Context Based Emotion Detection from Text Input

Jianhua Tao

National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, China
jhtao@nlpr.ia.ac.cn

Abstract

Emotion detection was normally conducted from the viewpoint of prosody and articulation features. There is still an opening question on how to extract the emotion from the text input. To solve the problem, the paper generates an emotion estimation net (ESIn), which combines the content words and emotion functional words to estimate the final emotion output. In the paper, emotion functional words are also classified into emotional keyword, modifier word and metaphor word. To make more detailed word classification, some context information was analyzed. Both experiments and cross tests show that the method could generate the good results for emotion detection from text input.

1. Introduction

Emotion can be expressed as joy, sadness, anger, surprise, hate, fear and so forth. The classification category is not defined clearly yet. So we focus on the related research about emotion in cognitive psychology domain. In 1988, Andrew Ortony, Gerald Clore and Allan Collins published the book called The Cognitive Structure of Emotion, in which they explain the emotion system provided. Their emotion structure model was named as OCC model[1]. According to OCC’s theory, all emotions can be divided into terms according to the emotion eliciting situations, consequences of events, actions of agents and aspects of objects. According to the classification of emotion-elicitating situations, all emotions can be divided into 3 classes, 6 groups and 22 types.

While the broad topic of emotion has been studied in psychology for decades, very little effort has been spent on attempting to detect emotion in text. Chuang[14] has developed a semantic network for emotion extraction from textual content, but there are less corpus to support the results. In the paper, we assume that the emotion reaction of an input sentence is essentially represented by its word appearance. To get emotion state, all of the words are divided into content words and Emotion functional words (EFWs). They are manually defined and used to extract emotion from the input sentence. All of the extracted emotional functional words have their corresponding connection to “basic emotion values” which are defined in the lexicon. For each input sentence, the basic emotion values are combined to give the final emotion output with emotion estimation net. Emotion functional words are also classified into emotional keyword, modifier word and metaphor word. To make more detailed word classification, some context information was analyzed. Both experiments and cross tests show that the method could generate the good results for emotion detection from text input.

The paper is organized as following. Section II describes the idea of emotion functional words, which are classified into emotional keywords, modifier words and metaphor words. With the classification and analysis, a lexicon which contains the emotion tagging and some semantic labeling was established in Section III. In Section IV, the context-based emotion extraction module and the integration of modules are detailed. Based on this, an emotion estimation net was established for emotion detection. Finally, the experimental results of the integrated emotion prediction system, from both neutral & emotion comparing and cross testing among emotion states, are shown in Section V and some conclusions will be given in Section VI.

2. Emotion Functional Words

There are two ways for emotion generation. One is from semantic information, human likes to presenting the emotion according to what he (or she) wants to say. The other is from the environment and psychology. Sometimes, the content is not very important, if he (or she) has a strong feeling to express the emotion while they are in different situations. To extract the appropriate emotion state from context information, each input sentence could be considered as the combination of content word and emotion functional words (EFWs), though most of them only contain content words. EFW is a kind of word which could be linked to the special emotion states or has some influence on them. In emotion detection, they supply the basic emotion values or connection. With the semantic relation, the emotion value will be propagated in the whole sentence.

2.1. Emotional keywords

The most important words in EFWs are emotional keywords, which provide the basic emotion value of the input sentence. Normally, it is not easy to classify the words to different emotion state, emotion hides in human’s experience long before the history of language. There are lots of ambiguities, most of them occur in anger and sadness. For example, the word “unhappy” may indicate “angry” or “sad”, according to the different personality and situation. To get more accurate description, we assign the weight for each of them which behave as the emotion intense. For example, we assigned the weight with 0.5 (for angry) and 0.5 (for sad) for the word “unhappy”. It means “unhappy” has equal possibilities for emotion state angry and sad. The final results are the combination and propagation with these values.

In Chinese, there are 462 emotional keywords in total. Most of them are noun, verb and adjective. To reduce the complexity, we only choose 6 types for the basic emotion states labeling, joy, sadness, anger, surprise, hate and fear.

2.2. Modifier words

Except for the emotional keywords, the emotion state could also be influenced by some words which behave as modifier.
in the sentence, such as “very, so, too much, not, etc.”. Normally, emotional keywords represent the major emotion reaction of the sentences related to a certain topic. The modifier words are normally used to enhance or weaken the mood. For example, “I’m so angry.” The phrase “so angry” denotes the key emotion state of the sentence, and is extremely emphasized. The effect of “intense” mood could be obviously represented by emphasizing the emotion keywords.

2.3. Metaphor words

The other part of EFWs seems to have no direct action on the emotion states but do have the latent influence on them. They denote the attitude and moral character of the person to make positive and negative influence on emotion states. For example, “asperity” is more like related to exaggerating and negative emotion, “anger” and “hate”, but “kindness” always concerns the gentle and positive emotion, “joy” or sometimes “neutral”. In our work, the metaphor words are further divided into two types, one is for spontaneous expressing, such as “anxious, deferential, ardent, fierce, etc.”, the others only denote personal character, such as “chipper, arrogant, etc”. The whole amount of metaphor words in our work is 440. Most of them are adjectival, among which 201 are related to positive feeling and the others are for negative expression.

3. Lexicon Structure

To generate a lexicon for emotion detection from context, we create a lexicon. It contains 65620 words. All of the word items in lexicon are classified into two major types, content word and EFW. If the word type is emotional keyword, it contains six tags of emotion states (joy, sadness, anger, surprise, hate and fear), and corresponding weights. For modifier word and metaphor word, they are more related to exaggerating emotion or introversive expression. We mark a coefficient in the lexicon for them.

In order to eliminate the error due to subjective judgment, all words are labeled by four people and then crossly validated by the other two people. For each word, if the results tagged by different people are close, the average of these values will be set as the basic emotion value of the word. If the three people can not reach a common consensus, an additional person will be asked to tag the word and the result will be taken into consideration. Experimentally, only few words need additional suggestion. The whole lexicon is organized in figure 1.

To get the relationship among content words and EFWs, some basic semantic tagging is introduced in the lexicon. Lexical semantics begins with a recognition that a word is a conventional association between a lexicalized concept and an utterance that plays a syntactic role. POS and semantic tree are the basic features in semantic tagging.

In order to reduce ambiguity in emotion and semantic integration, therefore, “word form” is used here to refer to the physical utterance or inscription and “word meaning” to refer to the lexicalized concept that a form can be used to express. Then the starting point for lexical semantics can be said to be the mapping between forms and meanings (Miller, 1986). A conservative initial assumption is that different syntactic categories of words may have different kinds of mappings. In our work, all of the semantic tagging is driven from HowNet.

4. Emotion Estimation

Sometimes, the emotion state can just be decided by emotion keywords and other labeling information, the ambiguity may appear in some situation. It mainly results from the multivocal of emotion keywords. The emotion reaction is firstly deduced from initial emotion assign according to word classification above, and also from the combination of the semantic relations. In order to get the context sensitive model, we propose a unified architecture based on Emotion eStimation Net (ESiN) that seamlessly integrates context-dependent probabilistic hierarchical sub-lexical modeling.

ESiN is composed of nodes and routes. Each node denotes a word which contains three attributes: emotion states, the corresponding weights and semantic tagging. The route of ESiN represents the propagation of the emotion. It has three attributes: direction, transmission probability \( P_{ij} \) (denote the probability from one node \( i \) to another node \( j \) ) and propagation decreasing coefficient \( \alpha \).

The emotion value of the node \( t \) is got from,

\[
\dot{E}_t = D(E_{t-1}) = \dot{E}_t + \dot{E}_t \exp(-\alpha \times t^2) + C_t, \tag{1}
\]

Where, \( \dot{E}_t = (e_{t,joy}, e_{t,sadness}, e_{t,anger}, e_{t,surprise}, e_{t,hate}, e_{t,fear}) \) contains the emotion values of all emotion states. \( \dot{E}_t \) is the emotion values generated in node \( t \). To get \( \dot{E}_t \) there are three different situation according to node type.

- **If the current node is content word, \( \dot{E}_t \) will be got from the semantic relation between two nodes.**

\[
\dot{E}_t = P(o_t | o_{t-1})\dot{E}_{t-1} \tag{2}
\]

- **If the node is emotional keyword, \( \dot{E}_t \) will be got from the initial emotion weights defined in the lexicon.**

\[
\dot{E}_t = (\omega_{t,joy}, \omega_{t,sadness}, \omega_{t,anger}, \omega_{t,surprise}, \omega_{t,hate}, \omega_{t,fear}) \tag{3}
\]

- **If the node is modifier or metaphor word, all of the \( \dot{E}_t \) should be multiplied by the coefficient \( \beta \).**

To make more detailed description, we give a sample as following.

“Mr.Wang is too introversion to speak out, though he feels very pleasure while he hears the news.”

Detection of emotion in text by ESiN is then followed by the following steps.
4.1. EFWs Detection (First Step)
In the first step the text is tagged with a POS tagger. The tagger learns sentence structures for a language as a set of transition rules. These rules are then applied to the text to label each word as a noun, verb, etc. Once words are labeled, they are checked for EFWs and assigned an emotional rating.

For above sample, the EFWs are emotion keyword “pleasure”, modifier word “too, very”, and metaphor word “introversive”.

4.2. Weight Assign and Link Construction (Second Step)
The second step is to assign the weight for emotional keywords and construct the link among EFWs. Here, “pleasure” is tagged as emotion “joy” with weight 0.9 from the lexicon.

There are two modifier words in the sample, but only “very” linked to emotion keyword “pleasure”. “very” makes the positive action to enhance the score in “pleasure”. On the contrary, metaphor word “introversive” could decrease valence score in the whole sentence.

4.3. Propagation, Collection and Decision (Final Step)
In propagation, the emotion keyword is considered as the propagating source. The scores are then summed across all sentences and finally run through a fuzzy-logic process to determine an overall score for the correspondence. In ESiN, valence is determined from a proprietary list of emotionally charged words, abbreviations, and emoticons. Administrators of the system are free to add new emotion words, or change charged words, abbreviations, and emoticons. Administrators of the system are free to add new emotion words, or change charged words, abbreviations, and emoticons. Administrators of the system are free to add new emotion words, or change charged words, abbreviations, and emoticons.

The aim of the emotion trigger is to integrate the non-zero emotion vector according to the emotion state history by path searching.

\[ M_i = \arg \max_u \left( \sum_{j=0}^{n} e_{uj} \right) \]  

(4)

5. Experiment and Discussion
To test the emotion estimation results, we make some experiments. The text is from a spontaneous speech corpus. There are 835 sentences, with about 5000 words. All of the sentences were labeled with emotion state manually. Table 3 shows the tagged results.

Table 1: Tagged results of testing samples

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy</td>
<td>136</td>
</tr>
<tr>
<td>sadness</td>
<td>3</td>
</tr>
<tr>
<td>anger</td>
<td>26</td>
</tr>
<tr>
<td>surprise</td>
<td>67</td>
</tr>
<tr>
<td>hate</td>
<td>17</td>
</tr>
<tr>
<td>fear</td>
<td>5</td>
</tr>
</tbody>
</table>

5.1. Neural vs. Emotional

The first comparing is conducted between neural states and emotional states. The testing results are shown in figure 2.

In figure 2, x-axis denotes the word number and y-axis denote the accurate. The solid line is the accurate rate of neural state and dash line reflects the accurate rate of neural state and dash line.

These results did not exhibit the coherence of the other results reported. As is exhibited in Figure 2 the maximums did not occur at the same point (although local maximums could be found at both points in most cases). This is probably due to the inability to develop adequate models for the less represented emotions because of their scarcity in the data. Nonetheless, performance peaked at over 80% for the Neutral class, and at over 70% average for precision of the Emotional class.

5.2. Cross-Testing
As a classification task, emotion detection should be evaluated with consideration on all the emotion states. Compared to manual labelling results, it is easy to generate a confusion matrix, represented as \( N_{ij} \). \( N_{ij} \) is the amount of emotion state whose manual labelling is \( i \) but predicted as \( j \).

Table 2: Matrix for cross testing

<table>
<thead>
<tr>
<th></th>
<th>joy</th>
<th>sadness</th>
<th>anger</th>
<th>surprise</th>
<th>hate</th>
<th>fear</th>
</tr>
</thead>
<tbody>
<tr>
<td>joy</td>
<td>136</td>
<td>4</td>
<td>23</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>sadness</td>
<td>0</td>
<td>60</td>
<td>11</td>
<td>0</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>anger</td>
<td>5</td>
<td>26</td>
<td>67</td>
<td>21</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>surprise</td>
<td>27</td>
<td>25</td>
<td>37</td>
<td>136</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>hate</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>37</td>
<td>8</td>
</tr>
<tr>
<td>fear</td>
<td>1</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>5</td>
<td>62</td>
</tr>
</tbody>
</table>

From Table 2, we can find there are still lots of confusion in emotion detection. The performance of the Emotional class studied in these experiments was not as significant as that of the Joy class. Nonetheless, it was substantially greater than chance and thus a significant result nonetheless. The lower result can be attributed to the fact that many of the less represented emotions are part of the emotional class. Thus, since these emotions are not well represented they could not be learned. This assumes that model is still looking for particular emotions in the emotional classification, though.

5.3. The analysis of syntactic structure
To get more accurate emotion detection, we still need to know how content words behave in the sentence. Do they also have the influence on the emotion? The answer is “YES”. A linguistic analysis for the emotion must consider the underlying syntactic structure in terms of its hierarchical organization, especially if the syntax of a given language allows different directions of branching as it is the case in an Object-Verb-language (OV) such as Chinese. In Chinese, the syntactic OV-parameter means that in structures with verb-final word order in most subordinate clauses, the verb takes its argument from its left. For reasons of explanatory adequacy, we make use of theories which consider the information structure. In Jacobs (1993), for both the so called ‘normally intonated’ sentences, widely focused sentences,
and sentences containing narrowly focused constituents, focus positions are predictable by terms of integration.

As we know, Chinese is a tonal language and each syllable has a tone, and has a relatively steady F0 contour. It is difficult to determine the emotion within the influence of various syllabic tone patterns. According to recent tonal sequence models (Reyelt, Grice, Benzmüller, Mayer, and Batliner 1996 for German), the main focus derived from the syntactic and information structure as described above serve as anchor points for the association of tonal sequences. In Chinese, emotion is assumed to be realized preferably by tonal variation.

To get the syntactic parsing results, Tomita algorithm with 326 emotion patterns were adopted in the paper. Both Semantic and modal lexicon were also used. Whereas, due to the lack of sophisticated syntactic and semantic parsing algorithm, it is still a very challenging work to determine the emotion with very high quality in real time.

5.4. Emotion Markup Language
As mentioned above, emotion can be influenced by the environment where we are living. For example, we may change mood while we see or hear different things, sometimes happy, sometimes sad. Recent experiments and discoveries have allowed for the creation of a model of the emotional mind that explains how much of what we do is emotionally driven and to what extent emotions have their own logic. To integrate the labeling information in the text input, we provide the emotion expression function to control emotion more conveniently. Contents provider can specify 22 types of emotion defined in OCC emotion model, to modify the action. The speech synthesis system expresses the emotion with performing different actions, changing the pitch, volume, speed, emphasis of the voice. For example, when the emotion type is specified as “pride”, it will speak loudly with the emphasis at the beginning of the sentence.

![Figure 3: Emotion Markup Language](image)

According to the above script, the system would give a self-introduction with the "pride" emotion activated at first.

6. Conclusions
The paper analyzes the context information for emotion detection from text input. Emotion functional words were defined in the paper to generate the emotion focus and emotion estimation net. The emotion output is the combination of the emotion value and propagation function. Experiments proved that human emotion was deeply influenced by the concept expressed by the speaker. As we know, it’s still hard to us to do the automatic concept analysis and semantic parsing with machine learning method. Nevertheless, part of syntactic information and the features of some environment parameters, such as background music, dialogue speech, facial detection, have been worked out in the system. The results show that we’ve got relatively good results for emotion detection from text input. With the function of environment awareness, the emotion can be easily adapted into new environments. Testing as well will be conducted on other dialogue driven domains such as drama and literature dialogue. Finally, models from one domain will be tested on another to show the extensibility of a model built using this method.

7. References

[3] Donna Erickson,1 Arthur Abramson, “Articulatory characteristics of emotional utterances in spoken English”, ICSLP2000