Are Emotions Enumerable or Decomposable?
And its Implications for Emotion Processing

Ying Chen, Sophia Y. M. Lee, and Chu-Ren Huang

Department of Chinese & Bilingual Studies,
The Hong Kong Polytechnic University
{chenying3176, sophiaym, churenhuang}@gmail.com

Abstract. Emotion is a complicated concept, and can be represented in different ways. In this paper, we discuss two kinds of emotion representations: the enumerative representation and the compositional representation. Compared to the enumerative representation, the compositional representation is the less rigid description of an emotion. However, from the perspective of emotion classification and detection, different representations often correspond to different emotion processing task. In the enumerative representation, emotion processing can be considered as single-label classification (detecting one and only one label); in the compositional representation, the task turns into the detection of a vector. In this paper, we explore the impact of these emotion representations in emotion processing, including the trade-off of these representations and the selection of technologies to process emotion.

Keywords: emotion, emotion classification, emotion processing, multi-label classification

1   Introduction

Emotions represent one of the most fundamental set of shared human experience, while the recognition and identification of emotions is one of the most crucial human cognitive ability. It is probably not an exaggeration to claim that most human activities are motivated by or designed to excite some emotion. And most events do activate emotion, regardless of whether they are designed to do so. Given the critical roles emotions play in human activities, it is not surprising that sentiment analysis, as coarse-grained account of emotional tendencies (positive, negative, and neutral) became one of the most popular topics in NLP and IE. What is surprising is that there were few studies on emotion computing, which would offer finer-grained information and will be universally applicable regardless of domain and product types.

With regard to emotion processing, some works (Tokuhisa et al., 2008; Mihalcea and Liu, 2006) have been done on text, and most of them use the resource from web, i.e. web blog and analysis that can be explored, such as emotion detection (Tokuhisa et al., 2008), emotion classification (Mishne, 2005; Mihalcea and Liu, 2006), and emotion trend prediction (Mishne & Rijke, 2005; Balog & Rijke, 2006). In this paper, we discuss a basic yet important question in emotion analysis: How to classify and represent emotions?

Although scientific study of emotion can be traced all the way back to early philosophers, both in the West and in the East, we still lack a standard theory of emotion classification today. In terms of emotion classification, the most urgent issue is the nature of emotion taxonomy. Should human emotions be treated as an enumerable, albeit rather large, set of atomic emotions? Or should human emotions be treated as decomposable as a set of primary emotions and their combinations? For example, in Turner's taxonomy (Turner, 2000), "pride" is decomposed into

"happiness + fear". This indicates that the two emotions: "happiness" and "fear", are more basic and can be combined to form complex emotions. However, for emotion classification in NLP, this compositional representation changes the content of classification (detect a vector, not one label), and a different classification technology is required. In this paper, we choose multi-label classification (each instance can have more than one label) to handle the vector detection task. We also discuss the trade-off between single-label classification (e.g. the detection of "pride" only) and multi-label classification (e.g. the detection of both "happiness" and "fear").

The paper is organized as follows. Section 2 gives some related work about emotion processing on text, and provides some background for multi-label classification. In Section 3, we first explain the objective of emotion processing in formal text, and then discuss the two emotion representations, namely the enumerative representation and the compositional representation. Section 4 describes our emotion system and our Chinese emotion corpus, and Section 5 explains the experiments of our study. Finally, a conclusion is made in Section 6.

2   Related Work

Although both sentiment and emotion belong to affective analysis, compared to sentiment task, emotion analysis in NLP, ranging from the corpus construction to the definition of emotion computing task, is still in its early stages. In this paper, we focus on emotion representations. Emotion representation seems to be a fundamental issue to deal with, but it involves many issues, such as data annotation, the selection of classification methods, and so on.

Most work (Mishne, 2005; Mihalcea and Liu, 2006) on English emotion use a blog corpus collected from LiveJournal. In LiveJournal, authors have an option to describe their mood with some words for each post, and those description words are either selected from the predefined list of 132 common moods or just enter free-text by 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) 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.

Although top 40 frequent emotion labels were 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 (each instance contains one and only one label), a common technology for classification in NLP, faces inherited conflict. Single-label classification assumes that the pre-defined labels are mutually exclusive and each instance belongs to one label only. However, this assumption is often invalid in emotion classification.

Alternatively, some emotion theories suggest that an emotion can be represented in a compositional way. In other words, an emotion is expressed by a vector with fixed dimensions. For example, Turner's emotion taxonomy (Chen et al., 2009) provides a way to decompose some complex emotions; Quan and Ren (2009) design a scheme to annotate an emotion corpus for robots, and in this scheme, an emotion is expressed with eight prototypical emotions with other accessory dimensions. Nevertheless, another problem for emotion processing emerges in this compositional representation. The task of the emotion classification turns into an assignment of weights in the given dimensions, not just a label detection. In this paper, we simplify this vector value detection problem in which each dimension has only a binary value (0 or 1). Then, we choose multi-label classification to solve this vector value detection problem.

For NLP, multi-label classification has been widely applied in text categorization (McCallum, 1999) as a document can often be assigned into more than one topic. The important difference between single-label classification and multi-label classification is that multi-label

classification requires to capture the relationships among different labels. Recently, many technologies have been developed to achieve this mutual information (Zhu et al. 2005; Ji et al. 2008; Tang 2009). However, not much work (Trohidis et al., 2008) has been done to use this technology for emotion analysis partially because emotion analysis is still controversial and is not as well-developed as text categorization.

3   Emotion Problem

In this paper, we limit emotion computing as the task of emotion detection and classification. It is straightforward to understand emotion detection task: differentiate emotion sentences and neutral sentences. However, the task of emotion classification is complicated, and some issues need to be discussed.

First, we discuss what kind of text needs emotion detection and classification. Emotion distribution varies in different kinds of corpora, such as blog, online chat, and news. Currently, we concentrate on emotions in formal written text. Then we present the two kinds of emotion representations, the enumerative representation and the compositional representation, and discuss the trade-off of these two representations for real applications.

3.1   The text for emotion detection and classification

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 (e.g. news) is more likely to be expressed by an emotion keyword. Therefore, it seems to be intuitive that emotion detection and classification is satisfactory if the collection of emotion vocabulary is comprehensive and an emotion taxonomy is given. Emotion detection and classification is just to detect the occurrence of those given emotion keywords. For example, a sentence containing the word "joyful" indicates the presence of "happiness" emotion. However, this intuitive approach cannot work well because of the following reasons:

1) As explained above, it is impossible to collect all emotion keywords not only because the size of emotion expressions is very large (Mishne, 2005) but also because emotion expressions evolve from time to time. For example, in Chinese online chat, "雷(lei)" becomes popular to express "shock" emotion;

2) Emotion keywords often have multi-senses, and hence the problem of word ambiguity cannot be avoided;

3) Emotion context also has the problem of context shift as the sentiment shift (Polanyi and Zaenen, 2004).

Therefore, even in formal text, emotion detection and classification is not a simple task, and it is a complicated job, which requires in-depth semantic understanding of texts.

3.2   The representation of emotion

For emotion classification, the difficult and important issue is the choices of emotion representation, which directly decides the content of emotion classification and its related technologies. Two popular emotion representations, namely the enumerative representation and the compositional representation, are discussed as follows.

3.2.1 The enumerative representation
The enumerative representation enumerates an emotion with a unique name, such as "pride", "jealousy", and so on. From the cognitive perspective, how to define and discern an emotion is a big problem, for example "envy" vs. "jealous". From the emotion processing perspective, several points should be taken into account. First, the enumerative 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 the data sparsity

problem (Mishne, 2005). Second, this kind of emotion analysis cannot handle emotions that are not in the emotion list as there is no way to represent it.

When adopting the enumerative representation, it is naïve to treat the emotion classification as a single-label classification problem as most of other NLP tasks do. However, some issues need to be considered.

1) It is possible that several emotions occur in a sentence simultaneously, and the size of all those combinations may be very large so that it is impossible for a classifier to train all of them.

2) It is difficult to capture the complicated relationships between different emotions. Most emotion theories admit that except for few prototypical emotions, an emotion often involves several other emotions. This indicates that an emotion has often inherited relationships with other emotions, and an emotion classification model should have a capability to detect or learn this kind of relationship. Unfortunately, a single-label classification cannot achieve this because of it underlying assumptions.

3.2.2 The compositional representation

Instead of numerating all possible emotions, some emotion theories suggest representing an emotion through a vector with small-scaled fixed dimensions. A simple case is that an emotion is represented by five primary emotions (five dimensions) based on Turner's emotion taxonomy (Chen et al., 2009). This compositional representation is a rather loose way of describing an emotion, and some information may be lost. Moreover, the conversion from an emotion to a vector with fixed dimensions is often inconvertible. For example, Kemper (1987) suggests that complex emotions are resulted from various aspects of social interaction, which are rather culture-specific. In such case, "guilt" (an emotion with the enumerative representation), for instance, apart from being decomposed into joy and fear, may involve other cultural-related moods which are lost in Turner's compositional representation. Hence, one realistic problem for the compositional representation is the selection of dimensions and the way to decompose an emotion so as to capture as much as possible information in an emotion. Most applications choose some prototypical emotions as dimensions and other complement dimensions specifically designed for the applications. The number of those prototypical emotions varies about from four to 12 in different emotion theories (Kemper, 1987).

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

4   The Emotion System

In this paper, we try to compare the above two emotion representations and their impact on emotion computing. First, we use the corpus collected by Chen et al. (2009), and then decompose an emotion according to Turner's emotion taxonomy. Second, we choose some popular classification methods including one single-label classification tool and three multi-label classification approaches to process emotion detection and classification.

4.1   Data

Chen et al., (2009) create a Chinese corpus for emotion detection and classification with an unsupervised method. Here, we briefly introduce it. The corpus includes two parts: emotion corpus (containing sentences with emotions) and neutral corpus (containing neutral sentences only). In the experiment, only a subset of the corpus is used, which totally contains 80,908 sentences (65,060 emotion sentences and 15,848 neutral sentences). To avoid data sparsity, we only focus on 14 kinds of emotions (represented by the enumerative representation). Note that

if an emotion sentence contains more than one emotion, it will occur repeatedly and each occurrence is labeled with only one emotion.

The neutral corpus is created in the following way: a sentence is considered as neutral only when the sentence itself and its contexts (i.e. the previous sentence and the following sentence) do not contain any of the focused emotion words. As Chen et al. work on formal text, the accuracy of this neutral sentence extraction is very high (about 98%). Comparatively, their emotion corpus creation is more complicated, and achieves a decent accuracy (about 77%). There are five steps to extract an emotion sentence:

1. Extract emotion sentences: for a given emotion keyword, the sentences contain this emotion keyword are extracted by keyword matching.
2. Delete ambiguous structures: to guarantee the annotation quality, some ambiguous sentences, which contain some structures, such as negative structure, modal structure and so on, are filtered out.
3. Delete some ambiguous emotion keywords: all sentences containing this ambiguous emotion keyword are filtered out.
4. Annotate with emotion tags: each remaining sentence is marked with its emotion label according to the emotion taxonomy.
5. Ignore the focus emotion keyword: for emotion computing, the emotion word is removed from each sentence.

The underlying foundation for their emotion corpus construction is that most emotion theories support that an emotion is provoked by a stimulus. This indicates one possible way to detect and classification emotions in text, i.e. the detection and classification of emotional stimulus, which is often provided in the text.

The focus of our work is to explore the possible impact of emotion representations for emotion computing. Therefore, although Chen et al.'s corpus by no means comparable with the real emotion corpus as Step (2) and (3) filter out a large size of sentences, we still use it to do emotion detection and classification only with context information.

In addition, Chen et al. (2009) also provide a way to decompose each focused emotion according to Turner's emotion taxonomy. In Turner's emotion taxonomy, an emotion is either a primary (prototypical) emotion or a complex emotion, and a complex emotion is decomposed into some involving primary emotions. The five primary emotions used in Chen et al. (2009) are "happiness", "sadness", "anger", "fear" and "surprise", and they correspond to five dimensions used in this compositional representation. For example, "envy" is decomposed into "fear + anger," which indicates that "envy" contains a greater amount of "fear" and a lesser amount of "anger".

Finally, given the set of data, there are two different tasks for emotion detection classification in our work.

1) Single-label classification: to avoid data sparsity, we choose 14 types of emotions (5 primary emotions + the top 9 complex emotions), and other emotions are re-labeled as "OtherEmotion". Taken the label of "neutral" (neutral sentence) into account, there are totally 16 labels in our single-label classification.

2) Multi-label classification: as explained, there are only 6 labels (5 primary emotions plus "neutral") in multi-label classification. And each complex emotion label is replaced with its involving primary emotions according to Turner's taxonomy. For example, "envy" is a kind of composition of "fear + anger" emotion, and therefore it is re-labeled as the two labels, "fear" and "anger". Notice, "neutral" and the five primary emotions are mutually exclusive, and the five primary emotions can co-occur simultaneously.

4.2 Emotion System

For NLP, single-label classification is well-studied, and there are a lot of choices, such as MaxEnt and SVM. In our system, we choose MaxEnt as our single-label classification. For multi-label classification, we select three popular multi-label classification methods. Here, we briefly describe the three multi-label classification methods, namely Binary Relevance (BR), Label Powset (LP), and Hybrid Label Powset (HLP).

First, we assume L, $|L| = l > 1$, is a set of disjoint labels, and an instance x, is tagged with a set of labels $\{y_1 \ldots y_i\} = Y \subseteq L$.

Binary Relevance (BR): it is one-vs-rest classification. For each label $y_i \in L$, train a classifier and the corresponding training data is collected as the following mapping: for an instance, if its labels contains y_i, retag it as y_i; otherwise, retag it as "others". For each test instance, run all classifiers, and keep the labels that are not "others". In our emotion classification, there are six classifier: 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", relabeled it as a unique label, "happiness+fear". If a testing instance get label "happiness +fear", decompose it into "happiness" and "fear". Notice, according to Turner's taxonomy, the order of a primary emotion involving in a complex emotion indicates its importance of this primary emotion playing in this complex emotion. For example, "fear + anger" means this complex emotion is closer to "fear", whereas "anger + fear" means this complext emotion contains more of "anger". Therefore, in our LP system, we treat "fear + anger" vs. "anger + fear" as different tags, and there are totally 20 labels.

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

In fact, all of these three methods (BR, LP and HLP) finally are converted into single-label classification, and there is no limitation on the choice of single-label classification methods. In this paper, we select MaxEnt as their underlying single-label classification method. The features used in single-label classification, BR and LP are word unigram (1-gram word) and word bi-grams (2-gram word) in the focus sentences.

5 Experiments

We reserve 80% of the 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 (Refer to Tsoumakas & Vlahavas (2007) for more details). We select 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 choose the enumerative representation and run MaxEnt, the performance is shown in Table 1 (the enumerative label). We notice that the performance is still low, which indicates that emotion analysis is a difficult task as explained in Mishne (2005). Then, we choose Turner's compositional representation. To test the plausibility of the decomposition of a complex emotion according to Turner's taxonomy, we design a simple task of single-label classification in this way: for each instance which has more than one label, only its first label remains. As explained, in Turner's taxonomy, the order of primary emotions involving in a complex emotion indicates its importance of the primary emotions in this 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 1 (the first primary emotion label). We find that the overall performances significantly improve, which prove that Turner's decomposition is plausible.

Then, we run the three multi-label classification methods. The performances are shown in Table 2. Overall, we find all of these multi-label classification methods outperform single-label classification (57.59% for the enumerative label and 62.88% for the first primary emotion label). This indicates that the compositional representation permits a classification to detect different facets in an emotion, which further help the emotion computing.

In Table 2, we also notice that LP has achieved the best performance, regardless of which measure is used, and therefore, we look closely at BP for its comparatively low performance. First, we divide the test data into two parts:
1) Simple test set: containing the instances whose label number is 1 (contains only primary emotions and neutral label)
2) Complex test set: containing the instances whose label number is greater than 1 (contains a complex emotion).

Then, we run our evaluation for the two data sets, and show the performances in Table 3. It is a little surprising that the performance (53.82%) for simple test set is much lower than the overall performance of BP (64.25%) in Table 2, and even much lower than the performance of the detection of the first primary emotion label (62.88%) in Table 1. 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 for the simple test set. As mentioned, in Turner's taxonomy, the primary emotions involving in a complex emotion play different roles. In our current BP classification, we do not include this information into account. This can also partially explain why LP outperforms BP.

Moreover, because complex emotion contains at least one primary emotion, therefore we make some analysis for the detection of the first primary emotion and the second primary emotion for complex test set. From the output of complex test set with BP, we found 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%, and however the overall Micro F1 for complex test set is 69.88% (in Table 3). This proves that each classifier in BP can detect a facet of a complex emotion, and these classifiers can complement each other.

Table 1: The performance of single-label classification

	Accuracy	Micro F1	Macro F1
The enumerative label	57.05	57.59	39.49
The first primary emotion label	62.88	62.88	54.17

Table 2: The performance of multi-label classification

	Accuracy	Micro F1	Macro F1
BP	30.48	64.25	61.74
LP	57.53	68.71	64.95
HLP	51.61	66.16	61.5

Table 3: The detailed performance of BP

	Accuracy	Micro F1	Macro F1
Simple test	31.74	53.82	43.96
Complex test	29.44	69.88	56.18

6 Conclusion

In this paper, we explore two different emotion representations and their impact on emotion processing. For the enumerative representation, each label contains rather complete information of an emotion, and the technology for emotion classification is not so complicated. As for the compositional representation, it is a rather loose way to represent an emotion, and therefore it requires more complicated technology for emotion classification. However, from our experiments, we found that the emotion classification becomes simpler for the compositional representation, and at the same time, the compositional way to represent an emotion also permits emotion classification to detect different facets of an emotion, which may be useful for real applications.

For our future work, we will explore more multi-label classification technologies for emotion detection and classification. We will also try to capture the sentence structures for emotion, such as causal event features.

References

Balog, Krisztian and Maarten de Rijke. 2006. Decomposing Bloggers' Moods. *3rd Annual Workshop on the Weblogging Ecosystem, at WWW 2006*.

Chen, Ying, Sophia Y. M. Lee, and Chu-Ren Huang. 2009. A Cognitive-based Annotation System for Emotion Computing. *The Third Linguistic Annotation Workshop (The LAW III), ACL 2009*. Singapore.

Godbole, S. and S. Sarawagi. 2004. Discriminative methods for multi-labeled classification. *PAKDD*.

Ji, Shuiwang, Lei Tang, Shipeng Yu, and Jieping Ye. 2008. Extracting shared subspace for multi-label classification. *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Las Vegas, Nevada, USA

Kempter, Theodore D. 1987. How many emotions are there? Wedding the social and autonomic components. *American Journal of Sociology*, 93, 263-289.

McCallum, A. 1999. Multi-label text classification with a mixture model trained by EM. *AAAI'99 Workshop on Text Learning*.

Mihalcea, R. and Hugo Liu. 2006. A Corpus-based Approach to Finding Happiness. In *Proceedings of AAAI*.

Mishne, Gilad. 2005. Experiments with Mood Classification in Blog Posts. *Style2005 – the 1st Workshop on Stylistic Analysis of Text for Information Access*, at *SIGIR 2005*.

Mishne, Gilad and Maarten de Rijke. 2005. Capturing Global Mood Levels using Blog Posts. *AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006)*, March 2006. Also presented at the *16th Meeting of Computational Linguistics in the Netherlands* (CLIN 2005).

Polanyi, Livia and A. Zaenen. 2004. Contextual Valence Shifters. In J.G. Shanahan, Y. Qu, and J. Wiebe (Eds.), *Computing Attitude and Affect in Text: Theory and Applications*, pp. 1-10.

Quan, Changqin and Fuji Ren. 2009. Construction of a Blog Emotion Corpus for Chinese Expression Analysis. *EMNLP*.

Tang, Lei, Suju Rajan, and Vijay K. Narayanan. 2009. Large scale multi-label classification via metalabeler. *Proceedings of the 18th International Conference on World Wide Web*.

Tokuhisa, R., K. Inui, and Y. Matsumoto (Eds.) 2008. Emotion Classification Using Massive Examples Extracted from the Web. *COLING*.

Turner, J. H. 2000. *On the origins of human emotions: A sociological inquiry into the evolution of human affect*. Stanford, CA: Stanford University Press.

Trohidis, K., G. Tsoumakas, G. Kalliris, and I. Vlahavas. 2008. Multilabel Classification of Music into Emotions. *Proc. 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pp. 325-330, Philadelphia, PA, USA.

Tsoumakas, G. and I. Vlahavas. 2007. Random k-Labelsets: An Ensemble Method for Multilabel Classification. *Proc. 18th European Conference on Machine Learning (ECML 2007)*, pp. 406-417, Warsaw, Poland, 17-21 September 2007.

Zhu, Shenghuo, Xiang Ji, Wei Xu, and Yihong Gong. 2005. Multi-labelled classification using maximum entropy method. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, Salvador, Brazil.