A Multi-modal Eliza Using Natural Language Processing and Emotion Recognition

Siska Fitrianie, Pascal Wiggers, and Leon J.M. Rothkrantz

Data and Knowledge Systems
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
p.wiggers@its.tudelft.nl
l.j.m.rothkrantz@its.tudelft.nl

Abstract. In the human machine interaction domain adaptive life-like agents are becoming a popular interface. In order to provide a natural conversation such agents should be able to display emotion and to recognize the user's emotions. This paper describes a computer model for a multi-modal communication system based on the famous Eliza question-answering system. A human user can communicate with the developed system using typed natural language. The system will reply with text-prompts and appropriate facial-expressions.

1 Introduction

Recently, there has been a lot of interest in adaptive life-like agents in the area of human-computer interaction. Examples include the German Smartkom project [1] and the CSLR reading tutor [2]. The advantages of such systems are obvious, they offer a much more natural conversation with a machine than traditional user interfaces taking human face-to-face communication as their source of inspiration. This is especially the case when interaction through multiple modalities including speech and pointing is supported. However, in human-to-human communication emotions play an important role. As indicated by Mehrebian [3] about 55 percent of the emotional meaning of a message is communicated through the non-verbal channel, which includes gestures, postures and facial expressions. Thus, in order to offer a natural interface software agents should also show the proper non-verbal reactions, like facial expressions. On the input side this requires of course recognition of the user's emotions.

In this paper we describe the design and implementation of a multi-modal question-answering system based on the famous Eliza program [4], which simulates a psychoanalyst who talks to a client using natural language. Our system recognizes emotion from natural language. It will show a facial expression for each sentence typed by the user. Subsequently, it will give a natural language reply together with an appropriate facial expression to convey emotional content.

2 System Architecture

The overall system architecture is based on the idea of message passing on a blackboard, as is illustrated in Fig. 1. All processing modules act as independent experts by taking

V. Matoušek and P. Mautner (Eds.): TSD 2003, LNAI 2807, pp. 394–399, 2003.

their input from and writing their output back to the blackboard. This ensures a flexible data-controlled architecture where modules may be executed in parallel. Modules can easily be added or changed, which also opens up the possibility of adding more modalities to the system. The input to the system is a user's text string and the results are the reply sentences and facial displays. The processing modules in the system can be subdivided into two layers, the first layer consists of modules for natural language processing and the second layer performs emotional recognition in order to construct facial displays. The next two sections will describe these layers in more detail.

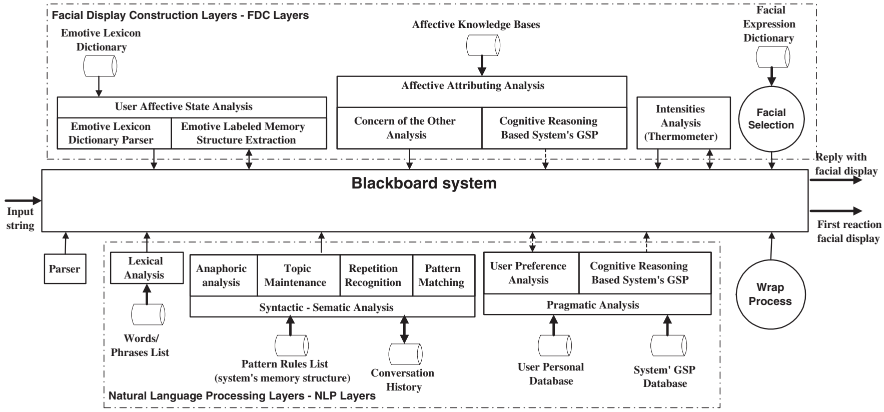

Fig. 1. System architecture.

3 Natural Language Processing

The natural language processing layer analyses a user's input string in order to construct a reply sentence. First, the *Parser* subdivides the string in a list of words. This list is the input to the *Lexical Analysis Module*, which checks the spelling, replaces abbreviations, slang words and codes with full words and uses a thesaurus to replace certain words by their synonyms to reduce the amount of variation subsequent modules have to deal with.

The *syntactic-semantic* analyzer performs shallow parsing by matching the input words with predefined patterns (that may contain wildcards) and composes a reply from the keywords found using reassemble rules attached to these patterns. For a given prompt the longest matching pattern containing the smallest number of wild cards is chosen. Preceding the pattern matching step the module performs a number of actions to ensure a more natural conversation. Anaphoric analysis guarantees that the system responds consistently on prompts by referring to the preceding conversation content. Repetition recognition makes sure that the dialog never gets into a loop and the system tries to stay within the current topic of conversation. The rules used are written in an extended-XML script specification called AIML (Artificial Intelligence Markup Language), defined by Wallace [5].

The final natural language module is the *Pragmatic Analysis*, which checks the reply composed by the previous module against the user preferences that are collected during the conversation and against the goals, states and preferences of the system. As the system simulates a psychoanalyst its main goal is to keep the conversation going. By distinguishing a dialog state as a certain dialog act like a question, statement, acknowledgment, or pause, the system determines which intermediate goal to pursue, for example answering a question, asking a user for explanation or reflecting a feeling. If a reply does not comply with the system goals it is rejected and the syntactic-semantic module is invoked again to formulate a new reply.

4 Emotion Recognition

How many and what kind of emotional expressions are to be used poses a non-trivial question. In this work we adopted the twenty-four categories of emotions defined by Ortany, Clore and Colling (OCC's theory [6, 7]). They are based on grouping human emotions by their eliciting conditions events, the consequences of their action, and the selections of computational implementation. Since classifications of some emotion eliciting factors are in a gray area, in this research, we add one emotion type: uncertainty. However, for emotion recognition our current prototype uses only the 7 universal emotions defined by Ekman [9]. These are shown in Table 1, together with the corresponding OCC emotions.

Table 1. Emotions.

Universal	OCC theory
Neutrality	Normal
Happiness	Joy, Happy-for, Gloating, Satisfaction, Relief, Pride, Admiration, Liking, Gratitude, Gratification
Sadness	Distress, Resentment, Sorry-for, Disappointment, Shame, Remorse
Disgust	Disliking, Hate
Surprise	Hope
Fear	Fear, Fears-confirmed
Anger	Reproach, Anger
Uncertainty	Uncertainty

4.1 Emotion Eliciting Factor Extraction

To extract emotion eliciting factors from the text prompts in the conversation both of the parsed user's string input and the systems reply sentence a shallow word matching parser, called *Emotive Lexicon Look-up Parser*, is used that utilizes a lexicon of words having 'emotional content' for each of the seven universal emotions. In total a list of 247 words was used compiled from three sets of emotional words described by [10–12]. For each of these words a natural number intensity value is given. To get the overall emotional content of the string a thermometer is defined for each of the seven emotions. When an emotionally rich word is found the thermometers are updated by:

$$\begin{aligned} &T_{i(t)} = T_{i(t-1)} + I_i.s \\ &\forall j \neq i \cdot T_{j(t)} = T_{j(t-1)} - distance[j,i] \end{aligned} \tag{1}$$

Where, i is the active emotion type, s is a summation factor; I is the emotion intensity and j ranges over all universal emotions defined in Table 1. The distance between two emotions follows from the work of Hendrix and Ruttkay [13] who defined the distance values shown in Table 2.

Table 2. Distance values between emotions.

	Happiness	Surprise	Anger	Disgust	Sadness
Happiness	0	3.195	2.637	1.926	2.554
Surprise		0	3.436	2.298	2.084
Anger			0	1.506	1.645
Disgust				0	1.040
Sadness					0

Each of the memory structures, that is a pattern and the corresponding rules, used by the syntactic-semantic modules is labeled with one or more emotion types in the *Emotive Labeled Memory Structure Extraction*. We achieved this by adding two additional tags in the AIML scheme. The `<affect>` tag that labels the user's affective situation and the `<concern>` tag that labels the system's reaction situation. Inside those two new tags, we define four emotive situation types: positive, negative, joking and normal/any.

Whether a certain goal, found during pragmatic analysis is appealing influences the system's affective state. As do the preferences defined for the system. The *Goal-Based Emotion Reasoning* also stores the user's personal data during conversation, e.g. name, birthday, favorite things and so on.

To determine the system's affective state two knowledge bases are used. One to determine the system's reaction affective state as stimulus response to the user's input string and on to determine the system's reaction affective state as the result of the cognitive process of the conversation content to convey its reply sentence. We have defined a set of so-called preference rules that specify the emotion recognition process of the system. Every rule in the set defines conditions of emotion eliciting factors and the affective thermometers to activate the rule and a preference that is expressed upon activation. The result from each knowledge-based system is one of twenty-four OCC's theory emotion types with addition of two emotion types: normal and uncertainty.

Fig. 2 shows two example preference rules. The first rule is a stimulus response preference rule for the reaction of joy. In this case the system will answer any questions from the user joyfully, because she enjoys the situation and she met the goal: making the user feel happy. The second rule is part of the cognitive process knowledge base. Here the system does not like the user making a joke while it feels sad.

```
IF (user is happy) AND (user asks question) AND (systems reply is sad) AND
(situation type of user is not negative) AND (highest thermo is happy) THEN
reaction is joy.

IF (user is sad) AND (systems reply is sad) AND (situation type of user is
joking) AND (situat ion type of the system is negative) AND (maximum affective
thermo is sad) THEN reply is resentment
```

Fig. 2. Preference rules.

4.2 Facial Display Selection

For the activation of an emotion, [6, 8] proposed the use of threshold values by counting all associated elicitation factors, excitatory (positive) and inhibitory (negative), from other emotions. They used an activation level range $[0, max]$ where max is an integer value determined empirically. All emotions are always active, but their intensity must exceed a threshold level before they are expressed externally. The activation process is controlled by a knowledge-based system that synthesizes and generates cognitive-related emotions in the system. To determine the intensity of the systems emotions as a reaction to the user's string input and the dialog content we define six affective thermometers classified by six Ekman's universal emotion types (neutrality is not considered here) as we did for the user emotions in the emotive lexicon. If an emotion is active, the system calculates all of thermometers T_i according to equation (1) given in the previous section.

The thermometer having the highest value is chosen as the systems emotion and depending on the intensity a facial display is chosen. Currently, the mapping from emotions to facial expressions is one-to-one where the emotions correspond to the 24 OCC emotions, uncertainty or neutrality.

5 Implementation and Future Work

Currently a web-based client server prototype of the model has been implemented for experimental purposes. The server provides the blackboard architecture, implemented in JESS, which is accessible over TCP/IP by the client application. Currently, the emotive lexicon contains: 48 lexemes for happiness, 170 lexemes for sadness, 34 lexemes for surprise, 33 lexemes for fear, 93 lexemes for disgust, and 69 lexemes for anger. This prototype has 1953 categories in its list of pattern rules. Its affective knowledge base contains 77 preference rules of stimulus response and 151 preference rules for the systems affective state. We can add new rules to these databases and knowledge bases while the server is still running.

The next step will be to extend the system with a speech interface instead of the typed text interface currently used. The emotions can then be extracted by shallow parsing from the spoken words and be combined with emotions deduced from prosodic cues to get a more accurate indication. Furthermore, the static facial displays used in the prototype will be replaced by a 3D animated talking face that is currently under development within our group [14].

References

1. Wahlster, W., Reithinger, N., Blocher, A.: SmartKom: Multimodal Communication with a Life-Like Character, Proceedings of Eurospeech 2001, Aalborg Denmark, 2001
2. Ma, J., Yan, J., Cole, R.: CU Animate Tools for Enabling Conversations with Animated Characters, Proceeding of ICSLP 2002, Denver, CO USA, September 2002.
3. King, Donnel: Nonverbal communication, http://www2.pstcc.cc.tn.us/ dking, 1997.
4. Weizenbaum, J.: ELIZA - A Computer Program for the Study of Natural Language Communication between Man and Machine, Communication of the ACM 9(1): p36-p45, 1966.
5. Wallace, Richard: Alicebot, http://www.Alicebot.org, 1995.
6. Bazzan, Bordini: A Framework for the Simulation of Agents with Emotions, Report on Experiments with the Iterated Prisoner's Dilemma, AGENT'01, Communication of ACM, p292-p299, Canada, 2001.
7. Elliott, Siegle: Variables Influencing the Intensity of Simulated Affective States, In AAAI Technical Report for Spring Symposium on Reasoning about Mental States: Formal Theories and Applications, 58-67, American Association for Artificial Intelligence, 1993.
8. Elliott: Using the Affective Reasoner to Support Social Simulations, In Proceeding of the 13th International Joint Conference on Artificial Intelligence, 194-200, Chambery, 1993.
9. Ekman, Friesen: Unmasking the Face, Prentice Hall, New Jersey, USA, 1975.
10. Davitz: the language of emotions, New York, Academic Press, 1969.
11. Frijda, N.H: The emotions, Cambridge University Press, Cambridge.
12. Fehr, B., Russel, J.A: Concept of emotion viewed from a prototype pespective, In journal of experimental psychology, 113, 464-486.
13. Hendix, Ruttkay: Exploring the Space of Emotional Faces of Subjects without Acting Experience, ACM Computing Classification System: H.5.2, I.5.3, J.4, 1998.
14. Wojdel, A., Rothkrantz, L.J.M.: A text based talking face, in proceedings of TSD 2000.