ISCA Archive
http://www.isca-speech.org/archive

7th International Conference on Spoken Language Processing [ICSLP2002]
Denver, Colorado, USA
September 16-20, 2002

EMOTION RECOGNITION FROM TEXTUAL INPUT USING AN EMOTIONAL SEMANTIC NETWORK

Ze-Jing Chuang and Chung-Hsien Wu
Department of Computer Science and Information Engineering,
National Cheng Kung University,
1, Ta-Hsueh Road, Tainan, Taiwan, R.O.C
bala@csie.ncku.edu.tw, chwu@csie.ncku.edu.tw

ABSTRACT

This paper presents an emotion recognition system with textual input. In this system, an emotional semantic network is proposed to extract the semantic information related to emotion. The semantic network is composed of two subnetworks: a static semantic network and a dynamic semantic network. The static semantic network is established from an existing Chinese knowledge base called HowNet and used to estimate the emotion trigger value of each word. The dynamic semantic network accepts the textual input and dynamically constructs the nodes and links, which represent the emotion carrier and the emotion propagator respectively. Initiated by the emotion trigger value, the emotion in the dynamic semantic network will propagate and finally converge to the final emotion output. Experimental results show an encouraging achievement was obtained.

1. INTRODUCTION

The techniques of emotion recognition have been developed for several decades and usually focused on the physiological characteristics, such as voice, facial expression, hand gesture, or body movement. Yanaru [1] took the footages of real speakers when they talked about some emotional keywords, and defined different emotion reactions to these keywords. Chan and Franklin [2] analyzed the input sentences and constructed a symbolic network to reduce the perplexity of each word. Woods [3] used the transition network to analyze the natural language. In other emotion research, speech and image signals are the most common information for emotion extraction [4][5], and there are also some research combining different emotion recognition models [6][7][8]. Recently, many encouraging results have been achieved, but very few researches tended to recognize the emotions from only textual input. However, natural language plays an important role in emotion recognition. In this paper, an emotion recognition system with textual input is proposed. This approach is hopefully to be integrated with other physiological characteristics to achieve a more robust emotion recognition.

Due to the diversity of natural language, it is more difficult to recognize the emotions from the textual information than from other kinds of information. The main problem is that the emotion representation of a single sentence is ambiguous and context-dependent. In order to handle the complexity of natural language, we propose an emotional semantic network in our emotion recognition system. Figure 1 shows the block diagram of our system. In this system, there are two subnetworks: a static semantic network and a dynamic semantic network. The static semantic network is established from HowNet and used to estimate the emotion trigger value of each word. The dynamic semantic network accepts the textual input and dynamically constructs the nodes and links, which represent the emotion carrier and the emotion propagator respectively. Initiated by the emotion trigger value, the emotion will propagate and finally converge to the final emotion output.

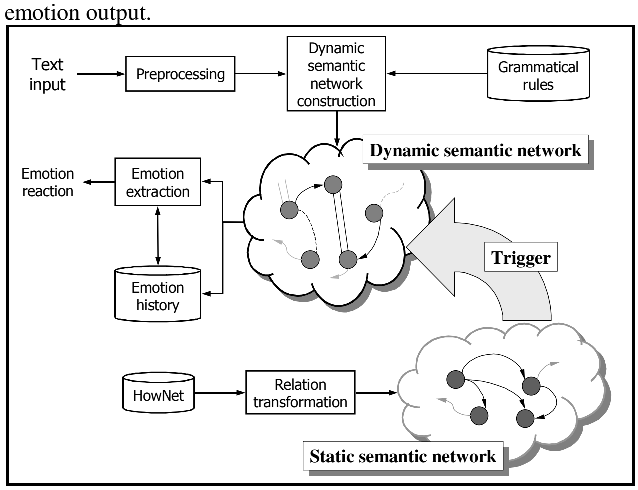

Figure 1: Block diagram of the emotion recognition system.

2. STATIC SEMANTIC NETWORK

2.1. Definition of emotion trigger value

In previous approaches, emotions were generally classified into six basic emotions: happiness, anger, surprise, sadness, fear, and antipathy. In this approach, these six basic emotions are further classified into two groups: positive emotions and negative emotions. The classification result is listed in Table 1.

Table 1 Emotion groups

Group	Emotions
Positive emotion	Happiness, Anger, Surprise
Negative emotion	Sadness, Fear, Antipathy

In general, the positive emotion presents stronger and higher reaction compared to the negative emotion. Corresponding to the emotion groups, we can also classify the verbs into two groups: positive verb and negative verb. The positive verb increases the positive emotion and decreases the negative emotion, while the negative verb decreases the positive emotion and increases the negative emotion. In order to quantify the emotional effect of different kinds of verbs, we define an emotion trigger value T_D. The value of T_D ranges from -1 to +1 indicating the effect from maximum negative level to maximum positive level. A verb with zero value of T_D shows that it is unrelated to emotions.

2.2. HowNet structure

HowNet is a conceptual relationship database for Chinese language.

The word concepts, or definitions, in HowNet are composed of basic sememes, such as "human," and sometimes accompanied with "pointer" that express certain kinds of relations. For example, the word "老師(teacher)" is associated with three definitions:

W_C=老師

W_E=teacher

*DEF=human|人,*teach|教,education|教育*

Where "*" denotes a relation of "agent or instrument of an event." So this definition can be translated as: "teacher" is essentially related to "human" and "education" and can do the action of "teach."

2.3. Emotion trigger value generation using HowNet

There are about 3,500 definitions in HowNet and only about 1,000 definitions have emotional effect. We tagged the emotion trigger value of each definition manually and calculated the emotion trigger values of all concepts according to the relationship in HowNet. The emotion trigger value is estimated as follows:

If a concept C is defined by k definitions (D_1 to D_k) with k relationships (r_1 to r_k), shown in Equation (1):

$$DEF = r_1\,D_1,\, r_2\,D_2,\, \ldots,\, r_k\,D_k \tag{1}$$

The emotion trigger value T_C of concept C is calculated as:

$$T_C = \frac{1}{k}\sum_{i=1}^{k} F\left(r_i, T_{D_i}\right) \tag{2}$$

Manually

Where T_D indicates the emotion trigger value of definition D and ranges between -1 and +1. The function $F()$ in Equation (2) is used to transform the emotion trigger value from definitions to concepts. For each relationship, the transformation function is defined and listed in Table 2.

Table 2: Transformation functions. (s is the sign of T_D)

Relationship symbols	Functions
NONE	$0.5\times T_D$
%	$s\times(2^{\lvert T_D\rvert}-1)$
&	$s\times(2^{\lvert T_D\rvert}-1)$
?	$s\times(2^{\lvert T_D\rvert}-1)$
*	$s\times\log_{10}\left(9\times\lvert T_D\rvert+1\right)$
$	T_D
@	0
#	$0.2\times T_D$
^	$-T_D$

3. DYNAMIC SEMANTIC NETWORK

The dynamic semantic network is the main component in our system and is used to estimate the emotion reaction of the textual input.

3.1. Parsing of Chinese language

Since we are only interested in the emotion representation of a sentence, a parser is designed to focus on the special syntactic structures useful for emotion recognition. There are three steps in the parsing process: (a) word segmentation, (b) POS to syntactic constituent mapping, and (c) grammatical rules mapping.

3.1.1. Word Segmentation

The goal of word segmentation is to convert the Chinese sentence structure from character sequence into word sequence. In our system, a general n-gram model is used to perform word segmentation. This system can extract the possible words in an input sentence and provide possible POSs of the words.

3.1.2. POSs to syntactic constituents mapping

Basically, different POS structures coveys different emotions. According to the analysis of Chinese POSs, we induce the following rules:

- The predicate is the main emotion propagator. Some predicates have two objects but some have only one. We use different link types to handle these two kinds of predicates.
- Although most of the prepositions have no effect on emotion reaction, some prepositions are still useful for emotion recognition and they are included in our grammatical rules.
- The auxiliary and interjection are two special POSs and play important roles in emotion reaction. These POSs are manually tagged and included.

In order to disambiguate the syntactic constituent of each word, we induce the mapping from POSs to syntactic constituents, which is shown in Figure 2.

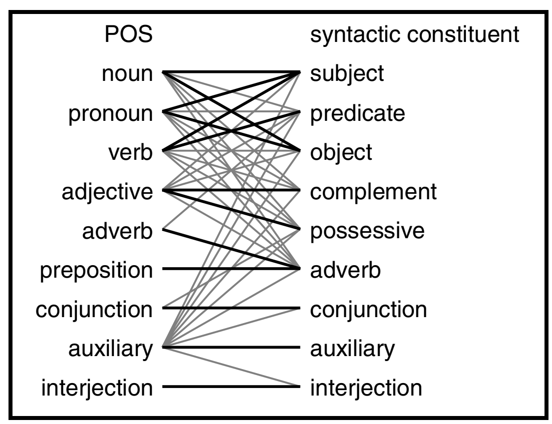

Figure 2: Mapping from POSs to syntactic constituents

The connective lines in Figure 2 show the available mapping from POSs to syntactic constituents. It is too complex to build all the grammatical rules. We only consider the most frequently used mapping, which are shown by the bold lines.

3.1.3. Chinese grammatical rules mapping

So far we have found the candidates of syntactic constituents for each textual input. The next step is to select the best one using Chinese grammatical rules. There are lots of grammatical rules in Chinese language, and most of them are unrelated to emotion reaction. We have collated 19 major rules and constructed the dynamic semantic network based on these rules. We use symbols S, P, O, C, and PP to indicate the most commonly used POSs "subject," "predicate," "object," "complement," and "junction," respectively. After the above preprocess, each input sentence has been converted into a sequence of syntactic constituents.

3.2. Network construction

The primary components in the dynamic semantic network are nodes and links. The nodes (emotion carrier) can carry the emotional information while the links (emotion propagator) can propagate the emotional information. Each node in the dynamic semantic network represents a subject or an object term in the input sentence. There are five components in a node:

- Word name: the words in the input sentence.
- Trigger value acceptor (T_C): The acceptor accepts the emotion trigger value from the static semantic network for emotion propagation.
- Initial emotion vector (E_i):
  The initial emotion vector records the initial trigger values from the static semantic network. This vector will keep unchanged until the end of the textual input.
- Emotion propagation value (E_p):
  The emotions are propagated when accepting a new sentence input. The emotion propagation values are recorded and put into the history. Finally the emotion propagation converges to give the final emotion results.
- Emotion vector (E_t):
  The emotion vector with six dimensions, each ranging from -1 to +1, represents the six basic emotions of a word. The emotion vector is recursively calculated from initial emotion vector, emotion propagation values, and emotion history.

The other important component is links. Each link in the dynamic semantic network indicates the predicate in the input sentence. According to different grammatical rules described above, there are different link types in the network. Totally we define four types of links as follows, shown in Fig. 3.

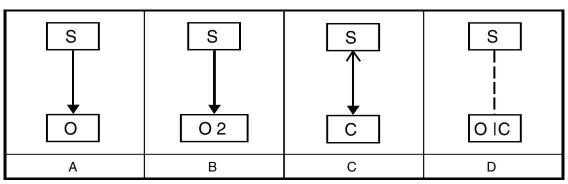

Figure 3: Link types in dynamic semantic network.

- Type A: Link from subject to object:
  Suitable for the predicate that has only one object. This is the simplest link, and the emotion propagation is directive.
- Type B: Link from subject to indirect object:
  Suitable for the predicates that have two objects: direct object and indirect object. The emotion propagation is directive.
- Type C: Bi-directional link between subject and complement:
  The complement is usually regarded as an embellishment of the subject, so we use the bi-directional link to present the bi-directional influence between these two POSs. The emotion propagation is directive but unbalanced.
- Type D: Equal link between subject and object:
  Suitable for the linking verb. The emotion propagation is directionless. Since the subject and object are equally related, the emotions can propagate without any decay.

There are three information contained in a link:

- Word name: the words in the input sentence.
- Direction: the source and the destination of the link.
- Emotion decay function: The decay function controls the emotion propagation.

4. EMOTION ESTIMATION

4.1. Emotion propagation

The propagation of emotions is triggered by a new input sentence and ends with the emotion propagation values below a threshold. In the emotion propagation process, the sources of the propagation are firstly decided according to three criteria as follows:

(a) The pre-tagged emotional keywords
(b) The pre-tagged predicates in the broadcast drama
(c) The words with emotion trigger value.

When an input sentence contains emotional keywords in case (a), these words are used as the propagation source. If the emotional keywords do not appear, we look for the pre-tagged predicates in case (b). If neither case (a) nor case (b) happens, the words with emotion trigger value are used as the propagation source.

Emotion propagation value E_p is then calculated from the source and combined with the emotion vector E_{t-1} using a transformation decay function $D()$. The decay function $D()$ is defined to ensure that the emotion can propagate appropriately. In order to ensure the convergence of the propagation, the emotion propagation will decay to zero after some propagation steps. According to the above criterion, we define the decay function as follows:

$$E_P = D(E_{t-1}) = \begin{cases} E_{t-1} & \text{, equal link} \\ \delta(l) \times W(E_{t-1} + 0.5 g \times T_C) & \text{, other wise} \end{cases}$$

$$\delta(l) = \exp(-0.05 \times l^2) \qquad W(x) = \begin{cases} 1, & -1 \le x \le 1 \\ 0, & \text{other wise} \end{cases} \tag{3}$$

Where E_{t-1} denotes the emotion vector that is calculated in the last propagation step. l is the number of propagation steps from the propagation source, and g is the sign of emotion vector E_{t-1}. Functions $\delta()$ and $W()$ limit the emotion propagation ranging from -1 to +1 and decaying to zero after ten links.

4.2. Emotion history

The current emotion vector E_t was calculated recursively from E_p and emotion history E_{t-1}. The emotion history calculated in Equation (4) is used to remain the continuity of the emotion.

$$E_0 = E_i \qquad E_t = E_{t-1} + 0.5^{(E_p - E_{t-1})} \tag{4}$$

Where E_t indicates the emotion reaction for the t-th input sentence. E_i and E_p indicate the initial emotion vector and the emotion propagation value respectively. The final emotion reaction is the average of the values of E_t over all nodes in the dynamic semantic network.

5. EXPERIMENTAL RESULTS

In order to obtain the natural emotion reaction, we collected the text data from a broadcast drama. The dialogues between the leading man and the leading woman were transcribed as our training corpus. There are 558 sentences contained in 137 dialogues from the leading man, and 453 sentences contained in 136 dialogues from the leading woman. We tagged the emotion reaction of each sentence manually. Table 3 shows the tagged results.

Table 3: Tagged emotion results.

	Male	*Female*
happiness	104	121
anger	92	80
surprise	84	66
sadness	137	92
fear	108	63
antipathy	33	31
total	558	453

In the broadcast drama, the numbers of the sentences for the six emotions are not equal. Table 4 and Figure 4 show the results of emotion recognition.

Table 4: Emotion recognition result.

	Male			Female			Average
	Tag	**Rec**	**Rate**	**Tag**	**Rec**	**Rate**	
happiness	104	72	69.23%	121	80	66.12%	67.67%
anger	92	71	77.17%	80	63	78.75%	77.96%
surprise	84	54	64.29%	66	49	74.24%	69.26%
sadness	137	83	60.58%	92	56	60.87%	60.73%
fear	108	62	57.41%	63	35	55.56%	56.48%
antipathy	33	14	42.42%	31	7	22.58%	32.50%
total	558	356	63.80%	453	290	64.02%	63.91%

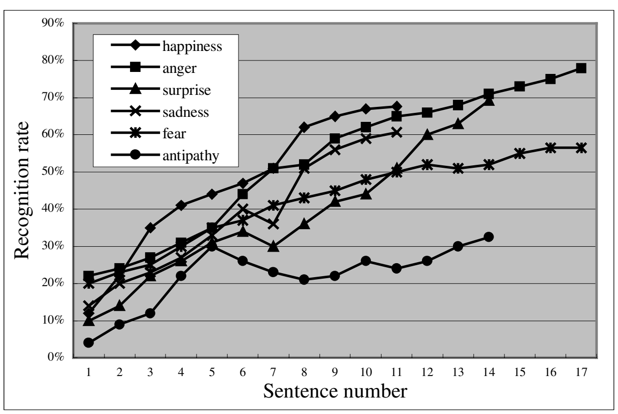

Figure 4: Emotion recognition rate as function of the number of input sentence

The recognition rates listed in Table 4 is calculated over all input sentences. The "Tag" column lists the number of pre-tagged sentence and the "Rec" column shows the number of correctly recognized sentences. This result shows that using this system, the emotion recognition of textual input achieved an accuracy of about 60%.

For further analysis, we analyzed the relationship between the recognition rate and the number of input sentences. Since the dynamic semantic network used in this approach considers the history of the sequential input sentences, the recognition rate increases proportionally to the number of input sentences. Figure 4 shows the recognition rate as a function of the number of input sentences. The x-coordinate indicates the number of input sentence and the y-coordinate indicates the recognition rate. This result presents the ability of the network structure to memorize and learn from the history. When there is only one sentence input, the system can only recognize the emotion by the keywords with no emotion propagation. And when there are more and more input sentences, the emotion recognition rate can be improved by considering the influence of emotional keywords, emotion history, and emotion propagation.

6. CONCLUSION

In this paper, an emotion recognition system with textual input is proposed. An emotional semantic network, consisting of a static semantic network and a dynamic semantic network, is proposed to recognize the emotion from textual input. The static semantic network is constructed from HowNet to generate the emotion trigger value. The dynamic semantic network is constructed from the analysis of Chinese grammatical rules and input sentences. Initiated by the emotion trigger value, the emotion in the dynamic semantic network will propagate and finally converge to give the final emotion output. According to the experimental results, the system can achieve an accuracy of about 60%. This approach is hopefully to be integrated with other physiological characteristics to achieve a more robust emotion recognition.

7. ACKNOWLEDGEMENTS

The authors would like to thank the Computer & Communications Research Labs of Industrial Technology Research Institute, Taiwan, R.O.C., for its financial support of this work, under contract No. T1-90043.

8. REFERENCES

[1] T. Yanaru, "An Emotion Processing System Based on Fuzzy Inference and Subjective Observations," *Artificial Neural Networks and Expert Systems*, 1995. Proceedings., Second New Zealand International Two-Stream Conference.

[2] Samuel W. K. Chan, James Franklin, "Symbolic connectionism in natural language disambiguation," *IEEE Transactions on Neural Networks*, Volume: 9 Issue: 5 , Sept. 1998 Page(s): 739 –755.

[3] W. A. Woods, "Transition network grammars for natural language analysis," *Communications of the ACM* October 1970 Volume 13 Issue 10.

[4] Tsuyoshi Moriyama, Shinji Ozawa, "Emotion Recognition and Synthesis on Speech," *IEEE International Conference, Multimedia Computing and Systems*, 1999.

[5] S. Fukuda, V. Kostov, "Extracting Emotion from Voice," Systems, Man, and Cybernetics, 1999. *IEEE SMC '99 Conference Proceedings*.

[6] Antonio Camurri, Alessandro Coglio, "An architecture for emotional agents," *IEEE Multimedia*, Volume: 5 Issue: 4 , Oct.-Dec. 1998, Page(s): 24 -33.

[7] Joseph Bates, "The role of emotion in believable agents," *Communications of the ACM*, July 1994, Volume 37 Issue 7.

[8] Ryohei Nakatsu, Joy Nicholson, Naoko Tosa, "Emotion recognition and its application to computer agents with spontaneous interactive capabilities," *IEEE 3rd Workshop on Multimedia Signal Processing*, 1999, Page(s): 439 -444.