A Chat System Based on Emotion Estimation from Text and Embodied Conversational Messengers

(Preliminary Report)

Chunling Ma
Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8656, Japan
macl@miv.t.u-tokyo.ac.jp

Alexander Osherenko
Institut für Informatik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin, Germany
osherenko@gmx.de

Helmut Prendinger
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
Tokyo 101-8430, Japan
helmut@nii.ac.jp

Mitsuru Ishizuka
Graduate School of Information Science and Technology, University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8656, Japan
ishizuka@miv.t.u-tokyo.ac.jp

Abstract

This short paper contains a preliminary description of a novel type of chat system that aims at realizing natural and social communication between distant communication partners. The system is based on an Emotion Estimation module that assesses the affective content of textual messages. Avatars associated with chat partners act out the assessed emotions of messages through multiple modalities, including synthetic speech and affect-related gestures.

1. Introduction and Motivation

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 and online chat systems, some recent systems employ embodied (like-like) agents [7] as a new multi-modal communication means.

Most prominently, the BodyChat system [2] employs embodied conversational avatars to mimic human–human face-to-face communication. The TelMeA system [8] 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 online communication, it remains within the responsibility of the user to carefully prepare the affective content of the (textual) message. A suggestive example is provided by Picard [6, p. 87]:

> "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.

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 [3], 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 short paper can be conceived as an alternative to the EmpathyBuddy system [3]. In our approach, (i) the affective textual content is recognized by an advanced keyword spotting technique, and (ii) an animated 2D agent (rather than a face caricature) per-

forms 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 the keyword spotting technique, i.e. the system divides a text into words and processes emotional estimations for these words (see [5] for an extensive discussion of this approach).

In the initial step of analyzing an emotional scenario, the system designer defines the emotions relevant to the application. In the chat application, emotions could be "gladness" and "sorrow" emotions. Then the designer adjusts the system database correspondingly, by including the input of possible emotional words and their emotional estimations (see Formula 1).

$$\vec{\epsilon}_w = \begin{bmatrix} c_{1_w} \\ \vdots \\ c_{n_w} \end{bmatrix} \quad (1)$$

where $\vec{\epsilon}_w$ is the emotional estimation of word w, the vector elements c_{1_w}, ... c_{n_w} are the empirical values entered in the database for the emotions $1...n$.

After emotional estimations are entered, the system can calculate the overall emotional estimation for the given text by summing the extracted estimations (see Formula 2).

$$\vec{E}_t = \sum_{i=1}^{n} \vec{\epsilon}_i \quad (2)$$

where $\vec{E}_t$ is the emotional estimation of an analyzed text, $\vec{\epsilon}_i$ is the emotional estimation of word i in the text, and n is the number of words in the text.

The $\vec{E}_t$ vector can be used for emotional appraisal. For instance, in case of two emotions ($n = 2$) the system calculates the ratio of the first to the second vector element and deduces from it a certain affective state (see Formula 3).

$$Appraisal = \begin{cases} gladness & \text{if } \frac{e_1}{e_2} \geq 1.7644; \\ sorrow & \text{otherwise.} \end{cases} \quad (3)$$

In this example the significant emotions are gladness and sorrow, $\vec{E}_t = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ is the emotional estimation calculated by the system for a given text according to Formula 2, value 1.7644 is an experimentally revealed constant that provides maximal success rate.

3. Embodied Conversational Messengers

Based on the engine for emotion estimation from text discussed in the previous section, we built a chat system that

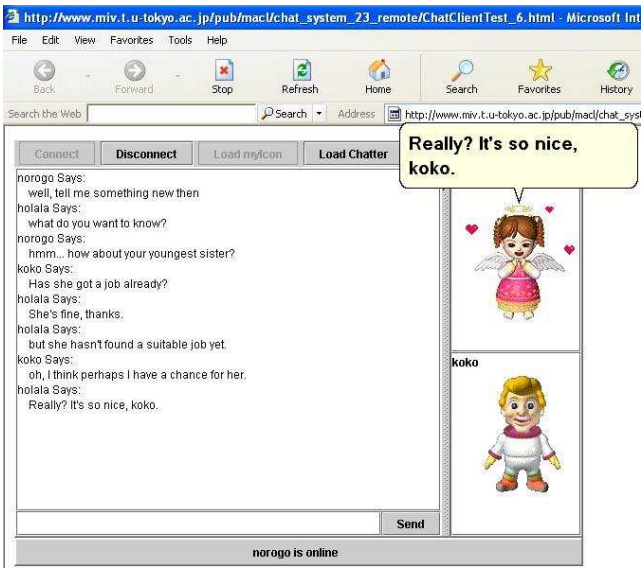

Figure 1. Chat client.

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. Figure 1 shows an example of a chat client.

In Fig. 1, there are three persons involved in a chatting activity. Among them, "norogo" refers to the user as one chat client, the other two (named "halala" and "koko") are displayed by their own avatar characters. When the chat partner called "halala" types the message "Really? It's so nice, koko.", her avatar character expresses the "gladness" emotion. The message consists of two sentences: "Really?" and "It's so nice, koko.". The relevant emotion word in those sentences is only the word "nice", which is defined as "gladness" in the emotion database. The words "really" and "so" add to the intensity of the emotion conveyed by "nice". Hence, the emotional content of the message "Really? It's so nice, koko." is expressed through the avatar by synthetic speech and an (exaggerated) non-verbal expression of gladness.

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. The analysis of emotion is based on an emotion database and the algorithm discussed in Sect. 2.

The chat system has been implemented using the Java platform, JavaScript, and the Microsoft Agent package [4]. The behaviors sequence and messages to be spoken out are transferred to JavaScript functions as parameters. Cur-

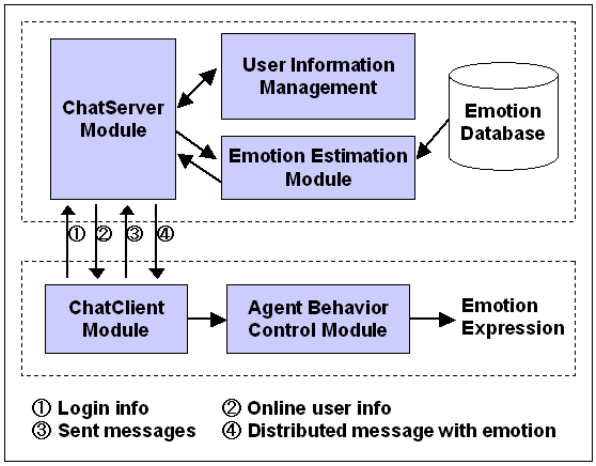

Figure 2. Architecture of the chat system.

rently we use a light client design in the system; that is, 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. Future Work

In our future work, we plan to improve the Emotion Estimation module, e.g. by integrating the recorded user (client) information into the analysis of emotions. Most importantly, past emotional states could be important parameters for deciding the affective meaning of the user's current message.

Acknowledgements

This research is partly supported by the JSPS Research Grant (1999-2003) for the Future Program.

References

[1] T. Bickmore, L. K. Cook, E. F. Churchill, and J. W. Sullivan. Animated autonomous personal representatives. In *Proceedings 2nd International Conference on Autonomous Agents (Agents-98)*, pages 8–15, Minneapolis, MN, 1998.

[2] J. Cassell and H. Vilhjálmsson. Fully embodied conversational avatars: Making communicative behaviors autonomous. *Autonomous Agents and Multi-Agent Systems*, 2:45–64, 1999.

[3] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings International Conference on Intelligent User Interfaces (IUI-03)*, pages 125–132, 2003.

[4] Microsoft. *Developing for Microsoft Agent*. Microsoft Press, Redmond, WA, 1998.

[5] A. Osherenko. *Modeling Emotions Using a Shallow Natural-Language Processing Technique*. Humboldt University Berlin, Institute of Informatics, 2004. Master's thesis.

[6] R. W. Picard. *Affective Computing*. The MIT Press, 1997.

[7] H. Prendinger and M. Ishizuka, editors. *Life-Like Characters. Tools, Affective Functions, and Applications*. Cognitive Technologies. Springer Verlag, Berlin Heidelberg, 2004.

[8] T. Takahashi, C. Bartneck, Y. Katagiri, and N. H. Arai. TelMeA—Expressive avatars in asynchronous communications. *International Journal of Human-Computer Studies*, 62:193–209, 2005.