Automatic Recognition of Emotion based on a Cognitively Motivated Emotion Annotation System

Ying Chen,1,* Sophia Yat Mei Lee,2 and Chu-Ren Huang3

1 College of Information and Electrical Engineering, China Agricultural University, China
  chenying3176@gmail.com
2 Language Centre, Hong Kong Baptist University, Hong Kong
  sophiaym@gmail.com
3 Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong
  churenhuang@gmail.com

Emotion recognition is very important for the extraction of expressive information. In this paper, we provide a robust and versatile emotion annotation scheme that can not only annotate explicit and implicit expressions of emotion, but also can encode different levels of information for a given emotion content. In addition, taking cognitive psychologists’ theories into account, large and comparatively high-quality emotion corpora are automatically created to allow for emotion recognition in Chinese and English. Our annotation scheme can easily be adapted for different kinds of applications dealing with emotion, and, being generic, can also be applied to other languages. We also discuss the two kinds of emotion representations used in our corpus, namely, holistic and componental representations. We find that the two representations have their own qualities and shortcomings and that choosing between them ultimately depends mainly on the application type.

Key words: emotion, annotation scheme, emotion recognition

* Corresponding author. Address: College of Information and Electrical Engineering, Room 731, China Agricultural University, Beijing, 100083, China. Tel.: +86 1062738792

©2011 Institute for Cognitive Science, Seoul National University
1. Introduction

Since emotions represent one of the most fundamental reactions of human experience, their recognition and identification is one of our most important cognitive abilities. It is probably no exaggeration to claim that most human activities are motivated by, or designed to excite, some emotion. And indeed, most events do activate such feelings, whether they are intended to or not. Given the critical roles emotions play in human activities, it is not surprising that sentiment analysis, as coarse-grained accounts of emotional tendencies (positive, negative, and neutral) has become one of the most popular topics in natural language processing (NLP) and Information extraction (IE). What is surprising though is the fact that there are few studies concerning emotion recognition in text, offering finer-grained information than most parsed texts do, which in addition are universally applicable regardless of the domain and the product types. In this article, we discuss a basic, yet important question concerning emotion analysis, the question of emotions should be classified and represented.

Since the concept of emotion is very complex and subjective compared to, say, POS annotation or word segmentation, emotion annotation is a real challenge and highly labor intensive, requiring careful human judgment. On the one hand, both explicit and implicit emotions must be recognized and tagged. On the other hand, automatic emotion recognition requires large and high-quality annotated data, something we still lack. This is not only because of the enormous labor required, but also because of the lack of a satisfying emotion annotation scheme, allowing for the robust and versatile annotation and recognition of emotion. In this paper, we provide an annotation scheme that is grounded in cognitive psychology. It combines psychological theories and semantic insights. Furthermore, concerning emotion recognition, our annotation scheme represents emotions in two ways: holistically (as a single lemma or toke) and componentially. For example, “pride” (the holistic representation) is decomposed into “happiness + fear” (the componential representation.) The grain-size representation of our emotion provides a means to view an emotion from different angles, revealing its different facets.
With such an emotion annotation scheme, a large and comparatively high-quality annotated emotion corpus has been built for emotion recognition via unsupervised learning. As Tokuhisa et al. (2008) have correctly pointed out, besides emotional-sentence corpora, neutral-sentence corpora (i.e. sentences containing no emotion) are needed for emotion recognition. This is why we have built (automatically) a high-quality neutral-sentence corpus based on contextual information. The two corpora are combined to form what we call an emotion corpus, as one of its primary tasks is to facilitate the recognition of emotions. While unsupervised learning cannot yet guarantee a perfectly, well-annotated corpus, it can be adapted for different kinds of emotion recognition.

Although a lot of work has been done on emotion recognition, many issues remain unsolved, such as the relationships between different emotion types, emotion-type selection, and so on. Hence, we still have a long way to go to solve the problem of emotion recognition, as there is more involved than simply relying on tagged emotion corpora.

In what follows we explore emotion recognition on the basis of a taxonomy of emotions, attempting to solve, be it only partially, the above mentioned issues. We look at emotions in our corpus from either of two perspectives (holistic or componential representations/occurrences). We also explore some critical issues concerning emotion recognition, such as whether human emotions should considered as enumerable, atomic emotions, or whether they should be decomposed into a fixed set of primary emotions. We develop different classification schema to process emotions, and our experiments show that each model is able to detect certain facets of an emotion.

However, in order to do this automatically, our representations need to be changed; that is, a different classification technology is required. In particular, we must be able to detect a vector rather than just a single label. In this paper, we choose multi-label classification (an instance can have several labels) to handle this vector detection task. We also discuss the trade-off between single-label classification (e.g. the detection of a single token, say, “pride”) and multi-label classification (e.g. the detection of its components “happiness” and “fear.”)

The remainder of this paper is organized as follows. In Section 2, we give
an overview concerning related work on emotion annotation, emotion recognition, and relevant theories from psychology and linguistics. In Section 3, we describe our emotion annotation scheme and its impact with respect to emotion recognition. Section 4 discusses how the corpus is created via unsupervised learning. Section 5 presents the pilot experiments for emotion recognition based on our corpus. Finally, in Section 6 we present our perspectives and conclusions.

2. Related Work

Although both sentiment and emotion belong to affective analysis, emotion analysis in NLP, ranging from the corpus construction to the definition of emotion recognition, is still in its early stages in comparison to sentiment task. Emotion representation is the fundamental issue to deal with as it involves many important tasks, such as data selection, the creation of annotation schemes, and emotion recognition modeling. In the following section, we first briefly introduce some theories of emotions related to our emotion annotation scheme, and we then compare two representations of emotions and their impact on the selection of emotion recognition models.

2.1 The analysis of the emotion problem

There is no clear consensus among psychologists and linguists on the concept of emotion. In this paper, we adopt the classic definition of emotions (Cannon, 1927): emotion is the felt awareness of bodily reactions to something perceived or thought.

Emotion is a complicated concept, and there are complicated relationships among different emotions. For example, the relationship between “discouraged” and “sad” is different from the one between “remorse” and “sad.” Since most theorists study emotions from the biological and psychological perspectives, they cannot easily apply them to NLP (Osherenko, 2011; Plutchik, 1980). In this article, we explore the following two analyses in cognitive emotion models in terms of emotion recognition: componential analysis (Osgood and Suci, 1957; Plutchik, 1980) and emotion cause analysis (Wierzbicka, 1996).

Regarding theories for componential analysis, Osgood and Suci (1957)
examine emotion representations using several dimensions (good-bad, active-passive, strong-weak, optimistic-pessimistic, beautiful-ugly, etc.). Their experiment shows that most of the variances in text affect judgment could be explained by three major factors: the evaluative factor (e.g. good-bad), the potency factor (e.g. strong-weak), and the activity factor (e.g. active-passive). Out of these three factors, they find that the most discriminative factor is the evaluative one. Plutchik (1980) proposes an emotion taxonomy, which divides emotion into two categories: primary emotions and complex emotions (i.e. the combinations of primary emotions).

With respect to theories of emotion causes analysis, Natural Semantic Metalanguage (NSM), one of the prominent cognitive models exploring human emotions, offers a comprehensive and practical approaches (Wierzbicka, 1996.) NSM describes complex and abstract concepts, such as emotions, in terms of simpler and concrete ones. In such a way, emotions are decomposed as complex events involving a cause and a mental state, which can be further described with a set of universal, irreducible cores called semantic primitives. This approach identifies the exact differences and connections between emotion concepts in terms of the causes, which provide an immediate cue for emotion recognition. We believe that the NSM model offers a plausible framework to be implemented for automatic emotion recognition.

2.2 Emotion representations for emotion recognition
For NLP, most studies of English emotions (Mishne, 2005; Mihalcea and Liu, 2006) use a blog corpus collected from LiveJournal. In LiveJournal, authors are allowed to describe their moods with some words for each post, and those description words are either selected from the predefined list of 132 common moods or else entered freely by the authors themselves. Mishne (2005) found that 54,487 unique mood words appear in 624,905 blog posts, and 46,558 (85.4%) mood words appear once only. The large size of mood words indicates that 1) it is impossible to collect all possible mood description words; 2) the problem of data sparsity cannot be avoided for some uncommon emotions; 3) a mood is so subjective that there are various ways to describe it. The work of Mishne (2005) also indicates the importance of emotion representation because an appropriate representation
can sometimes partially solve the above problems.

Furthermore, although the 40 most common emotion labels are examined in Mishne (2005), do clear boundaries exist to differentiate those emotion labels? Some emotion theories argue that emotions evolve like colors, and it is hard to discern emotions. As there are overlaps between those focused emotions, single-label classification (i.e. each instance contains one and only one label), a common technology for classification in NLP, faces inherent conflict. Single-label classification assumes that the predefined labels are mutually exclusive and each instance belongs to one label only. However, this assumption is often invalid in emotion recognition.

Alternatively, some emotion theories suggest that an emotion can be represented in a componential way (Osgood and Suci, 1957; Plutchik, 1980). Each emotion is expressed by a vector with fixed dimensions. For componential emotion representation, the task of the emotion recognition turns into an assignment of weights in the given dimensions, instead of single label detection. In this article, we simplify the issue of vector value assignment in which each dimension has only a binary value (0 or 1). We then explore different multi-label classification schema to solve this vector value assignment problem.

3. The Emotion Corpus

In this section, we first discuss the text type involved for emotion recognition in our study. We then explore the relationships among different emotion types based on the proposed emotion taxonomy, which combines psychological theories and linguistic features. Based on the emotion taxonomy, a robust and versatile emotion annotation scheme is designed to extract emotion sentences from a large corpus. Finally, we create a Chinese emotion corpus for facilitating emotion recognition.

3.1 The text for emotion recognition
In this article, we concentrate on the emotion analysis of formal written text. Compared to emotions in spoken data (intonation is a key indicator for emotion analysis) and informal text (e.g. blogs and online chat), emotions in formal text, such as news, are more likely to be expressed by emotion
keywords. Therefore, it seems that emotion recognition is satisfactory if the collection of emotion vocabulary is comprehensive and an emotion taxonomy is given. This can be done by detecting the given emotion keywords. For instance, a sentence containing the word “joyful” indicates the presence of “happiness” emotion. However, this intuitive approach is found to have certain limitations:

1) It is impossible to collect all emotion keywords, not only because of the large size of emotion expressions (Mishne, 2005) but also because of its dynamic nature. For example, “雷(lei)”, which originally means “thunder”, is now widely used to express the “shock” emotion in online chat rooms;
2) Emotion keywords often have multiple senses which inevitably lead to the problem of word ambiguity;
3) Emotion context has the problem of context shift, which is similar to sentiment shift (Polanyi and Zaenen, 2004.)

With these constraints, emotion recognition requires an in-depth semantic understanding of texts, even for formal texts.

3.2 The emotion taxonomy
Most theories of emotion agree that emotion can be divided into primary and complex emotions. In this article, we adopt Turner’s taxonomy (Turner, 2000), which is the extension of Plutchik’s work. The two main points are emphasized in this approach.

First, each primary emotion is divided into three levels according to its intensity: high, moderate, and low. Besides “happiness,” “sadness,” “anger,” and “fear,” Turner also suggests that “disgust” and “surprise” can be primary emotions (Turner, 1996; Turner, 2007.) In Chinese, the character “惊” (“surprise”) is productive in forming emotion words, such as 惊喜 (surprise and happiness), and 惊吓 (surprise and fear), which is consistent with the explanation of “surprise” by Plutchik (1991). Therefore, in our annotation scheme, we consider “happiness,” “sadness,” “anger,” “fear,” and “surprise” as primary emotions.

Second, complex emotions can be divided into first-order emotions (con-
consisting of two primary emotions), second-order emotions (consisting of three primary emotions), and so on, according to the number of primary emotions that involves in the complex emotion. For example, “pride” (happiness + fear) is a first-order complex emotion, which contains a greater amount of “happiness” and a lesser amount of “fear.”

Tables 1 and 2 show some keywords in Turner’s taxonomy. Table 1 lists the most common English keywords and their corresponding primary emotions, where the symbol “//” is used to separate different levels of intensity. Table 2 lists the English keywords and their corresponding complex emotions, where the symbol “//” is to separate different types of emotions. Several emotion keywords, which express similar meanings, are grouped under an emotion type. For example, the emotion keywords “awe, reverence, veneration” are grouped under the emotion type “awe.” For a complex emotion, the order of primary emotions indicates the importance of those primary emotions for that complex emotion. Thus, “envy” is “fear + anger,” which contains a greater amount of “fear” and a lesser amount of “anger” whereas “awe” is “fear + happiness,” which contains a greater amount of “fear” and a lesser amount of “happiness.”

Such a cognitive taxonomy of emotions is versatile for different lan-

Table 1. Primary emotions and some corresponding keywords

<table>
<thead>
<tr>
<th>Primary Emotions</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td><strong>High</strong>: ecstatic, eager, joy, enthusiastic, happy//<strong>Moderate</strong>:</td>
</tr>
<tr>
<td></td>
<td>cheerful, satisfy, pleased, enjoy, interest//<strong>Low</strong>: sanguine,</td>
</tr>
<tr>
<td></td>
<td>serene, content, grateful</td>
</tr>
<tr>
<td>Fear</td>
<td><strong>High</strong>: horror, terror//<strong>Moderate</strong>: misgivings, self-conscious,</td>
</tr>
<tr>
<td></td>
<td>scare, panic, anxious//<strong>Low</strong>: bewildered, reluct, shy, puzzles, confuse</td>
</tr>
<tr>
<td>Anger</td>
<td><strong>High</strong>: dislike, disgust, outrage, furious, hate//<strong>Moderate</strong>:</td>
</tr>
<tr>
<td></td>
<td>contentious, offend, frustrate, hostile, angry//<strong>Low</strong>:</td>
</tr>
<tr>
<td></td>
<td>contemptuous, irritable, annoy, impatient</td>
</tr>
<tr>
<td>Sadness</td>
<td><strong>High</strong>: deject, despondent, sorrow, anguish, despair//<strong>Moderate</strong>:</td>
</tr>
<tr>
<td></td>
<td>gloomy, dismay, sad, unhappy, disappoint//<strong>Low</strong>: dispirit, downcast,</td>
</tr>
<tr>
<td></td>
<td>discourage</td>
</tr>
<tr>
<td>Surprise</td>
<td><strong>High</strong>: astonish//<strong>Moderate</strong>: startled, amaze, surprise</td>
</tr>
</tbody>
</table>
Automatic Recognition of Emotions

Table 2. First-order complex emotions and some corresponding keywords

<table>
<thead>
<tr>
<th>Combinations</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness + Fear</td>
<td>Wonder: wonder, wondering, hopeful//Pride: pride, boastful</td>
</tr>
<tr>
<td>Happiness + Anger</td>
<td>Vengeance: vengeance, vengeful//Calm: appeased, calmed, calm, soothed//Bemused: bemused</td>
</tr>
<tr>
<td>Happiness + Sadness</td>
<td>Yearning: nostalgia, yearning</td>
</tr>
<tr>
<td>Fear + Happiness</td>
<td>Awe: awe, reverence, veneration</td>
</tr>
<tr>
<td>Fear + Anger</td>
<td>Antagonism: antagonism, revulsed//Envy: envy</td>
</tr>
<tr>
<td>Fear + Sadness</td>
<td>Worried: dread, wariness, pensive, helpless, apprehension, worried</td>
</tr>
<tr>
<td>Anger + Happiness</td>
<td>Unfriendly: snubbing, mollified, rudeness, placated, apathetic, unsympathetic, unfriendly, unaffectionate//Sarcastic: sarcastic</td>
</tr>
<tr>
<td>Anger + Sadness</td>
<td>Depressed: bitter, depression//Intolerant: intolerant</td>
</tr>
<tr>
<td>Sadness + Happiness</td>
<td>Acceptance: acceptance, tolerant//Solace: moroseness, solace, melancholy</td>
</tr>
<tr>
<td>Sadness + Fear</td>
<td>Hopeless: forlorn, lonely, hopeless, miserable//Remorseful: remorseful, ashamed, humiliated</td>
</tr>
<tr>
<td>Sadness + Anger</td>
<td>Discontent: aggrieved, discontent, dissatisfied, unfulfilled//Boredom: boredom//Grief: grief, sullenness</td>
</tr>
<tr>
<td>Surprise + Happiness</td>
<td>Delight: delight</td>
</tr>
<tr>
<td>Surprise + Sadness</td>
<td>Embarrassed: embarrassed</td>
</tr>
</tbody>
</table>

guages although there is no one-to-one mapping. We explore the Chinese emotion taxonomy by first selecting emotion keywords from the feeling words listed in Xu and Tao (2003). We then map those emotion keywords to Turner’s taxonomy with adaptation in some cases. Lastly, some polysemous emotion keywords are removed to reduce ambiguity, and 226 Chinese emotion keywords remain.
3.3 The representation of emotions

One of the most difficult and important issues for emotion recognition is the selection of emotion representation, which directly decides the content of, as well as the related technologies for, emotion recognition. Our emotion annotation scheme provides two of the most common emotion representations, namely the holistic and componential representations. In this sub-section, we discuss the differences between these two representations and the advantages and disadvantages of each representation in relation to emotion recognition.

3.3.1 The holistic representation

The holistic representation enumerates an emotion with a unique name, such as “pride” and “jealousy.” From the cognitive perspective, how to define and discern an emotion has been a great challenge, as in the case of “envy vs. jealousy.” In terms of emotion recognition, several points should be taken into account in this respect. First, the holistic representation covers only partial emotions. Very often, the focusing emotions are selected or designed according to specific applications. Very fine-grained or a large size of emotions will lead to a data sparsity problem (Mishne, 2005.) Second, this kind of emotion analysis cannot handle emotions that are not in the keyword list as there is no way to represent it.

When adopting the holistic representation, it is undesirable to treat emotion recognition as a single-label classification problem, as most of other NLP tasks do. However, some issues need to be considered. First, it is possible for several emotions to occur in a sentence simultaneously, and the size of all those combinations may be very large, making it impossible for a classifier to train all of them. Second, it is difficult to capture the complicated relationships between different emotions. Most emotion theories admit that, with the exception of a few prototypical emotions, an emotion often involves several other emotions. This indicates that an emotion has often inherited relationships with other emotions, and a good emotion recognition model should have the capability to detect or learn this kind of relationship. Unfortunately, a single-label classification model does not have this capability because of its underlying assumptions (the labels are inde-
3.3.2 The componential representation

Instead of enumerating all possible emotions, some emotion theories suggest that an emotion should be represented through a vector with small-scale fixed dimensions. For example, in our emotion annotation scheme, an emotion is represented by five primary emotions (five dimensions) based on Turner’s emotion taxonomy; Osherenko (2011) discusses affect sensing through multimodal fusion using lexical and acoustic data. This componential representation is a rather loose way of describing an emotion in which some information may be lost. Moreover, it is often not possible to convert an emotion to a vector with fixed dimensions. For example, Kemper (1987) suggests that complex emotions result from various aspects of social interaction, which are rather culture-specific. For example, “guilt” (an emotion with the holistic representation), apart from being decomposed into “joy” and “fear,” may involve other cultural-related moods which are lost in Turner’s componential representation. Hence, one problem for the componential representation is the selection of dimensions and the manner in which an emotion is decomposed so as to capture as much information as possible for an emotion. Most previous studies choose some prototypical emotions as dimensions and other complement dimensions specifically designed for the applications (Quan & Ren, 2009). The number of those prototypical emotions varies from four to 12 in different emotion theories (Kemper, 1987.)

Comparing these two representation methods, we find that the problem arising from the holistic representation can be avoided in the componential representation. However, we should admit that the holistic representation is capable of containing more information of an emotion than the componential representation. Finally, as explained in Section 2, single-label classification is not compatible with the componential representation in processing emotions. Therefore, we choose multi-label classification.

3.4 Emotion annotation scheme

With the proposed emotion annotation scheme, a large and comparatively high-quality annotated emotion corpus, including an emotional-sentence
Ying Chen, Sophia Yat Mei Lee, and Chu-Ren Huang

corpus and a neutral-sentence corpus, can be built for emotion recognition through an unsupervised approach.

Similar to most corpora, our corpus creation is designed to satisfy the requirements of real emotion recognition. Emotions can be expressed with or without emotion vocabulary in the text. As we have explained, the intuitive method based on a mapping of emotion keywords to emotions does not perform well. Moreover, the detection of emotions in a context without emotion keywords is very challenging. To deal with these problems, we built the emotion-sentence corpus, which is motivated by the NSM theory.

According to the NSM theory, an emotion is provoked by a stimulus. This indicates that detecting the emotion stimulus is one possible way to detect emotions in text. In other words, the emotion-sentence corpus is a collection of emotion stimuli. Since emotion is subjective, the stimulus-based approach works only when its context is provided. For example, the stimulus “build a gym for this community” may cause different emotions, such as “surprise,” “happy” and so on, depending on its context. We also notice that the text containing an emotion keyword may contain emotion stimulus and its context. Thus, a natural corpus creation approach emerges.

In our annotation scheme, a pattern-based approach is used to collect the emotion-sentence corpus, which is similar to the one used in Tokuhisa et al. (2008). However, we do not limit it to event-driven emotions (Kozareva et al., 2007); rather we adjust our rules to improve the quality of emotion annotation. Our emotion sentence annotation includes the following three steps, with Steps (2) and (3) functioning to improve the annotation quality:

1. Extract emotion sentences: sentences containing emotion keywords are extracted by keyword matching.
2. Delete ambiguous structures: some ambiguous sentences, which contain structures such as negation or modals, are filtered out.
3. Delete ambiguous emotion keywords: if an emotion keyword is very ambiguous, all sentences containing this ambiguous emotion keyword are filtered out.

Polanyi and Zaenen (2004) addressed the issue of polarity-based sentiment context shift. Similar phenomenon exist in emotion expressions. In creating
our corpus, two kinds of contextual structures are considered: the negation structure and the modal structure. To ensure the quality of the emotion-sentence corpus, in Step (2), all sentences containing a negation or a modal structure are removed using some rules plus a list of keywords (negation polarity words for the negation structure and modal words for the modal structure).

To overcome the significant ambiguity of some emotion keywords, in Step (3), five sentences of each emotion keyword are randomly selected and annotated by two annotators. If the accuracy of the five sentences is lower than 40%, this emotion keyword is removed from our emotion taxonomy. Finally, 149 Chinese keywords and 645 English keywords remain.

Tokuhisa et al. (2008) find that a big challenge for emotion recognition is to collect neutral sentences. Since neutral sentences are unmarked and hard to detect, we developed a naïve yet effective algorithm to create a neutral corpus. A sentence is considered as neutral only when the sentence itself and its context (i.e. the previous and subsequent sentence) do not contain any of the given emotion keywords.

One of the main advantages of our unsupervised emotion annotation approach is that it can easily be adapted for different languages and different emotion applications. Another advantage is that it can avoid the controversy regarding emotion annotation. Emotion is subjective, and therefore disagreement for emotion types often arises if the emotion is not expressed through an explicit emotion keyword.

3.4 Emotion corpus

We ran our emotion sentence extraction and neutral sentence extraction on two corpora, the Sinica Corpus and the Chinese Gigaword Corpus, and created two emotion-sentence corpora and two neutral-sentence corpora separately. The Sinica Corpus is a balanced Chinese corpus, which includes documents in 15 kinds of genres. The Chinese Gigaword Corpus is a huge collection of news reports.

To estimate the accuracy of our emotion/neutral sentence extraction, we randomly selected about 1000 sentences from each of the two emotion-sentence/neutral-sentence corpora, and asked two annotators to check it. Table 3 lists the accuracy of those sentences (emotion-sentence corpus/neutral-
sentence corpus). From Table 3, the high accuracy of the neutral corpus proves that our approach is effective in extracting neutral sentences from the document-based corpus containing contextual information. Although the accuracy of the emotion-sentence corpus is lower, it is still much higher than the one Kozareva et al. (2007) reports (i.e. 49.4).

Overall, the annotated corpus created by the unsupervised approach has a comparatively high quality and is suitable for emotion recognition. As the size of the neutral-sentence corpus is much bigger than its corresponding emotion-sentence corpus, we randomly selected some neutral sentences from the neutral corpus, combining them with their corresponding emotion sentences to form a complete emotion corpus, so as to avoid model bias.

4. Emotion Recognition

To compare the above two emotion representations encoded in our emotion annotation scheme in terms of emotion recognition, we designed two kinds of emotion recognition tasks, single-label classification and multi-label classification. In terms of the implementation of multi-label classification, we chose three popular multi-label classification approaches.

4.1 The definition of emotion recognition tasks

According to the emotion representation in our emotion corpus, we define two different tasks of emotion recognition as follows.

(1) The single-label classification task: an emotion is encoded by the holistic representation. To avoid the problem of data sparsity, we label 14 types of emotions (5 primary emotions + the top 9 complex emotions), while other emotions are re-labeled as “Other Emotion.” Together with “neutral” sentences, there are in total 16 labels in our single-label classification. Note that if an emotion sentence contains more than one emotion, it will occur repeatedly and each occurrence is labeled with

<table>
<thead>
<tr>
<th>emotion corpus</th>
<th>neutral-sentence corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gigaword</td>
<td>77.52</td>
</tr>
<tr>
<td>Sinica</td>
<td>73.59</td>
</tr>
</tbody>
</table>
only one emotion.

(2) The multi-label classification task: an emotion is expressed by the componential representation. As explained, there are only 6 labels (5 primary emotions plus “neutral”) in our emotion corpus, and each complex emotion label is replaced with its associated primary emotions according to Turner’s taxonomy. For example, “envy” is the combination of “fear + anger”, and therefore it is re-labeled as “fear” and “anger.” Notice that “neutral” and the five primary emotions are mutually exclusive, whereas the five primary emotions can co-occur simultaneously.

4.2 The selection of classification models

Single-label classification is well-studied in NLP, and there are many approaches, such as MaxEnt and SVM. In this study, we chose MaxEnt in the package Mallet\(^1\) for the single-label classification task. For the multi-label classification task, we selected three popular multi-label classification methods, namely, Binary Relevance (BR), Label Powset (LP), and Hybrid Label Powset (HLP). Assuming \(L, |L| = l > 1\), is a set of disjoint labels, and an instance \(x\) is tagged with a set of labels \(\{y_1...y_i\} = Y \subset L\).

**Binary Relevance (BR).** BR is a one-vs-rest classification method. For each label \(y_i \in L\), a single-label classifier is trained and the corresponding training data is collected as the following mapping: for an instance, if its labels contain \(y_i\), it is retagged as \(y_i\); otherwise it is retagged as “others.” For each test instance, all classifiers are run and the labels that are not “others” are kept. In our emotion recognition there are six classifiers; five classifiers correspond to the five primary emotions (“happiness,” “sadness,” “anger,” “fear,” and “surprise”), and one for “neutral” label.

**Label Powset (LP).** We treat each possible combination of labels appearing in the training data as a unique label, and hence convert multi-label classification to single-label classification. For example, if the labels for a training instance are “happiness” and “fear”, it is relabeled as a unique label, “happiness + fear.” If a testing instance gets the label “happiness + fear”, it is decomposed into “happiness” and “fear.” Notice that, accord-

\(^1\) http://mallet.cs.umass.edu/
According to Turner’s taxonomy, the order of the primary emotions involving in a complex emotion indicates the importance the primary emotions are playing in this complex emotion. Therefore, in our LP system, we treat “fear + anger” vs. “anger + fear” as different tags, and there are 20 labels in total.

**Hybrid Label Powset (HLP).** HLP is a combination of BP and LP. Besides the given features used in LP, the predicted labels from BP are added as a new set of features (refer to Godbole and Sarawagi (2004) for details.)

From the above descriptions of the three multi-label classification methods, we found that all of them eventually are converted into a single-label classification method, and there is no limitation with regard to the implementation of single-label classification methods. In this study, we also selected MaxEnt in Mallet as the underlying single-label classification method. Moreover, two common features in NLP used in the two tasks, the single-label classification task and the multi-label classification task, are word unigram and word bigram in the focus sentences.

### 4.3 Experiments

We merged the two emotion-sentence corpora and the two neutral-sentence corpora into one emotion corpus for our experiments, reserving 80% of this emotion corpus as the training data, 10% as the development data, and 10% as the test data.

As an instance may have several labels, multi-label classification requires more evaluation measures than single-label classification (Tsoumakas and Vlahavas, 2007). We selected three common measures: accuracy (extract match ratio), Micro F1, and Macro F1. The calculation of Micro F1 takes the instance distribution into account, while Macro F1 does not. These three measures can certainly be applied to the evaluation of single-label classification.

First we worked on the single-label classification task. The performance, which is shown in Table 4 (the holistic label), is low, indicating that emotion analysis is a difficult task as explained in Mishne (2005.) To test the plausibility of the decomposition of a complex emotion we designed a simple task of single-label classification based on Turner’s componential representation:
for each instance with more than one label, only its first label remains. As explained above, the order of primary emotions involved in a complex emotion indicates the importance of the primary emotions in that complex emotion. The first label (the first primary emotion) is the typical primary emotion to represent its complex emotion. The performance is shown in Table 4 (the first primary emotion label.) We found that the overall performances significantly improved, which demonstrates that Turner’s decomposition is plausible.

We then ran the three multi-label classification classifiers for the multi-label classification task; the performances are shown in Table 5. Overall, we found that all of these multi-label classification methods outperform single-label classification (57.59% for the holistic label and 62.88% for the first primary emotion label in Table 4.) This indicates that the componential representation permits a classification to detect different facets in an emotion, which in turn helps emotion recognition.

Table 5 also shows that LP achieved the best performance, regardless of which measure is used. Therefore, we looked closely at BP for its comparatively low performance. First, we divided the test data into two parts:

1) Simple test set: containing the instances whose label number is 1 (contains only primary emotions or neutral label)
2) Complex test set: containing the instances whose label number is greater than 1 (contains a complex emotion).

We ran our evaluation for the two data sets; the performances are shown in Table 6. It is a little surprising that the performance for simple test set (53.82%) is much lower than the overall performance of BP (64.25%) in Table 5, and even much lower than the performance of the detection of the first primary emotion label (62.88%) in Table 4. This poor performance may be attributed to the fact that, in BP, the way to merge instances with a complex emotion for each classifier adds noise, which further hinders the detection of emotion in the simple test set. In Turner’s taxonomy, the primary emotions involved in a complex emotion play different roles. However, this information does not taken into account in the BP classification. This can also partially explain why LP outperforms BP.
Moreover, since complex emotions contain at least one primary emotion, we analyze the detection of the first primary emotion and the second primary emotion in the complex test set. From the output of the complex test set with BP (Table 6), we find that the Micro F1 both for the detection of the first primary emotion and for the detection of the second primary emotion are about 45%. However, the overall Micro F1 for the complex test set is 69.88%. This proves that each classifier in BP can detect a facet of a complex emotion, and these classifiers can complement each other.

5. Conclusion

In this study, we designed a cognitive-based emotion annotation scheme based on existing theories of emotions. It is robust and versatile and can encode different levels of emotion information for different types of emotion recognition. Moreover, motivated by NSM, we developed an unsupervised approach to create a large and comparatively high-quality corpus for emotion recognition, the usefulness of which was proven in our pilot experiments.
Automatic Recognition of Emotions

experiments. Furthermore, we explored two different emotion representations (holistic and componential) and their impact on emotion recognition modeling.

Some issues remain unsolved. For corpus construction, we will explore emotion distribution in other kinds of corpora, such as informal and spoken data. We will also analyze ambiguous emotion sentences, such as negation and modal structures. This article presents a pilot experiment for emotion recognition; more work needs to be done on issues such as feature extraction.

References

mental Psychology, 50, 325-338.