

# Emotion Estimation and Reasoning Based on Affective Textual Interaction

Chunling Ma<sup>1</sup>, Helmut Prendinger<sup>2</sup>, Mitsuru Ishizuka<sup>1</sup>

<sup>1</sup>Graduate School of Information Science and Technology, University of Tokyo  
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan  
{macl, [ishizuka](mailto:ishizuka@miv.t.u-tokyo.ac.jp)}@miv.t.u-tokyo.ac.jp

<sup>2</sup>National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan  
[helmut@nii.ac.jp](mailto:helmut@nii.ac.jp)

**Abstract.** This paper presents a novel approach to Emotion Estimation that assesses the affective content from textual messages. Our main goals are to detect emotion from chat or other dialogue messages and to employ animated agents capable of the emotional reasoning based on the textual interaction. In this paper, the emotion estimation module is applied to a chat system, where avatars associated with chat partners act out the assessed emotions of messages through multiple modalities, including synthetic speech and associated affective gestures.

## 1. Introduction

An important issue in meeting the needs of the spatially distributed knowledge society is to provide natural and intuitive communication tools. In order to improve textual methods such as e-mail, online chat systems and dialog system, some recent systems are based on like-like embodied agents as a new multi-modal communication means [9]. Most prominently, the BodyChat system [2] employs embodied conversational avatars to mimic human-human face-to-face communication. The TelMeA system [12] uses embodied agents to deliver messages in an asynchronous online community system. Other work employs agents as personal representatives to express the user's point of view of (personal) documents [1].

Although avatars may improve the expressivity of online communication, it remains within the responsibility of the user to carefully prepare the affective content of the textual message. Picard [8] provides a suggestive example: “ ‘How many of you have lost more than a day's work trying to straighten out some confusion over an email note that was received with the wrong tone?’ A majority of hands usually go up when I ask an audience this question. Email is an affect-limited form of communication.” [8, p. 87]

In order to increase the ‘affective bandwidth’ of computer-mediated exchange, the internet community typically uses special ASCII symbol combinations, so-called ‘emoticons’, to express the emotional tone of the message (e.g. “:-)” for “happy”). As a

complementary technique, work on ‘textual affect sensing’ proposes to analyze the textual message itself for affective qualities. In the e-mail composer EmpathyBuddy [5], emotional content of text is processed by an approach based on large-scale real-world knowledge. The assessed emotion is then attached to the textual message in the form of a caricature face that displays the relevant emotion.

The concept presented in this paper can be conceived as an alternative to the EmpathyBuddy system [5]. Our approach is based on the following methods: (i) the affective content of the textual message is recognized by an advanced keyword spotting technique, (ii) syntactical sentence-level processing is applied for detection of affective meaning, and (iii) animated 2D full-body agents perform the emotional coloring of the message using synthetic affective speech and appropriate gestures.

## 2. Calculating Emotion Estimation from Text

The approach for providing emotional estimations for natural-language texts is based on a keyword spotting technique, i.e. the system divides a text into words and performs an emotional estimation for each of these words, as well as a sentence-level processing technique, i.e. the relationship among subject, verb and object is extracted to improve emotion estimation.

The initial step of analyzing an emotional scenario is to define the emotions relevant to the application scenario. In our research, we use the six (basic) emotions from Ekman’s research [3]: *happiness*, *sadness*, *anger*, *fear*, *surprise* and *disgust*.

### 2.1 Affective Words

We employ WordNet-Affect Database [13] of ITS-irst (The Center for Scientific and Technological Research of Autonomous Province of Trento, Italy) with WordNet 1.6 [4] to first find synonyms sets of affective words. We select the emotion “seed” words according to these six emotion categories, i.e., “happy”, “sad”, “angry”, “surprise”, “fear”, “disgusted”, then use key words recognition and synonyms to choose related words from WordNet based on the affective words set. The emotional weight of these words is given based on their *sense* that represents a meaning of a word in WordNet database. Normally, a word has several meanings, but not all the meanings of it are emotional. For example, the verb “beat” has 5 emotional senses among 23 senses in WordNet. So the emotional weight of verb “beat” is 0.22, i.e. the proportion of emotional senses among total senses. The range of emotional weight for each emotion category is [0, 1]. The words are categorized with their part of speech (POS) tag. So we have four affective words database according to WordNet database categories --- *verb*, *adjective*, *adverb* and *noun*. For example, both “joy” and “joyful” belong to “happiness” emotion category, but they are existing in two different database, “noun” and “adjective”. The noun “joy” has its emotion

vector as [1.0 0.0 0.0 0.0 0.0 0.0] while adjective “joyful” has the same emotion vector of [1.0 0.0 0.0 0.0 0.0 0.0]. The examples of affective words extracted from WordNet are shown in Table 1.

**Table 1.** examples of affective words extracted from WordNet (adjective database)

Word	POS	Emotion Vector
exciting	a	[0.76 0.0 0.0 0.0 0.0 0.0]
fortunate	a	[0.72 0.0 0.0 0.0 0.72 0.0]
happy	a	[1.0 0.0 0.0 0.0 0.0 0.0]
healthy	a	[0.66 0.0 0.0 0.0 0.0 0.0]
shaken	a	[0.0 0.0 0.0 0.0 1.0 0.0]
sad	a	[0.0 1.0 0.0 0.0 0.0 0.0]
tearing	a	[0.0 1.0 0.0 0.0 1.0 0.0]
...	.	...

## 2.2 Word-Level Processing

After building the affective lexicon, we use the word spotting technique to estimate emotion of word phrase. The system will calculate the overall emotional estimation for the phrase by summing the extracted estimations. For example, the nouns phrase “a surprise, terrible news” can get the two affective word vectors: “surprise [0.0 0.0 0.0 0.0 1.0 0.0]” and “terrible [0.0 1.0 0.0 0.0 0.0 0.0]”, so the final emotion vector will be [0.0 1.0 0.0 0.0 1.0 0.0]. The emotion of this phrase is “surprise” and “sadness”.

## 2.3 Sentence-level Processing

The word spotting method is too simple to deal with sentences such as “I think that he is happy” or “the lake near my city was very beautiful, but now it is polluted” since here, the speaker is not necessarily happy or the speaker’s emotion depends on the semantic of the sentence deeply. We hence perform the following steps on *sentence-level processing*.

### 2.3.1 Sentence Splitting

In this step, multiple-sentence text is spited into single sentences. Each sentence is estimating the emotion separately.

### 2.3.2 POS (part of speech) Tagging and Sentence Structure Recognition

In this step, the syntactic phrase types are derived out from parse trees that are generated by the parser. The phrase types express the semantic roles in the sentence, for examples, noun phrase (NP), verb phrase (VP) and clause phrase (S or SBAR).

First, we eliminate 'non-emotional' sentences first: (i) Questions, (ii) Clause Phrase beginning with "when" "after" "before" or "if".

Second, we perform the sentence processing. In the top-level of the sentence, the syntactic subject and verb can be derived from noun phrase and verb phrase. Then we recognize some structure patterns of the verb phrase as following types.

- Verb + Adjective Phrase.

If the subject is first person pronouns or with the noun terms that are related to the speaker, we calculate the emotion of Adjective Phrase with the word-level method mentioned in 2.2 above. Otherwise, it is treated as non-emotional. For example, "The book I bought yesterday is very good." In the sentence, "The book" is related to the speaker because "I bought yesterday", so the emotion estimation of adjective phrase "very good" can be applied to the subject and treated as the emotion of speaker. Therefore, the emotion of this sentence is "happiness" with the emotion vector [1.0 0.0 0.0 0.0 0.0 0.0].

- Verb + Noun Phrase.

If the verb exists in the emotion database, after extracting the emotion estimation of the verb, we analyze the noun phrase of the verb phrase: if the noun terms in the noun phrase are related to the speaker, the emotion estimation of the verb will apply to the speaker. Otherwise, it is non-emotional for the speaker. If the verb doesn't exist in the emotion database, we search the knowledge database to get the emotion effect of the verb and the emotion will affect to the speaker if the subject is related to the speaker. The knowledge base is extracted from OMCS (Open Mind Common Sense) knowledge base [7], which includes some affective rules, such as "buy something" is a way to "get something", one of the affect of "get good thing" is "being happy". Therefore, the sentence "I bought a very good book yesterday" can be treated as "happiness".

- Verb + Clause Sentence.

The rules described above are used to analyze the Clause Sentence and the emotion estimation of the Clause sentence is returned back. Then the rules of "Verb + Noun Phrase" pattern are used again. For example, "I think he is good there", "he is good there" will return non-emotional, so apply to "I" and "think", the final result is non-emotional.

### 2.3.3 Negation Detection

In this step, we detect 'negation' in sentences. Since negatively prefixed words such as "unhappy" are already included in the emotion database, they do not have to be considered. On the other hand, negative verb forms such as "have not", "was not", "did not" are detected and flip the polarity of the emotion word.

### 3. Embodied Conversational Messengers

Based on the engine for textual emotion estimation from text, we built a chat system that extracts the emotion from the user's input sentence. In the following, we briefly describe our chat system where animated life-like agents with synthetic speech and gestures serve as user avatars and conversational messengers.

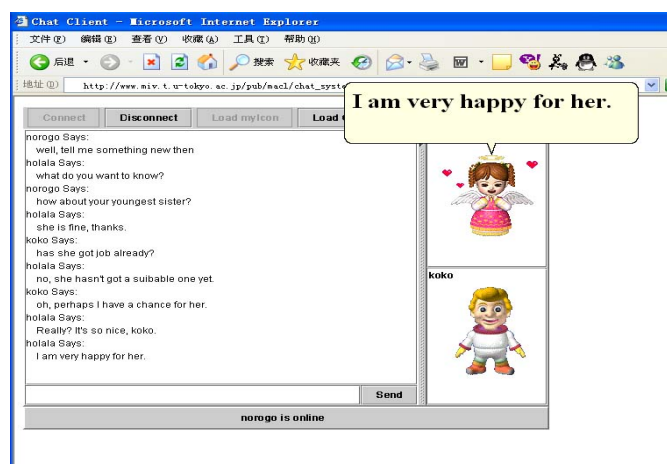


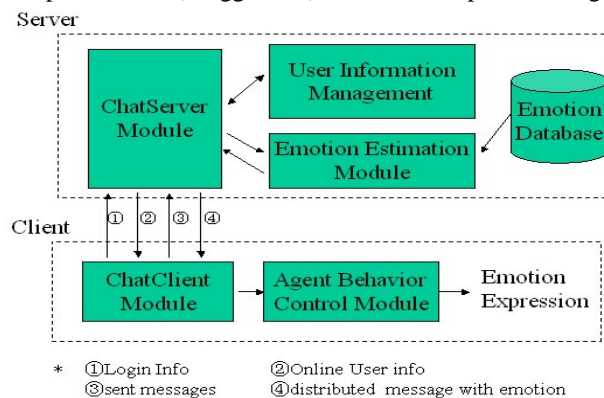
Fig. 1. Chat Client Interface demonstration

Fig. 1 shows an example of a chat client, where three persons are involved in a chatting activity. Among them “norogo” refers to the user as one of the chat clients, the other two (named “holala” and “koko”) are displayed by their own avatar characters. The female character is the avatar of “holala”; the male character is the avatar of “koko”. In this scenario, the following chatting was processing:

*norogo Says(to holala): how about your youngest sister?*  
*holala Says(to norogo): she is fine, thanks.*  
*koko Says(to holala): has she got job already?*  
*holala Says(to koko): no, she has't got a suitable one yet.*  
*koko Says(to holala): oh, perhaps I have a chance for her.*  
*holala Says(to koko): Really? It's so nice, koko.*  
*holala Says(to koko): I am very happy for her.*

When the chat partner called “holala” types the message “I am very happy for her”, her avatar character expresses the “gladness” emotion. The relevant emotion word in the sentence is the word “happy”, which is defined as “gladness” in the emotion database.

The words “very” add to the intensity of the emotion conveyed by “happy”. The emotional content of the message “I am very happy for her.” is expressed through the avatar by synthetic speech and a (exaggerated) non-verbal expression of gladness.



**Fig. 2.** Architecture of the chat system

The architecture of the chat system is depicted in Fig. 2. On the server side, the ChatServer module is used to listen to the clients' connection and incoming messages. The Emotion Estimation module analyzes the emotion tendency of the incoming messages and returns the result back to the ChatServer Module. In this system, emotion detection is context independent, i.e. we currently do not consider the current mood or emotion of a chat partner. Once the emotion of a message is estimated, the avatar will perform the message with affective expression. The analysis of emotion is based on an emotion database and the algorithm mentioned in Sect. 2. The chat system has been implemented using the Java platform, JavaScript, the Microsoft Agent package [6], and Stanford javaNLP API [10]. The behavior sequence and messages to be spoken out are sent to JavaScript functions as parameters. Currently we use a light client design in the system; i.e. the client side essentially sends connection requests to the server module, and sends or receives messages. When the connection is established, the client will request the server to send its information, e.g. to a particular agent character. The server will update its online list of the clients for each client. Then the user can choose the chat user to talk with in the animated style. After a chat user is chosen, the Agent Behavior Control module is called to generate a behavior expression corresponding to the emotion estimation value. On the server side, we also maintain a user information database.

#### 4. Conclusions and Future Work

This paper presents an Emotion Estimation approach based on textual interaction. In our approach, a word spotting technique and syntactical sentence-level processing are

proposed. The issue of emotional reasoning based on affective text has been studied and implemented in different contexts and approaches [5,11,13], which include the use of affect dictionaries, simple natural language processing techniques, and commonsense knowledge of real world. Our approach is based on the affect lexicon and simple rules of commonsense knowledge with sentence-level processing. The difference of our approach to other studies is that our aim is to decide the emotional content of the online textual messages in real time, so that we may trace the relationship between speaker and conversational content. Emotion assessment is based on the role of a verb in a sentence. The emotion reasoning procedure is bottom-up, i.e. the word spotting technique is used to decide the emotion of basic phrase type, and then the distilled emotion category is transferred to the upper layer phrase. This approach allows for more efficient reasoning than approaches based on the real-world knowledge, and we can get better results than by using a simple word spotting approach.

In our future work, we plan to improve the Emotion Estimation module, specifically by combining past emotional states as a parameter for deciding the affective meaning of the user's current message. In addition, the speaker's mood will be associated with the topic of the conversation. In this way, we hope to obtain a better identification of the relation between speaker and topic terms in the sentence.

## References

1. Bickmore, T.W., (ed.): Animated autonomous personal representatives. In Proceedings 2nd International Conference on Autonomous Agents (Agent-98), Minneapolis, MN (1998) 8-15
2. Cassell, J., Vilhjálmsdóttir, H.: Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems* (1999) 2:45–64
3. Ekman, P.: Facial Expression and Emotion. *American Psychologist* (1993) 48, 384-392
4. Fellbaum C.: *WordNet: An Electronic Lexical Database*. MIT Press. (1982)
5. Liu, H., Lieberman, H., Selker, T.: A model of textual affect sensing using real-world Knowledge. In Proceedings International Conference on Intelligent User Interfaces (IUI-03) (2003) 125–132
6. Microsoft. *Developing for Microsoft Agent*. Microsoft Press. Redmond, WA (1998)
7. MIT Media Lab: *Open Mind Common Sense* (2005) URL: <http://commonsense.media.mit.edu/>
8. Picard., R. W.: *Affective Computing*. The MIT Press, Cambridge, MA (1997)
9. Prendinger, H., Ishizuka, M.: editors. *Life-Like Characters.Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg (2004)
10. Stanford NLP Group (2005) URL: <http://nlp.stanford.edu>
11. Subasic, P. and Huettner, A.: Affect Analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems* (2001) 9: 483-496
12. Takahashi, T. (ed.): *TelMeA—Expressive avatars in asynchronous communications*. *International Journal of Human-Computer Studies* (2005) 62:193–209
13. Valitutti, A., Strapparava, C., Stock, O.: *Developing Affective Lexical Resources*. *PsychNology Journal*. (2004) Volume 2, Number 1, 61-83