlabeled corpus used is People's Daily[3](人民日报1997∼2004). After word segmentation and POS tagging, we chose three POS's (i,a,l)[4]. The nouns were not included to limit the scale of word space. We set the cooccurrence window to a sentence, and removed the duplicate occurrences of words. Any pair of words in a sentence will contribute a unit weight to the edge which connects the pair of words.

4.1.2 Similarity based on a synonym dictionary

We used the Chinese synonym dictionary (哈工大同义词词林扩展版[5]) for this matrix. In this dictionary, the words in a synonym set are presented in one line and separated by spaces, so there is no need to perform word segmentation and POS tagging. Any pair of words in one line will contribute a unit weight to the edge which connects the pair of words.

4.1.3 Similarity based on a semantic dictionary

We used The Contemporary Chinese Dictionary (现代汉语词典) to construct the third simi-

[1] downloaded from http://www.ictclas.org/

[2] Words are selected after word segmentation and POS tagging, see section 4.1.1∼4.1.3 for selection of words in details.

[3] http://icl.pku.edu.cn/

[4] i=Chinese idiom, a=adjective, l=Chinese phrase

[5] http://ir.hit.edu.cn/

larity matrix. Since word segmentation may segment the entries of the dictionary, we extracted all the entries in the dictionary and store them in a file whose words ictclas3.0 was required not to segment. Furthermore, for an entry in the dictionary, the example sentences or phrases appearing in its gloss may contain many irrelevant words in terms of emotions, so they were removed from the gloss.

After word segmentation and POS tagging[6], we set the cooccurrence window to one line (an entry and its gloss without example sentences or phrases), and removed the duplicate occurrences of words. An entry and any word in the modified gloss will contribute a unit weight to the edge which connects the pair of words. This constructing was a bit different, since we did not consider the similarity between words in modified gloss.

4.1.4 similarity based on a heuristic rule

In Chinese, a word is composed of one or several Chinese characters. A Chinese character is normally by itself an independent semantic unit, so the similarity between two words can be inferred from the character(s) that they share. For example, the Chinese word 欣 (happy) appears in the word 欣然 (readily). Since 欣然 and 欣 share one Chinese character, they are regarded as similar. Naturally, the larger the proportion that two words share, the more similar they are. In this way, the fourth weighted matrix was formed. To avoid incurring noises, we exclude the cases where one Chinese character is shared, with the exception that the Chinese character itself is one of the two Chinese words.

4.1.5 Fusion of four similarity matrices

After processing all the lines (or sentences), the weighted matrices are normalized as in algorithm 1, then four similarity matrices are linearly fused with equal weights (1/4 for each matrix).

4.2 Select seed emotion words

In our experiments, we chose emotions of *happiness, sadness, anger, fear and surprise* which are widely accepted as basic emotions[7]. Empirically, we assigned each emotion with seed words given in Table 1.

Emotion	**Seed words**
喜(happiness)	高兴, 愉快, 欢乐, 喜悦, 兴高采烈, 欢畅, 开心
怒(anger)	愤怒, 不满, 恼火, 生气, 愤恨, 恼怒, 愤懑, 震怒, 悲愤, 窝火, 痛恨, 恨之入骨, 义愤填膺, 怒气冲天
哀(sadness)	悲伤, 沮丧, 痛苦, 伤心, 难过, 悲哀, 难受, 消沉, 灰心丧气, 悲戚, 闷闷不乐, 哀伤, 悲愤, 悲切, 悲痛欲绝, 欲哭无泪
惧(fear)	恐惧, 惧怕, 担心, 提心吊胆, 害怕, 惊恐, 疑惧, 畏惧, 不寒而栗, 望而生畏
惊(surprise)	惊讶, 大吃一惊 ,震惊, 惊恐, 惊异, 惊骇, 惊, 出乎意料, 惊喜, 惊叹

Table 1: Seed emotion words

4.3 Evaluation of our method

We obtained five ranking lists of words using the method in section 3. Following the work of (Riloff and Shepherd, 1997), we adopted the following evaluation setting.

To evaluate the quality of emotion ranking lists, each list was manually rated by two persons independently. For each emotion, we selected the top 200 words of each ranking list and presented them to judges. We presented the words in random order so that the judges had no idea how our system had ranked the words. The judges were asked to rate each word on a scale from 1 to 5 indicating how strongly it was associated with an emotion, 0 indicating no association. We allowed the judges to assign -1 to a word if they did not know what it meant. For the words rated as -1, we manually assigned ratings that we thought were appropriate.

The results of judges are shown in figures 1-5. In these figures, horizontal axes are the number of reviewed words in ranking lists and vertical axes are number of emotion words found (with 5 different strength). The curve labeled as $> x$ means that it counts the number of words which are rated

[6]since we do not segment entries in this dictionary, all POS's are possible

[7]Guidelines for identifying emotions is in section 5, before that, we understand emotions through common sense.

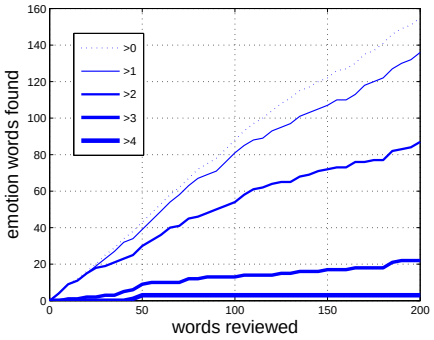

Figure 1: happiness

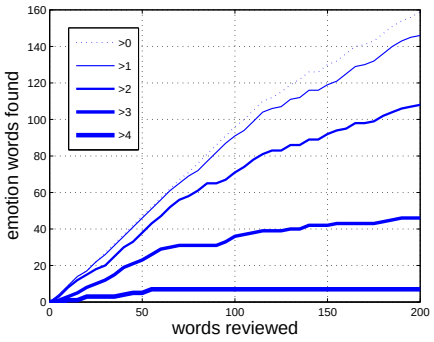

Figure 2: anger

greater than x by either judge.

Curves ($> 0, > 1, > 2$) display positive slopes even at the end of the 200 words, which implies that more emotion words would occur if more than 200 words are reviewed. By comparison, curves ($> 3, > 4$) tend to be flat when they are close to the right side, which means the cost of identifying high-quality emotion words will increase greatly as one checks along the ranking list in descendent order.

It is observed that words which both judges assign 5 are few. In *surprise* emotion, the number is even 0. Such results may reflect that emotion is harder to identify compared with topical categories in (Riloff and Shepherd, 1997).

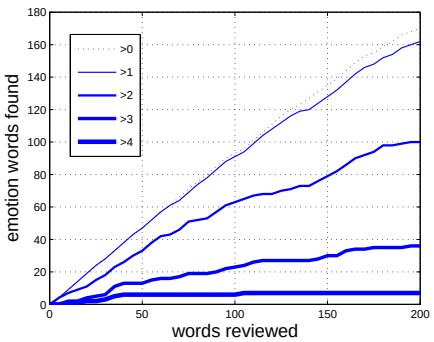

Figure 3: sadness

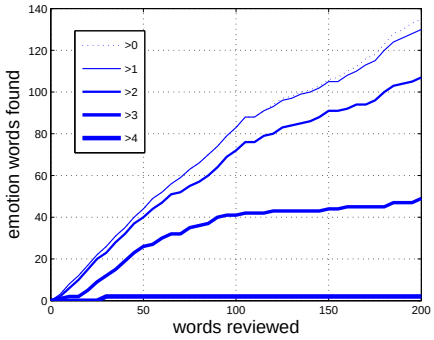

Figure 4: fear

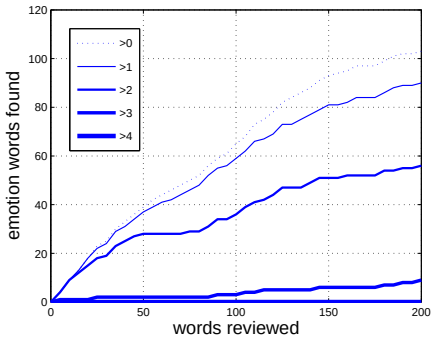

Figure 5: surprise

From the semantic dictionary, our method found many low-frequency emotion words such as 忭 (pleasant, glad), 蘧然 (surprise and happy), 忉怛 (sad), or those used in Chinese dialects such as 毛咕 (fear), 挂气 (angry). Such emotion words are necessary for comprehensive emotion lexicons.

Because more POS's than adjectives and verbs are included in our experiments, some emotion words such as the noun 冷门 (unexpected winner),and the adverb 竟然 (to one's surprise) are also spotted, which to some extent implies the generality of our method.

5 Construct emotion lexicons

The above section introduced a method to rank words with a few seed emotion words. However, to build emotion lexicons requires that we manually remove the noises incurred by the automatic ranking method. Accordingly, guidelines for identifying emotions are needed, and also some linguistic consideration in identifying emoting words should be given.

5.1 An iterative feedback to denoise

In our experiments, we observed that noises incurred by similarity matrices are almost unavoidable. For example, in the unlabeled corpus, 国事访问 (state visits) always co-occurred with 高兴 (happy) or 愉快 (happy), so in happiness emotion, 国事访问 acquired a high ranking position (174th); in terms of the heuristic rule, 意料 (expected) shares two Chinese characters with 出乎意料 (unexpected, surprised), however they have opposite meaning because 出乎 (exceed, beyond) is a negative word. 意料 unfavorably ranked high (88th) in surprise emotion; from the semantic dictionary, the gloss of 年画 (Chinese Spring Festival pictures) contains 欢乐 (happy), thus in happiness emotion, 年画 ranked high (158th).

So after each ranking of an emotion, in the descendent order of ranking scores, we manually revised some scores in about top 500. Several criteria (see 5.2 and 5.3) were given to guide if a word has a specified emotion. For those words surely bearing the specified emotion, we assigned 1 to them ,and left others unchanged. Seeing the words newly revised to be 1 as new seed emotion words, we run the ranking algorithm again. After such feedback was repeated 2∼3 times, we collected all the words labeled with 1 to form the final emotion lexicons. In (Zhou et al. , 2004), the author also suggested such *iterative feedback* to extend the query (seed) set and improve the ranking output. Commonly, the size of an emotion lexicon is small, so we do not have to check too many words.

The human revising procedure is sensitive to annotators' background. To improve the quality of the emotion lexicons, experts with linguistic or psychology background will help.

Furthermore, the ranking algorithm used in our paper is clearly sensitive to the initial seed words, but since we adopt an iterative feedback framework, the words not appearing in the initial set of seed words will show up in next iteration with high ranking scores. We also performed experiments which selected emotion seed words based on the Chinese synonym dictionary and the emotion words in (Chen et al. , 2009), similar results were found.

5.2 Guidelines for identifying emotions

The same as (Chen et al. , 2009), we used the definition that emotion is the felt awareness of bodily reactions to something perceived or thought. Also, we were highly influenced by the structure of the affective lexicon presented by (Ortony et al. , 1987), and used the Affective states and Affective-Behavioral conditions in the structure to identify emotion words in our paper[8].

With such guidelines, 胆小 (cowardice, relates more to external evaluation) is not an emotional word of fear. We also intentionally distinguish between emotions and expression of emotions. For example, 大笑 (laugh), 哈哈 (haw-haw) are seen as expression of happiness and 颤抖 (tremble) as of fear, but not as emotion words. In addition, we try to distinguish between an emotion and the cause of an emotion, see 5.3 for an example.

For each emotion, brief description is given as below[9]:

1. **Happiness**：the emotional reaction to something that is satisfying.
2. **Anger**：do not satisfy the current situation and have a desire to fight or change the situation. Often there exists a target for this emotion.
3. **Sadness**：an emotion characterized by feelings of disadvantage, loss, and helplessness. Sadness often leads to cry.
4. **Fear**：the emotional response to a perceived threat. Fear almost always relates to future events, such as worsening of a situation, or continuation of a situation that is unacceptable.
5. **Surprise**：the emotional reaction to something unexpected.

5.3 Linguistic consideration for identifying emotion words

If a word has multiple senses, we only consider its emotional one(s). For example, 生气 (as a verb, it means *be angry*, but means *vitality or spirits* as a noun) will appear in the emotion lexicon of anger.

[8] According to (Ortony et al. , 1987), *surprise* should not be seen as a basic emotion for it relates more to cognition. However, our paper focuses on the building of emotion lexicons, not the disputable issue of basic emotions

[9] we mainly referred to http://en.wikipedia.org/wiki

If one sense of a word is the combination of emotions, the word will appear in all related emotions.

We mainly consider four POS's, namely nouns, verbs, adjectives and adverb[10]. If a word has multiple POS's, we normally consider its POS with strongest emotion (Empirically, we think the emotion strength ranks in decedent order as following: adjectives, verbs, adverbs, nouns.). So we consider the verb of 恐惧 (fear) when it can be used as a noun and a verb in Chinese. The 生气 example above also applies here.

For each of four POS's, instruction for emotion identification is given as below:

Nouns: For example, 怒火 (rage, anger), 喜气 (joy or jubilation), 冷门 (an unexpected winner) are selected as emotion words. We distinguish between an emotion and the cause of an emotion. For example, calamity often leads to sadness, but does not directly contain the emotion of sadness. 冷门 appears in the surprise lexicon because we believe it contains surprise by itself.

Adverbs: The adverbs selected into emotion lexicons contain the emotions by themselves. For example, 竟然 (unexpectedly), 欣欣然 (cheerily), 气哼哼 (angrily), 蓦地 (unexpectedly), 伤心地 (sadly) etc.

Verbs: As in (Ortony et al. , 1987), Chinese emotion verbs also fall into at least two distinct classes, causatives and noncausatives. Both classes are included in our emotion lexicons. For example, 动肝火 (be angry), 担心 (fear) are noncausative verbs, while 激怒 (enrage), 震惊 (to make someone surprised) are causative ones. Probably due to the abundant usage of 令人/让人/使人 (to make someone) etc., causative emotion verbs are few compared to noncausative ones in Chinese.

Adjective：Quite a lot of emotion words fall in this POS, since adjectives are the natural expression of internal states of humans. For example, 高兴 (happy), 惊讶 (surprised), 愤怒 (angry) etc.

For any word that it is hard to identify at first sight, we used a search tool[11] to retrieve sentences which contain the word, and then identify if the word is emotional or not by its usage in the sentences.

[10]For Chinese idioms, we only considered those used as these four POS's, omitted those used as a statement, such as 哀兵必胜 (an army burning with righteous indignation is bound to win)

[11]provided by Center for Chinese Linguistics of Peking University, http://ccl.pku.edu.cn

5.4 Comparison with existing Chinese emotion resources

诧、骇、惊、讶、矍、蘧、愕、遽、骇然、赫然、竟然、居然、蘧然、愕然、愕然、矍然、爆冷、爆冷门、不料、不意、不虞、诧异、吃惊、出乎意料、出乎意外、出乎预料、出冷门、出其不意、出人意料、出人意外、触目惊心、错愕、大吃一惊、大惊失色、大惊小怪、怪讶、骇怪、骇然、骇人听闻、骇异、好家伙、赫然、赫然而怒、黑马、惊诧、惊呆、惊服、惊骇、惊慌、惊慌失措、惊惶、惊惶失措、惊魂未定、惊悸、惊惧、惊恐、惊恐万状、惊奇、惊人、惊世骇俗、惊叹、惊悉、惊喜、惊喜交集、惊喜万分、惊吓、惊羡、惊讶、惊疑、惊异、惊厥、惊愕、竟然、竟是、竟至、竟自、居然、冷不丁、冷不防、冷孤丁、冷门、没成想、猛不防、猛孤丁地、纳罕、始料不及、始料未及、受宠若惊、受惊、谁料、谁知、突如其来、未料、闻所未闻、想不到、心惊、心惊胆颤、心惊胆战、讶异、一语惊人、意料之外、意外、意想不到、又惊又喜、震惊、蓦地

Table 2: The emotion lexicon of surprise

Under the guidelines for manually identifying emotion words, we finally constructed five Chinese emotion lexicons using the iterative feedback. The newly constructed emotion lexicons were also reported as resources together with our paper. The emotion lexicon of *surprise* is shown in Table 2. In this part, we compare our lexicons with the following counterparts, see Table 3.

Ours1 in the table is the final emotion lexicons, and Ours2 is the abridged version that excludes the words of single Chinese character and Chinese idioms.

Chinese Concept Dictionary (CCD) is a WordNet-like semantic lexicon(Liu et al. , 2003).

	喜	怒	哀	惧	惊
CCD nouns	22	27	38	46	10
(Xu and Tao, 2003)	45	12	28	21	12
(Chen et al. , 2009)	28	34	28	17	11
(Xu et al. , 2008)	609	187	362	182	47
Ours1	95	118	97	106	99
Ours2	52	77	72	57	65

Table 3: Compare various emotion lexicons

We only considered the noun network which is richly developed in CCD, as in other semantic dictionaries. For each emotion, we chose its synset as well as the synsets of its hypernym and hyponym(s). In fact, most of words in the emotion nouns extracted can be used as verbs or adjectives in Chinese. However, since CCD is not designed for emotion analysis, words which are expression of emotions such as 哭泣 (cry) or evaluation such as 胆小 (cowardice) were included.

Selecting nouns and verbs, Xu and Tao (2003) offered an emotion taxonomy of 390 emotion words. The taxonomy contains 24 classes of emotions and excludes Chinese idioms. By our inspection to the offered emotion words in this taxonomy, the authors tried to exclude expression of emotions, evaluation and cause of emotions from emotions, which is similar with our processing[12]. Ours2 is intentionally created to compare with this emotion taxonomy.

Based on (Xu and Tao, 2003), Chen et al. (2009) removed the words of single Chinese character; let two persons to judge if a word is an emotional one and only those agreed by the two persons were seen as emotion words. It is worth noting that Chen et al. (2009) merges 怒 (anger) and 烦 (fidget) in (Xu and Tao, 2003) to form the 怒 (anger) lexicon, thus 讨厌 (dislike) appears in anger lexicon. However, we believe 讨厌 (dislike) is different with 怒 (anger), and should be put into another emotion. Also, we distinguish between 恨 (hate) and 怒 (anger).

Xu et al. (2008) constructed a large-scale affective lexicon ontology. Given the example words in their paper, we found that the authors did not intentionally exclude the expression of emotions such as 面红耳赤 (literally, red face and ear), 笑眯眯 (literally, be smiling). Such criteria of identifying emotion words may partially account for the large size of their emotion resources.

[12]Xu and Tao (2003) included words such as 情愿/愿意 (be willing to), 留神 (be careful) in their happiness lexicon, which we think should not be classified into happiness.

6 Conclusion and future work

In this paper, aiming to build Chinese emotion lexicons, we adopt a graph-based algorithm and incorporate multiple resources to improve the quality of lexicons and save human labor. This is an initial attempt to build Chinese emotion lexicons, the quality of constructed emotion lexicons is far from perfect and is supposed to be improved step by step.

The method in this paper can be further extended to subjectivity/polarity classification and other non-sentimental tasks such as word similarity computing, and can be also adapted to other languages. The more resources we use, the more human cost can be saved and the higher the quality of built emotion lexicons is.

In the future work, we want to construct other emotion lexicons such as 好 (like, love), 恶 (dislike), 欲 (desire) etc. using the same method.

Acknowledgement This research is supported by National Natural Science Foundation of China (No.60973053, No.90920011)

References

A. Blum and T. Mitchell. 1998. Combining labeled and unlabeled data with co-training。 *In Proceedings of the 11th Annual Conference on Computational Learning Theory*, 92-100.

Ying Chen, Sophia Y. M. Lee, and Churen Huang. 2009. A Cognitive-based Annotation System for Emotion Computing. *Proceedings of the Third Linguistic Annotation Workshop (LAW III)*.

Andrew B. Goldberg, Xiaojin Zhu. 2006. Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. *Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing on the First Workshop on Graph Based Methods for Natural Language Processing*.

Y. Liu and et al. 2003. The CCD Construction Model and Its Auxiliary Tool VACOL. *Applied Linguistics, 45(1):83-88*.

A. Ortony, G. L. Clore, and M. A. Foss. 1987. The referential structure of the affective lexicon. *Cognitive Science, 11, 341-364*.

Delip Rao and D. Ravichandran. 2009. Semisupervised polarity lexicon induction. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 675-682.

Ellen Riloff and Jessica Shepherd. 1997. A Corpus-Based Approach for Building Semantic Lexicons. *In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing, pages 117-124*.

V. Sindhwani, P. Niyogi, and M. Belkin. 2005. A co-regularization approach to semisupervised learning with multiple views. *Proc. ICML Workshop on Learning with Multiple views*.

H. Tong, J. He, M. Li, C. Zhang, and W. Ma. 2005. Graph based multi-modality learning. *In Proceedings of the 13th Annual ACM international Conference on Multimedia. MULTIMEDIA '05. ACM*, New York, NY, 862-871.

Peter D. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *ACL 2002*, 417-424.

Xiaojun Wan and Jianguo Xiao. 2009. Graph-Based Multi-Modality Learning for Topic-Focused MultiDocument Summarization. *IJCAI 2009*, 1586-1591.

Linhong Xu, Hongfei Lin, Yu Pan, Hui Ren and Jianmei Chen. 2008. Constructing the Afective Lexicon Ontology. *JOURNAL OF THE CHINA SOCIETY F0R SCIENTIFIC AND TECHNICAL INFORMATION Vo1.27 No.2, 180-185*.

X. Y. Xu, and J. H. Tao. 2003. The study of affective categorization in Chinese. *The 1st Chinese Conference on Affective Computing and Intelligent Interaction. Beijing, China*.

Hongbo Xu, Tianfang Yao, and Xuanjing Huang. 2009. The second Chinese Opinion Analysis Evaluation(in Chinese). *COAE 2009*.

D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Scholkopf. 2004. Learning with local and global consistency. *Advances in Neural Information Processing Systems 16. MIT Press, Cambridge, MA.* .

D. Zhou and C. J. C. Burges. 2007. Spectral clustering and transductive learning with multiple views. *Proceedings of the 24th international conference on Machine learning*.

X. Zhu and Z. Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation. *Technical Report CMUCALD02107*. CMU.