A rule-based approach to emotion cause detection for Chinese micro-blogs

Kai Gao [a,b,1], Hua Xu [a,*,1], Jiushuo Wang [a,b,1]

[a] State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[b] School of Information Science and Engineering, Hebei University of Science and Technology, Hebei 050018, China

ARTICLE INFO

ABSTRACT

Emotion analysis and emotion cause extraction are key research tasks in natural language processing and public opinion mining. This paper presents a rule-based approach to emotion cause component detection for Chinese micro-blogs. Our research has important scientific values on social network knowledge discovery and data mining. It also has a great potential in analyzing the psychological processes of consumers. Firstly, this paper proposes a rule-based system underlying the conditions that trigger emotions based on an emotional model. Secondly, this paper extracts the corresponding cause events in fine-grained emotions from the results of events, actions of agents and aspects of objects. Meanwhile, it is reasonable to get the proportions of different cause components under different emotions by constructing the emotional lexicon and identifying different linguistic features, and the proposed approach is based on Bayesian probability. Finally, this paper presents the experiments on an emotion corpus of Chinese micro-blogs. The experimental results validate the feasibility of the approach. The existing problems and the further works are also present at the end.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In the era of information explosion, the social network applications present a platform for people to share various news and information sources. As for the micro-blog in China, it has gradually become more popular, and millions of people present and share their opinions every day. The micro-blogs can convey almost all aspects of public opinions, including the description of emergencies, incidents, disasters and some other hot events. Perhaps some of them are full of emotions and sentiments. As a result, many researchers in the field of natural language processing pay more attention to Chinese micro-blog textual emotion processing (Liu, 2012), especially the emotion classification (Desmet & Hoste, 2013; Huang, Peng, Li, & Lee, 2013; Kontopoulos, Berberidis, Dergiades, & Bassiliades, 2013; Yang & Yu, 2013), and the corresponding social relations (Hu, Tang, Tang, & Liu, 2013). Generally, individual emotion generation, expression and perception are influenced by many factors. The emotion cause is considered to be the event or condition that can trigger the corresponding emotion, so emotion cause analysis is essential to mine the public opinion and knowledge discovery. In previous related work, the cause events can be composed of verbs, nominalizations and nouns, and these can evoke the presence of the corresponding emotions by some linguistic cues (Lee, Chen, Huang, & Li, 2013a). For example, as for the sentence (1), the causative verb is *"rang4"* (make), and the emotional keyword is *"Kai1 Xin1"* (happy). With the help of the linguistic rules, we can infer that the emotion cause event is *"Zhe4 Ci4 Chun1 You2"* (this spring outing). Meanwhile, Li and Xu (2014) applied the technology of emotion cause extraction to the micro-blog textual emotion classification.

(1) *Zhe4 Ci4 Chun1 You2 Rang4 Wo3 Hen3 Kai1 Xin1.*[2] (This spring outing makes me very happy.)

In this paper, we combine the relevant knowledge in the field of computer science, emotion psychology and the technology of natural language processing together to explore the emotion causes effectively. Firstly, this paper proposes an emotion model with cause events, and the model describes the conditions that trigger

---

* Corresponding author.
    E-mail addresses: gaokai68@139.com (K. Gao), xuhua@tsinghua.edu.cn (H. Xu), wjs8906@163.com (J. Wang).
    [1] Indicates equal contributions from these authors.

---

[2] It is the original Chinese sentence, and the corresponding literal English translation is shown below.

bloggers' emotions in the progress of cognitive evaluation. Meanwhile, all of the sub-events are extracted from micro-blogs. Then this paper detects the corresponding cause events with the help of the proposed rule-based algorithm. Finally, the proportions of different cause components under different emotions are calculated by constructing the emotional lexicon from the corpus and combining different linguistic features.

This paper is organized as follows: Section 2 discusses the related work on various aspects of emotion processing. Section 3 focuses on the method of emotion cause component analysis. Section 4 presents the performance evaluation and the experimental results. In the end, we discuss the remaining challenges and possibilities for the future works.

2. Related work

2.1. Emotion psychological model

As for the cognitive psychology, it focuses on the senior psychological process of human, such as the OCC model which was first proposed by Ortony, Clore and Collins in their book "The Cognitive Structure of Emotions" (Andrew, Clore, & Allan, 1988). The OCC model provided a psychological model of the eliciting conditions of emotions. But it fell short of capturing the logical structure. And Steunebrink, Dastani, and Meyer (2009) proposed a new inheritance-based view of the logical structure of emotions of the OCC model by identifying and clarifying several of the ambiguities. Latter, Steunebrink, Dastani, and Meyer (2012) proposed a formal model of emotion trigger. First, it captured the conditions which triggered emotions in a semiformal way and the main psychological notions used in the emotion model. After that, they proposed a BDI-based framework (belief–desire–intention) to mine the corresponding emotion.

2.2. Emotion classification

In the traditional algorithm on emotion classification, some researchers mainly focus on the following aspects: text processing (e.g., segmentation, part-of-speech tagging, named entity recognition, dependency parsing, etc.), feature extraction and the classification algorithm (e.g., rule-based and machine learning-based methods, etc.). In He (2013), the performances of the three methods (i.e., naive Bayesian, SVM, and SMO) were compared in micro-blog emotion classification. Moraes, Valiati, and Neto (2013) presented an empirical comparison between SVM and ANN for document-level sentiment analysis. Liu, Ren, Sun, and Quan (2013) analyzed the emotion of micro-blog hot events by making use of kernel and SVM method. Wen and Wan (2014) proposed an approach based on class sequential rules to classify the given micro-blog texts into seven emotion types. Li, Li, Li, and Zhang (2014) proposed an approach to generate a multi-class sentiment lexicon by using HowNet, NTUSD and Sina Micro-blog posts. The posts were represented as the lexical vectors based on the lexicon. Then the Semi-GMM and KNN by using symmetric KL-divergence were proposed to classify the lexical vectors for sentiment classification. Liu and Chen (2015) proposed a multi-label classification based approach for emotion analysis, including text segmentation, feature extraction and multi-label classification.

2.3. Construction of the emotional lexicon and multi-language features extraction

As for the emotional lexicon, it can be used to process and identify the emotional words. It usually can be constructed by manual process and acquiring automatically from the corpus (Xu, Liu, Pan,

Ren, & Chen, 2008). In the process of emotion analysis, identifying the multi-language features in micro-blog posts is necessary to compute the emotion intensity scores. As for the related work, Zhai, Xu, Kang, and Jia (2011) exploited effective features for Chinese sentiment classification, such as sentiment words, substrings, substring-groups, and key-substring-groups features. Li, Pan, Jin, Yang, and Zhu (2012) expanded a few high-confidence sentiment and topic seeds in target domain by the given RAP algorithm. Ren and Quan (2012) made an analysis on emotion expressions, including the following factors for emotion change: negative words, conjunctions, punctuation marks and contextual emotions. Quan, Wei, and Ren (2013) combined sentiment lexicon and dependency parsing for sentiment classification, and they extracted the evaluation objects based on the dependency and calculated the similarity between the words based on HowNet. Zhang, Xu, and Xu (2015) focused on the semantic features between words by clustering the similar features on the basis of word2vec.

Unlike the related methods, this paper incorporates the method of Chi-squared test, PMI and word2vec into the construction of the emotional lexicon based on Chinese micro-blog corpus.

2.4. Emotion cause analysis

In the aspect of emotion cause analysis, emotions can be invoked by the cause events. Lee, Zhang, and Huang (2013b) presented an event-based emotion corpus to analyze the interaction between event structures and emotions in the text. Lee et al. (2013a) constructed a Chinese emotion cause annotated corpus and presented seven groups of linguistic cues and two sets of generalized linguistic rules for the detection of emotion causes. Li and Xu (2014) proposed and implemented a novel method for identifying emotions of micro-blog posts, and tried to infer and extract the emotion causes by using knowledge and theories from other fields such as sociology. Nguyen, Phung, Adams, and Venkatesh (2013) extracted events using behaviors of sentiment signals and burst structure in social media, and these events often caused the behavioral convergence of the expression of shared emotion. Rao, Li, Mao, and Liu (2014) proposed two sentiment topic models (i.e., Multi-label Supervised Topic Model (MSTM) and Sentiment Latent Topic Model (SLTM)) to extract the latent topics that evoked emotions of readers, then the topics were seen as the causes of emotions as well.

Unlike the above works, this paper presents a novel emotion cause component analysis method for Chinese micro-blog posts. The innovation of this paper lies in mining the micro-blog data by analyzing the corresponding emotion on the basis of an emotion model. And then this paper also presents the subsystem for detecting and extracting the cause events by the designing rules in fine-grained emotions. Thirdly, constructing the emotional lexicon and identifying the multi-language features in micro-blog posts are used to analyze the emotion intensity scores. Finally, this paper presents the proportions of different cause components.

3. Emotion cause component analysis

The main flow on emotion cause component analysis is shown in Fig. 1. On the basis of the corresponding processing, the proportions of different cause components under different emotions will be obtained.

3.1. ECOCC model construction

Based on the cognitive theory, this paper improves the structure about the eliciting conditions of emotions in accordance with the OCC model referred in Andrew et al. (1988) and presents an