

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

Chunling Ma¹, Helmut Prendinger², Mitsuru Ishizuka¹

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

² National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
helmut@nii.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 chat messages and avatars associated with chat partners that 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 and online chat systems, 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 [11] 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 (see [7] for an extensive discussion of this approach), 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 the chat system, we use the six (basic) emotions from Ekman's research [3]: *happiness*, *sadness*, *anger*, *fear*, *surprise* and *disgust*. We employ WordNet-Affect Database [12] 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 these six emotion categories and to assess their emotional weight, and then compute the weight of a sentence by combining the weights of its parts. However, the word spotting method is too simple to deal with sentences such as "I think that he is happy" since here, the speaker is not necessarily happy.

We hence perform the following two types of *sentence-level processing*. First, we eliminate 'non-emotional' sentences: (i) sentences without emotional words, (ii) questions, (iii) sentences without first person pronouns (as the example sentence above). Second, 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.

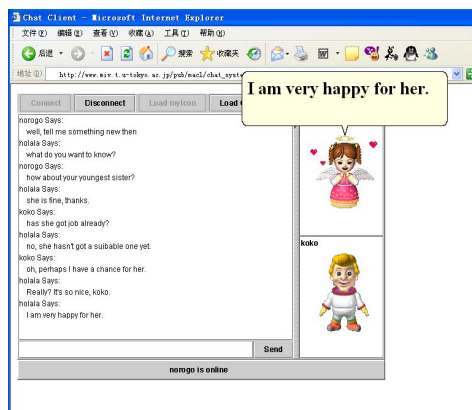


Fig. 1. Chat Client

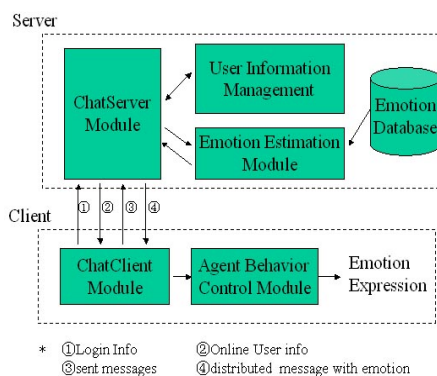


Fig. 2. Architecture of the chat system

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 chat client, the other two (named “holala” and “koko”) are displayed by their own avatar characters. 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.

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 our current 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; 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, 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., Villhj'almsson, 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. Osherenko., A.: *Modeling Emotions Using a Shallow Natural Language Processing Technique*. Humboldt University Berlin, Institute of Informatics (2004) Master's thesis.
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. Takahashi, T. (ed.): TelMeA—Expressive avatars in asynchronous communications. *International Journal of Human-Computer Studies* (2005) 62:193-209
12. Valitutti, A., Strapparava, C., Stock, O.: Developing Affective Lexical Resources. *PsychNology Journal*. (2004) Volume 2, Number 1, 61-83