1B1-08

Automatic First Utterance Creation Based on the Strongest Association Retrieved from WWW

Rafal Rzepka*1 Kenji Araki*1

*1 Graduate School of Information Science and Technology, Hokkaido University

In this paper we propose an algorithm that can be used by talking systems to create a list of natural replies to a user's utterance starting any dialog without restricting any domain. In our method we use Internet resources, commonsense processing and confront them with affective features to achieve a linguistic reaction of the highest naturalness as possible. We briefly introduce our approach to the dialog processing itself and suggest the need for investigating Pavlovian-like human linguistic behaviors.

1. Introduction

During developing our approach to dialogue processing, we decided to basic instincts that motivate people to talk, to reply questions, to react to statements. For that reason we concentrated on the very first utterance generation which we claim to be crucial in human-machine interaction which beginning depends mostly of the machine's purpose. Nowadays, most of the interfaces are for a specific task where reaction to the user's input is quite predictable, but as for the future challenge we will have multi-task open-domain talking robots which will be hard to put into one linguistic behavior frame. There also will be robots for elders which need to trust accompanying machines more than younger generations which produce these robots. It is not questionable that the understanding is the base for trusting someone but as we know, computers have problems with comprehension especially in open domains. On the contrary, being flexible to talk on every topic or have a natural comment on every utterance is what users think as "natural dialogue capability". This is probably why the famous systems as ELIZA [1] were taken as quite "human-like" - they were behaving naturally even if being considered boring or silly. Our experiments show that even a very simple keyword-spotting-based method using commonsensical knowledge sounds more natural and is a way more interesting than ELIZA which starts all its' conversation with basically one-pattern.

2. Automatic Commonsense Retrieval

As we have already confirmed [Rzepka 03abc] that the commonsense knowledge can be retrieved from the WWW – we showed measuring methods for Usualness and Positiveness. This time we went a step further and showed that using commonsensical (S)VO-then-V (Verb-Object and following Verb) and (S)VO-if-(S)VO phrases retrieved from homepages improves the user's acceptation for talking systems. In our commonsense retrievals we are inspired by Minskian frames [Minsky 75], Schankian scripts [Schank:77] and casual theories of Fillmore [Fillmore 68].

The beginning of this century showed us quite a big range of applications that are using the WWW as a corpus [Keller 03][Santamaria 03] but there is still one fundamental difference - human users need to retrieve information which they do not have - here the machines mine knowledge which is obvious for humans but not computers.

2.1 Technical Solutions

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

- A noun of is assigned for a **keyword** (any keyword spotting method can be applied);
- The system uses our web corpus for frequency check to retrieve **3 most frequent verbs** following the keyword noun;
- The **most frequent particle** between noun keyword and 3 most frequent verbs is discovered;
- For creating bi-gram the system retrieves a list of most frequent **verbs** occurring **after** the previously chosen **verb**;
- 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**;
- If yes - the VO-then-V and VO-if-VO **units are stored**:

$\mathbf{VO_{then}V} = \mathbf{N} + \mathbf{P}_{\max} + \mathbf{V}_{\max 1} + \mathbf{V}_{\max 2}$ *N:Triggering noun (keyword);*
$P_{\max}$*:most frequent particle joining noun and verb;*
$V_{\max 1}$*:most frequent verb occurring after the N;*
$V_{\max 2}$*:most frequent verb occurring after* $V_{\max 1}$*;*
$\mathbf{VO_{if}V} = \mathbf{N_1} + \mathbf{P_{1\max}} + \mathbf{V_{1\max}} + if + \mathbf{N_2} + \mathbf{P_{2\max}} + \mathbf{V_{2\max}}$
N_1*:Triggering noun (keyword);*
P_1max*:most freq. particle joining first noun with a verb;*
V_1max*:most freq. verb after the* N_1P_1*;*
N_2max*:most freq. noun after* $N_1P_{1\max}V_{1\max}$ *and "if";*
P_2max*:most freq. particle joining* N_2 *and* V_2*;*
V_2max*:most freq. verb after* N_2P_2*;*

Contact: Language Media Laboratory, Research Group of Information Media Science and Technology, Division of Media and Network Technologies, Graduate School of Information Science and Technology, Hokkaido University, Kita-ku Kita 14 Nishi 9, 060-0814 Sapporo, Japan. TEL: (+81)(11)706-6535, FAX: (+81)(11)706-6277, {kabura,araki}@media.eng.hokudai.ac.jp

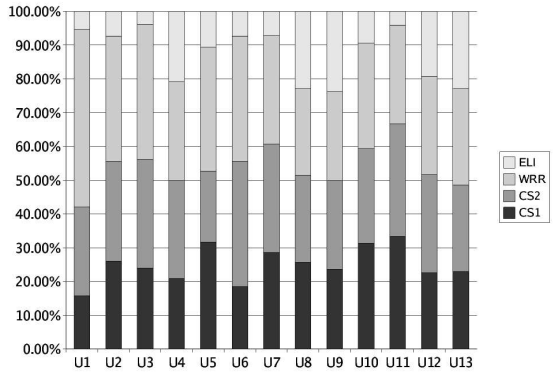

Figure 1: Interest Level Evaluation

3. Experiments

As we wanted 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.

4. Results

As we expected, in "continuation will degree" ELIZA 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 session. We also will talk about the other subtopics which we mentioned in the abstract and the introduction.

5. Conclusions

We proved that users of human-machine interfaces prefer talking to a grammatically and contextually imperfect system than one which is contextually and grammatically correct but lacks of commonsensical knowledge. We also showed that changing the "level of being common" and using "interest keywords" during web-mining may simply make the conversation joyful for the user.

6. Near Future Challenge

For the perfect comparison with Eliza abilities, we need to develop turn-taking algorithm next. We want to confirm that joining our method with even the simplest context-based dialogue system will outperform the classic open-domain programs.

References

[1] Weizenbaum, J.: Computer Power and Human Reason, W.H. Freeman and Co., New York (1976)

[Fillmore 68] Fillmore, J.C.: The Case for Case. E. Bach & R.T.Harms, eds., Universals in Linguistic Theory, New York: Holt, Rinehart & Winston, (1968) pp. 1-88

[Keller 03] Keller, F., Lapatay, M.: Using the Web to Obtain Frequencies for Unseen Bigrams. Computational Linguistics, Vol 29, No 3, September, (2003) pp. 459-484

[Minsky 75] Minsky, M.: A Framework for Representing Knowledge. In: P. H. Winston (ed.), The Psychology of Computer Vision. New York: McGraw Hill (1975) pp. 211-77

[Rzepka 03abc] Rzepka R., Araki K., Tochinai, K.: Bacterium Lingualis - the Web-based Commonsensical Knowledge Discovery Method. Lecture Notes in Artificial Intelligence series of Springer-Verlag, volume 2843. (2003) pp. 460-467

[Santamaria 03] Santamaria, C,, Gonzalo, J., Verdejo, F.: Automatic Association of Web Directories with Word Senses. Computational Linguistics, Vol 29, No 3, September (2003) pp. 485-502

[Schank:77] Schank, R., and Abelson, R. Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Erlbaum.(1977)