Multi-humoroid: Joking System That Reacts With Humor To Humans' Bad Moods (Extended Abstract)

Pawel Dybala
Hokkaido University
Kita 14 Nishi 9, Kita-ku, 060-0814 Sapporo,Japan
tel. +81-11-706-7389
paweldybala@media.eng.hokudai.ac.jp

Michal Ptaszynski
Hokkaido University
Kita 14 Nishi 9, Kita-ku, 060-0814 Sapporo,Japan
tel. +81-11-706-7389
ptaszynski@media.eng.hokudai.ac.jp

Rafal Rzepka
Hokkaido University
Kita 14 Nishi 9, Kita-ku, 060-0814 Sapporo,Japan
tel. +81-11-706-7389
kabura@media.eng.hokudai.ac.jp

Kenji Araki
Hokkaido University
Kita 14 Nishi 9, Kita-ku, 060-0814 Sapporo,Japan
tel. +81-11-706-7389
araki@media.eng.hokudai.ac.jp

ABSTRACT

This paper contributes to the field of humoroids – humor-equipped conversational agents. We present our multi-agent system (multi-humoroid), which tells jokes according to users' emotions, in order to make them feel better. We briefly describe the components and the design of the system, and present results of two experiments showing that the system with humor was evaluated as generally better than a baseline dialogue agent.

Categories and Subject Descriptors

I.2.11 [Artificial Intelligence] Distributed Artificial Intelligence - *intelligent agents*

I.2.7 [Artificial Intelligence] Natural Language Processing – *language generation*

General Terms

Performance, Human Factors

Keywords

Description: experimental; Inspiration: Artificial Intelligence; NLP; Focus: Agent-Human Interaction. Other: humor processing, conversational agents

1. INTRODUCTION

During AAMAS'09, we proposed to define a new class of conversational agents – "humoroids", agents able to use humor while talking with humans [1]. We presented our own simple humoroid for Japanese, called Pundalin, which tells puns at every third turn of conversation. The evaluation experiments showed that even such a simple humor-equipped agent was evaluated as generally better that a similar one without humor. The agent, however, had several flaws, one of which was artificial timing of jokes (told at every third turn of conversation).

This paper is a contribution to this field, and describes the next stage in research on humoroids. By combining a conversational agent, a humoroid and an emotiveness analysis agent, we created a "multi-humoroid" - multiagent joking system, able to use humor in conversation according to interlocutors' emotional states. This is important in several ways: first, we acquired a system that makes an effort to make users feel better (in an "intentional" way), and second, it is an improvement in terms of timing of jokes, which, instead of using the every-third-turn rule, is now based on the analysis of users' emotions.

2. OUR AGENTS

In this research we used three different agents: Maru-chan, a baseline chatterbot (see 2.1), Pundalin, a pun generator (see 2.2), and ML-Ask, an emotiveness analysis agent (see 2.3). They were combined to construct MAS-Punda - a multi-agent joke telling conversational system (multi-humoroid), able to tell jokes according to users' feelings (see 2.4).

2.1 Baseline chatterbot

The first agent in our research is Maru-chan, a keyword-based conversational agent for Japanese, developed by Takahashi [2]. The agent (below referred to as "baseline chatterbot") uses the Internet as a source of linguistic knowledge, in order to extract word associations for users' utterances, and then uses them to generate a relevant response. For more details, see [2].

2.2 Pun generator

The second agent used in this research is Pundalin, a pun generator for Japanese [3], which also uses the Internet. The input is a user's utterance, from which the agent selects a pun base word (a word that will be transformed into a pun). Then, for the base word, the agent generates a list of phonetic candidates, which are next queried in the Goo Internet search engine. The candidate with the highest hit rate is selected and integrated into an input-related response. For more details, see [3].

2.3 Emotion detector

The third agent used in our research is Ptaszynski et al.'s ML-Ask Emotive Elements/Emotive Expressions Analysis Agent [4]. Besides performing the decisive role in the multi-humoroid, it is also used in automatic evaluation of the system (see 3.2).

The agent uses databases, containing emotive expressions and emotive elements for Japanese. It performs three types of text analysis: 1) it detects if a sentence is emotive (if it contains emotive elements, like interjections, exclamations, or vulgar language), 2) it determines the types of conveyed emotions, and 3) it determines whether the emotions are positive / negative and activated / deactivated. In the second step, when no emotive expressions from the database are found in the input, the agent employs Shi et al.'s [5] web mining technique to extract emotive associations for the sentence. For more details, see [4] and [5].

2.4 Multi-humoroid

The three agents presented in 2 were merged to create a "multi-humoroid" – multiagent humor-equipped joking conversational system. Basing on existing research and literature (summarized in [6]), we assumed that in an interaction between humans and computer agents, the latter can use humor to make the interlocutor feel better. In order to do that, first it has to detect the human's emotions (this is performed by ML-Ask, without the web mining procedure, as it was too time-consuming), and then on this basis make a decision whether a joke should be told.

To summarize the decision rules used in this experiment, the assumptions were that: 1) if a human's emotive state is negative, the agent can use humor in order to make him / her feel better, and 2) if a human's emotive state is neutral (non-emotive), humor can be used to induce positive emotions.

Cite as: Multi-humoroid: Joking System That Reacts With Humor To Humans' Bad Moods (Extended Abstract), Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, Kenji Araki, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1433-1434

If ML-Ask decides that a joke should be told, the response is generated by the joke generator. If ML-Ask decides otherwise, the response is generated by the baseline chatterbot.

3. EVALUATION EXPERIMENTS

The performance of the multi-humoroid was evaluated in two experiments: first person oriented and automatic.

3.1 First person oriented

The first experiment in our study was first person (user) oriented. We asked 13 university students (age 21-30) to perform two conversations: one with the multi-humoroid and one with the base chatterbot, in a randomized order. No topic restrictions were made. All conversations were text-based. The interactions were conducted continuously, one immediately after another.

The questions included in the questionnaire were:

A) Did you get an impression that the agent was human-like?
B) Did the agent try to make the conversation more interesting?
C) Did you find the conversation interesting?
D) Did the agent try to make your feelings better / more positive?
E) Did the agent use humor in appropriate moments?
F) Describe your feelings towards the agent after the interaction
G) Which agent would you choose for a friend?
H) Which agent do you think was better?

Answers for questions A-E were given in 5-point scales. For question F, the evaluators could answer freely. For questions G and H, the evaluators had to choose between the two agents.

As only one of the agents used humor (intentionally), answers for question E, which was directly related to the issue of humor, also included the option "the agent did not use humor".

Answers for question F were compared with our emotive expressions database, to check the valence of each emotion listed by users. Every positive emotion counted as +1, and every negative as -1.

As we wanted to investigate the users' will to continue the interaction, after 10 turns of conversation the agents asked the experiment participant if he/she wanted to continue the conversation. If the answer was "yes", another 5 turns of conversation were performed. After that, the dialogue ended automatically (similarly if the answer was "no").

The statistical significance of the results was calculated. For questions A-D, we used the Two Paired Sample Wilcoxon Signed Rank Test (as the data was paired and not distributed normally). For the question E, we used the Mann–Whitney U test (as the data was unpaired and not distributed normally).

The results are summarized in Table 1.

Table 1. User-oriented evaluation experiment results

quest.	A	B	C	D	E	F	G	H	cont?
baseline	1.85	1.69	2.08	1.69	1.00	- 9	31%	38%	2
multi	2.15	2.85	2.69	2.69	2.45	+8	69%	62%	5
diff.	0.30	1.16	0.61	1.00	1.45				
P val.	0.090	0.006	0.026	0.007	0.001				

3.2 Automatic (emotive analysis based)

The chat logs acquired in the user experiment were next analyzed with ML-Ask (this time with the web mining procedure). Results of the analysis allowed us to compare the dialogues of the two systems (with and without humor) in three dimensions: 1) general emotiveness, 2) valence changes (positivity/negativity), 3) arousal level (activation/deactivation). We also analyzed users' final emotions in conversation, to investigate what mood they finished the interaction in – see Table 2 and Figure 1.

Table 2. Automatic evaluation experiment – results

	baseline			multi-humoroid		
emotiveness	91 (average: 7.0 per utt.)			125 (average 9.6 per utt.)		
val. changes	to positive	to negative		to positive	to negative	
	68%	32%		94%	6%	
last emotion	pos.	neg.	neutr.	pos.	neg.	neutr
	69%	31%	0%	85%	0%	15%

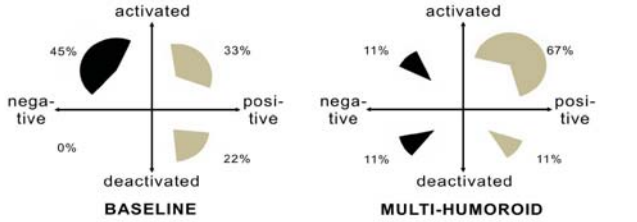

Figure 1. Results of analyzing users' emotion types

4. SUMMARY AND CONCLUSIONS

The results presented above show that the multi-humoroid was evaluated as generally better that the baseline chatterbot, in both user-oriented and automatic evaluation experiments. Although not all differences were outstandingly large or significant (e.g. question A in the user experiment), we can say that there is a general tendency to appreciate the joking system more than the one without humor (see questions G and H). The level of interest was higher for the multi-humoroid (questions B and C), and users' feelings toward it were generally more positive and activated than towards the other system (see questions D, F and results of the automatic experiment). It is also worth mentioning that users' final emotions in the conversations were generally more positive for the multi-humoroid. Finally, we can see that the timing of humor (question E) was also evaluated relatively well.

To conclude, the results are fairly satisfying, but they could be improved. Our plans for the future include further specification of the timing algorithm (i.e. checking what particular emotions can be followed by humor) and individualization of the system's humor sense (adapting to users' preferences).

5. REFERENCES

[1] Dybala, P., Ptaszynski, M., Higuchi, S., Rzepka, R., and Araki, K. 2009. Humoroids – Conversational Agents That Induce Positive Emotions with Humor, In Proc. of AAMAS 09, Budapest, Hungary, 1171-1172.

[2] Takahashi, M. 2009. *Web ni yoru kyoukihindo to N-gram moderu wo mochiita hatsuwabunnseiseishuhou* (Utterance Generation Method Using Web Search Results and Word N-grams), Bachelor dissertation, Hokkaido University, Sapporo, Japan.

[3] Dybala, P., Ptaszynski, M., Higuchi, S., Rzepka, R., and Araki, K. 2008. Humor Prevails! - Implementing a Joke Generator into a Conversational System. In Proc. of AI-08, Wobcke, W. and Zhang, M. (eds), Auckland, New Zealand, 2008. LNAI 5360, 214-225, Springer Berlin & Heidelberg.

[4] Ptaszynski, M., Dybala, P., Higuchi, S., Rzepka, R., and Araki, K. 2008. Affect as Information about Users' Attitudes to Conversational Agents, In Proc. of WI-IAT'08, HAI'08, Sydney, Australia, 459-500.

[5] Shi, W., Rzepka, R., and Araki, K. 2008. User Textual Input Using Causal Associations from the Internet, FIT2008, 267-26.

[6] Dybala, P., Ptaszynski, M., Rzepka, R., and Araki, K. 2009. When Should Computers Joke? Concept of Emotiveness Analysis Based Timing Algorithm for Humor-Equipped Conversational Systems, NLP2009, Tottori, Japan, 542-545.