Serious Processing for Frivolous Purpose – A Chatbot Using Web-mining Supported Affect Analysis and Pun Generation

Rafal Rzepka
Hokkaido University
kabura@media

Wenhan Shi
Hokkaido University
shibuka@media

Michal Ptaszynski
Hokkaido University
ptaszynski@media

Pawel Dybala
Hokkaido University
paweldybala@media

Shinsuke Higuchi
Hokkaido University
shin_h@media

Kenji Araki
Hokkaido University
araki@media

ABSTRACT

By our[1] demonstration we want to introduce our achievements in combining different purpose algorithms to build a chatbot which is able to keep a conversation on any topic. It uses snippets of Internet search results to stay within a context, Nakamura's Emotion Dictionary to detect an emotional load existence and categorization of a textual utterance and a causal consequences retrieval algorithm when emotive features are not found. It is also able to detect a possibility to make a pun by analyzing the input sentence and create one if timing is adequate.

Author Keywords

chatbot, web-mining, affect analysis, pun generation

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Natural language

INTRODUCTION

Encouraged by the attendees of EMNLP 2008 we have decided to present our basic system [1] enhanced by algorithms of its users - Ptaszynski and Shi's affect recognizer [2] and Dybala's PUNDA puns generator [3]. In this brief description we will introduce basics of each algorithm and propose one combined unlimited topic dialog system, which we want to demonstrate at IUI 2009. Without using any vocal characteristics, this system is able to recognize existence and a type of emotional load which helps when adding modality

---

[1]All authors belong to Hokkaido University Graduate School of Information Science and Technology, 062-0814 Sapporo, Kita-ku, Kita 14, Nishi 9, Japan. Contact phone number: +81-11-706-6535, all e-mail addresses end with @media.eng.hokudai.ac.jp - abbreviated due the lack of space.

to a preposition created from Web-based association lists. It also decides if it is good timing for a joke or not when a pun generator discovers a chance to use a word play. Although the combined system is very fresh at the moment of sending this demo proposal and we have done only preliminary experiments, the efficiency of its parts alone was highly evaluated what can assure that it is worth presenting even if the cooperation of all elements appears not perfect in a bigger scale evaluation. It also should gather participants' interest as an easily extendable platform which can be built by using freely accessible sources.

MODALIN - THE BASIC SYSTEM

This simple conversation agent was designed as a platform for further experiments with machine learning, dialog modeling, user preferences, topic shifting, affective computing, humor processing, machine ethics, etc. The system generates replies by extracting noun keywords from user utterance, extracting word associations lists from the Web, generating sentence proposition using the extracted associations and by adding modality to the sentence proposition. By "preposition" we mean information conveying part of a sentence as "smoking is unhealthy" and "modality" we define as an emotion conveyer - "**Oh**, smoking is unhealthy, **isn't it?!**". All the details of the Modalin system are described in Higuchi et al paper [1] and will be not mentioned here for the obvious reason of this paper length restrictions. Also the specifications of below described emotion recognition system [2] and pun generator [3] are introduced in separate papers which have just appeared in 2008. To give the reader a hint on the Modalin's output level we cite a part of one conversation below:

user: *What food do you like ?*
system: *Well, food tastes good.*
user: *Recently, I've begun to like alcohol, too.*
system: *But, I can't drink much.*
user: *Really...? Do you get drunk quickly?*
system: *Well, it's amazing.*
user: *I have a rather high tolerance for alcohol.*
system: *Whow, strong, I can say.*

AFFECT ANALYSIS

This method allows to analyse an affect of an utterance by using Nakamura's "Emotion Dictionary" [4] and a support of Web mining technique when the dictionary is not enough to detect the emotion category[2]. It was proved that Web-mining improved the accuracy of the algorithm in extracting the specific types of emotions by retrieving causal consequences and checking their emotional load. For example a sentence "I found a cheap supercomputer" has no emotive elements and emotion classifying words from Nakamura's dictionary, however searching for "found a supercomputer (and|then|therefore|etc)" give us a set of following sentences which can be processed for emotion type deduction. If the most of causal phrases is automatically labeled as "amazement" then this becomes a label describing the emotional meaning of the input sentence. The evaluation experiments achieved 75% of human level in recognizing emotions but categorization skills, though outperforming other systems, achieved over ten percent less in determining the emotion types.

PUNDA - A PUN GENERATOR

As we plan to demonstrate our system by using a Japanese-English machine translation, this part of the chatbot will be hard to evaluate as depending on the presenter interpreting skills. However, it is an important part of the system and a very original contribution, so we decided to describe it here. The PUNDA algorithm is capable of finding word similarities and choose a word which is similar and probably funny. For instance:

user: *Windy days donft seem too nice.*
system: *Speaking of wind, it got me pinned!*
In this example, the Japanese word "kaze" (wind) became a trigger for finding a word "kazen" meaning "as expected". In Japanese this kind of word play is called *oyaji gyagu* (old pop's gag) and is not evaluated highly but still considered funny.

THE CHATBOT ALGORITHM

With already started trials of combining systems described above, we decided to combine them all and make the next step on the path to a naturally chatting agent. For trying the combination and presenting the used algorithms we decided to implement following simple algorithm (see Fig. 1): first, the emotive recognizer tries to categorize the emotion included inside the user's utterance. If it is recognized and it is not "fear", "relief" or "surprise" - the Punda Generator tries to find a pun fitting the keywords. If it is not possible, the Modalin creates a preposition and statistically chooses a modality using the Internet Relay Chat logs while the original program adds modality almost randomly. The chatlogs are automatically tagged by affect analyzer for emotion types and for sentence endings usually describing grammatical functions and dialog acts[3].

---

[2]Nakamura categorizes emotions into ten types: joy (delight), anger, sorrow (sadness), fear, shame (shyness, bashfulness), liking (fondness), dislike (detestation), excitement, relief and surprise (amazement).

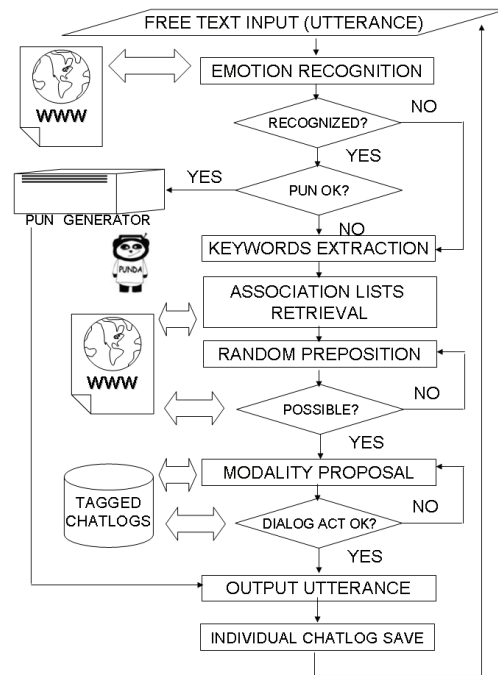[3]For determining the ending types we used a grammar book for Japanese language learners.

Figure 1. Three systems combined into one chatbot.

CONCLUSIONS AND FUTURE WORK

We have yet no proofs to confirm that our assumptions made while planning the merging of systems were correct. It is still unknown when the system should and should not be funny, it is not sure if statistical approach of "majority is always correct" helps to generate more natural modality. However, we think the parts of the systems are advanced enough to demonstrate their abilities to the Intelligent User Interfaces community and to spark a creative discussion on interfaces "emotional intelligence".

REFERENCES

1. Shinsuke Higuchi, Rafal Rzepka and Kenji Araki, "A Casual Conversation System Using Modality and Word Associations Retrieved from the Web", Proc. EMNLP 2008, USA, 2008, pp. 382-390.

2. Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, "Disentangling emotions from the Web. Internet in the service of affect analysis", Proc. KEAS'08, Japan, 2008, pp. 51-56.

3. Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka and Kenji Araki, "Humor Prevails! - Implementing a Joke Generator into a Conversational System", Proc. AI'08, New Zealand, 2008, pp. 214-225.

4. Akira Nakamura. 1993. Kanjo hyogen jiten [Dictionary of Emotive Expressions] (in Japanese). Tokyodo Publishing, Tokyo. 1993