Naturalness of an Utterance Based on the Automatically Retrieved Commonsense

Rafal Rzepka, Yali Ge and Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
Language Media Laboratory
Kita-ku Kita 14 Nishi 9, 060-0814 Sapporo, Japan
kabura,geyali,araki@media.eng.hokudai.ac.jp

Abstract

In this research we investigated user's behavior while facing a system coping with common knowledge about keywords and compared it with not only classic word-spotting method but also with random text-mining. We show how even a simple implementation of our idea can enrich the conversation and increase the naturalness of computer's utterances. Our results show that even very commonsensical utterances are more natural than classic approaches and also methods we developed to make a conversation more interesting. For arousing opinion exchange during the session, we will also briefly introduce our idea of combining latest NLP achievements into one holistic system where the main engine we want to base on commonsense processing and affective computing.

1 Introduction

During one of the NLP community meetings in Japan we attended a discussion about evaluating the freely (meaning "in the open domain") talking computer systems. The participants agreed that such systems are almost ignored by computer scientists while so called "chatter-bots" as Parry [Colby, 1975] or ELIZA [Weizenbaum, 1976] seem to be very often the only known yield of AI for other scientists and non-scholars. But today, even if passing the Turing test is still questionable, we have faster and faster access to massive corpora, which we believe may incline researchers to rethink the need of naturally talking systems and we claim they might be done without much labor and cost. Such renaissance would not only be important for the public recognition of the field but also could bring more young scholars fascinated by the fact that Internet has a chance to replace programmer's will and change the program's behavior depending on the facts discovered in millions of web-pages. Already Rzepka et al.[**?**] has noticed such possibility proposing an automatically developing average personality of "Mr. Internet" and has confirmed that commonsense knowledge can be retrieved from the WWW. We want to go a step further and show that using even simple usage of commonsensical (S)VO-then-V (Verb-Object and following Verb) and (S)VO-if-(S)VO (Subjects vary depending on the conversation processing so ignored here) phrases text-mined from homepages improve the user's acceptation for talking systems.

2 Commonsense Knowledge Retrieval

2.1 Collected Data Structure

For the set of experiments we performed, we used only nouns as for decades they were usually the main part of open-domain chat programs, although we claim that te commonsensical context is the future of dialogue processing. Using a nouns list and Larbin robot we created 1,907,086 sentences WWW corpus which was a base for a verb, a noun and VO n-gram dictionaries. The noun and verb dictionaries consist of 79,460 verbs and 134,189 nouns extracted with help of ChaSen Analyzer. For creating VO phrases automatically, our system had to search for the relationships between verbs and nouns and also between verbs. In this step, we used the verbs and nouns which had the highest occurrence and are common, as they are used in everyday live, for example [pour/drink/melt]-*water*, [listen/switch on/enjoy]-*music* or [go to/buy at/enter]*convenient store*. We used Japanese language, which has useful grammar features like *rentaikei* where the verb suffix *te* usually joins verbs in a time sequence e.g. *gohan wo tabe*-**te** *neru* (to sleep after having a dinner) or *tara*, *eba* and *to* "if" forms which are able to distinguish different causal connotations. By these useful grammar features we are able to web-mine commonsensical knowledge as "it is usual that some people buy sweets at convenient store even if they didn't wanted". Until now such data had to be collected manually[Singh, 2002] but full automatizing of such knowledge collecting brings new opportunities not only for dialogue but also for storytelling, question answering, machine translation and many other fields.

2.2 Architecture Overview

Basically, our system's architecture for creating commonsensical data can be summarized into the following processing steps:

a) A noun of is assigned for a keyword;

b) The system uses our web corpus for frequency check to retrieve 3 most frequent verbs following the keyword noun;

c) The most frequent particle between noun keyword and 3 most frequent verbs is discovered;

d) For creating bi-gram the system retrieves a list of most frequent verbs occurring after the previously chosen verb;

e) By using Yahoo search engine, the system checks if the noun-particle unit occurs with new verb-verb unit for time-sequence actions and verb-if unit for casual dependencies;

f) If yes - the VO-then-V and VO-if-VO units are stored:

$$\mathbf{VO_{then}V} = \mathbf{N} + \mathbf{P}_{\max} + \mathbf{V}_{\max1} + \mathbf{V}_{\max2}$$

N:*Triggering noun (keyword);*
$P_{\max}$:*most frequent particle joining noun and verb;*
$V_{\max1}$:*most frequent verb occurring after the* N*;*
$V_{\max2}$:*most frequent verb occurring after* $V_{\max1}$*;*

$$\mathbf{VO_{if}V} = \mathbf{N_1}+\mathbf{P_1}_{\max}+\mathbf{V_1}_{\max}+if+\mathbf{N_2}+\mathbf{P}_{2\max}+\mathbf{V}_{2\max}$$

N_1:*Triggering noun (keyword);*
$P_1\max$:*most freq. particle joining first noun with a verb;*
$V_1\max$:*most freq. verb after the* N_1P_1*;*
$N_2\max$:*most freq. noun after* $N_1P_{1\max}V_{1\max}$ *and "if";*
$P_2\max$:*most freq. particle joining* N_2 *and* V_2*;*
$V_2\max$:*most freq. verb after* N_2P_2*;*

3 Experiments

In order to see user's perception of the basic commonsense knowledge included in a utterance, we performed a set of experiments basically using four kinds of utterances following input with one noun $keyword$:

- ELIZA's output [ELI] (input sentence structure changing to achieve different outputs)
- WWW random retrieval output [WRR] (a shortest of 10 sentences retrieved by using $keyword$ and query pattern "did you know that?")
- WWW commonsense retrieval output "high" [CS1] (sentences using common knowledge of highest usualness (most frequent mining results)
- WWW commonsense retrieval output "low" [CS2] (sentences using common knowledge of the lowest usualness (least frequent mining results).

Typical ELIZA answer is "why do you want to talk about smoking" if the $keyword$ is "smoking". For the same $keyword$ WRR retrieved a sentence "did you know that people wearing contact lenses have well protected eyes when somebody is smoking?". An example of CS1 is "you will get fat when you quit smoking" and CS2 is "smoking may cause mouth, throat, esophagus, bladder, kidney, and pancreas cancers". We selected 10 most common noun keywords of different kinds (water, cigarettes, subway, voice, snow, room, clock, child, eye, meal) not avoiding ones often used in Japanese idioms (voice, eye) to see if it influences the text-mining results. 13 referees were evaluating every set of four utterances in two categories – "naturalness degree" and "will of continuing a conversation degree" giving marks from 1 to 10 in both cases.

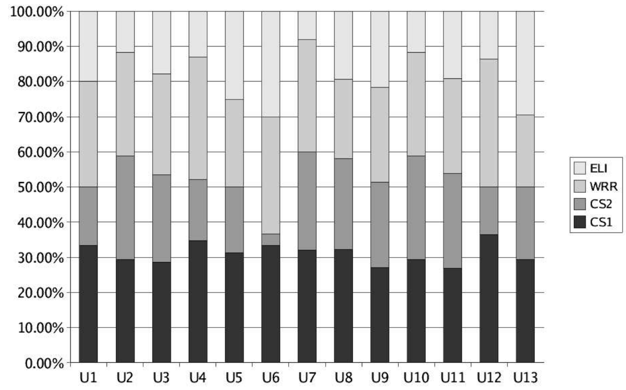

Figure 1: Naturalness Level Evaluation: CS1:29.5%, CS2:21%, WRR:24.5%, ELI: 25%

4 Results and Conclusions

In "continuation will degree" ELIZA unsurprisingly achieved 452 points out of 2919 for four systems (only 15.48%). But the performance of commonsenseical utterances was surprisingly high (CS1:25.38%), CS2:27.14%) which suggest that interlocutor prefers a machine saying "smoking is bad" than one naturally asking questions. The highest result of WRR (32%) tells us how simple tricks can help on keeping up the conversation. On the contrary, the naturalness degree results (see Fig.1) show that the "tricks" of ELIZA and WRR and information overload of CS2 are less natural than the ordinary truth statements. Due to the lack of space, more specific results analysis and graphs we are going to provide during the poster session.

References

[Weizenbaum, 1976] Joseph Weizenbaum. *Computer Power and Human Reason*. W.H. Freeman and Co., New York, 1976.

[Colby, 1975] Kenneth Colby. *Artificial Paranoia: A Computer Simulation of Paranoid Process* Pergamon Press., New York, 1975.

[Rzepka et. al, 2003] Rafal Rzepka, Kenji Araki and Koji Tochinai. Ideas for the Web-Based Affective Processing. *Proceedings of the Seventh Multi-Conference on Systemics, Cybernetics and Informatics; vol. XIV Computer Science, Engineering Applications*, vol XIV:376–381, August 2003.

[Singh, 2002] Push Singh. The public acquisition of commonsense knowledge. *Proceedings of AAAI Spring Symposium on Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Palo Alto, CA: AAAI, 2002.