Commonsense Retrieval as an Aid for Easier Conversation-based Language Acquisition

Rafal Rzepka and Kenji Araki

Graduate School of Information Science and Technology,
Hokkaido University,
Kita-ku Kita 14-jo Nishi 9-chome, 060–0814 Sapporo, Japan
{kabura,araki}@media.eng.hokudai.ac.jp

Abstract. In our paper we will introduce ideas for auto-evaluation of commonsensical causations that are being learned inductively not only from the dialogue history but also from samples which are automatically retrieved from Japanese homepages. Such support is mainly based on the knowledge text-mined from the Web and widely used for decreasing human utterer's labor of proofreading the language-learning system and at the same time to make the conversation more interesting and what is important for "zero level knowledge at the start" approaches which quickly bore the supervisor.

1 Introduction

In our research we are trying to develop language acquisition methods to be used in talking agents. Learning through the conversation has some disadvantages and the biggest ones seem to come from the psychical tiredness of a human user who the machine learns language from. Especially in cases as ours, where the goal is to achieve the best possible algorithm working from the zero level knowledge meaning no data when the learning process begins, the user gets bored with monotonous repetitions, explanation demands and only basic questions. As there were situations where the program lacked further questions having problems with keeping up the conversation, Kimura at al.[1] decided to change these breaks of conversational flow by using Eliza algorithm[2] which main purpose was to keep a conversation in the most natural way as possible. Eliza helped to avoid conversation crashes and to raise the performance but the learning session time was still unsatisfactory as the supervised was getting bored with the system. All chatbots, as the open-domain, light-conversations programs are called were almost as predictable as their Weizenbaum's ancestor, or were unpredictable at all as Colby's paranoid Parry program[3] which also was based on Eliza. Therefore we decided to find a way to make the conversation interesting even if an utterance was only slightly on topic (user's reactions on off-topic utterances are also an important part of language acquisition using Araki's Inductive Learning[4]). To achieve that goal we made four sets of experiments to show how the Internet resources could solve our problem. We compared Eliza typical utterances and three WWW-based utterances of different levels of peculiarity. .

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 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" (We also decided to concentrate on one language as we noticed that general beliefs which are part of commonsense depend on cultures). Until now such data had to be collected manually[6] 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:

$$\mathbf{VO_{then}V} = \mathbf{N} + \mathbf{P}_{\max} + \mathbf{V}_{\max 1} + \mathbf{V}_{\max 2}$$

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);
$P_1\max$: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;

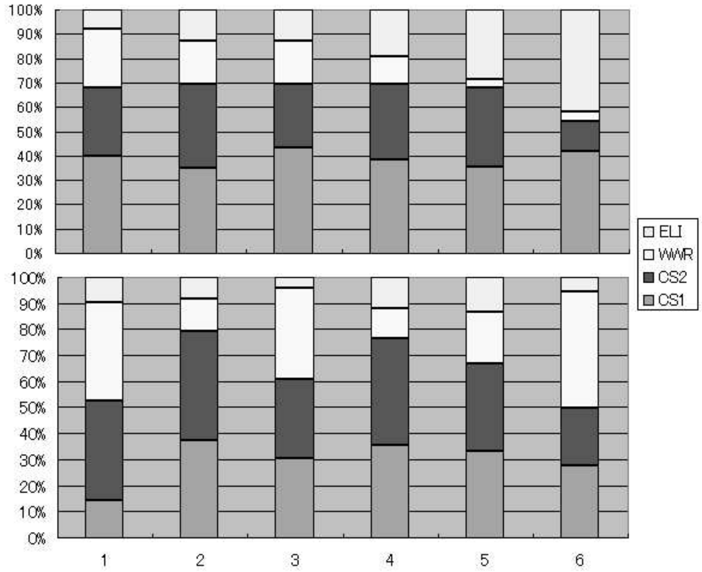

Fig. 1. Naturalness (up) and Interest (down) Levels Evaluation; $keyword$: "child", 6 referees

3 Inductive Learning

For language acquisition we use Inductive Learning proposed and evaluated for various NLP applications by Araki et al.[4]. It is based on cross-referential approach of combining common and different parts of input-output pairs used for learning. In our case we used "if" sentences as inputs and commonly following them causations as outputs. To achieve as many common parts as possible to enrich learning process, we concentrated on strictly commonsensical knowledge retrieved from the WWW. To decrease the user's labor we also propose a simple web-based proofreader which allows to avoid very basic errors which might easily irritate the user.

3.1 Idea of Self-Proofreading

We are working on a simple statistical method which is combined with a categorization in order to reach as high accuracy as possible without losing the aspect of being fully automatic. For instance if the Inductive Learning method creates new rule saying that if somebody is sick goes to doctor, by using Japanese EDR definition collection, we can translate the rule into a higher level of abstraction: if PERSON sick then go to PROFESSION. Simple set of search engine queries can easily avoid creating rules as: if *an idea* sick then go to PROFESSION (English simplified to resemble Japanese).

4 Users Reactions for Commonsensical Utterances

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

- ELIZA's[2] 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. Thirteen 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. The system comparison results proved that ELIZA does not eager users for continuing the chat but is still useful to keep the utterance naturalness. However, we proved that using commonsense even of the highest usualness is more natural than famous classic system (ELI 46%, CS1 54%). We also confirmed that query-based web-mining (WRR) results have slightly better user's acceptance than less common causal knowledge (CS2) which we find useful for creating a method for automatic category-based query formation depending of user's input.

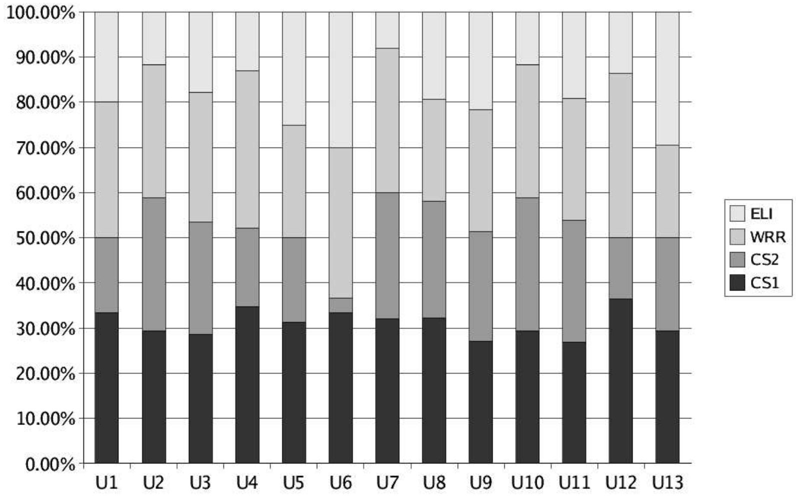

Fig. 2. Naturalness Level Evaluation

5 Results

The system comparison results proved that ELIZA usual utterance does not eager users for continuing the chat but is still useful to keep the utterance naturalness. However, we proved that using commonsense even of the highest usualness is more natural (ELI 270pts, CS1 304pts). We also confirmed that query-based web-mining (WWR) results have slightly better user's acceptance than less common casual knowledge (CS2) which give us a hint to find a method of automatic category-based query forming depending of user's input. The experiments on web-based self-proofreading are not finished yet but preliminary results were satisfactory enough to make a decision of including this part into the draft paper and describe the achievment state in the final paper.

6 Conclusions

In our experiments we have 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. In our research we proved how easily commonsensical causations can be discovered in enormous, mostly chaotic, data resources as WWW.

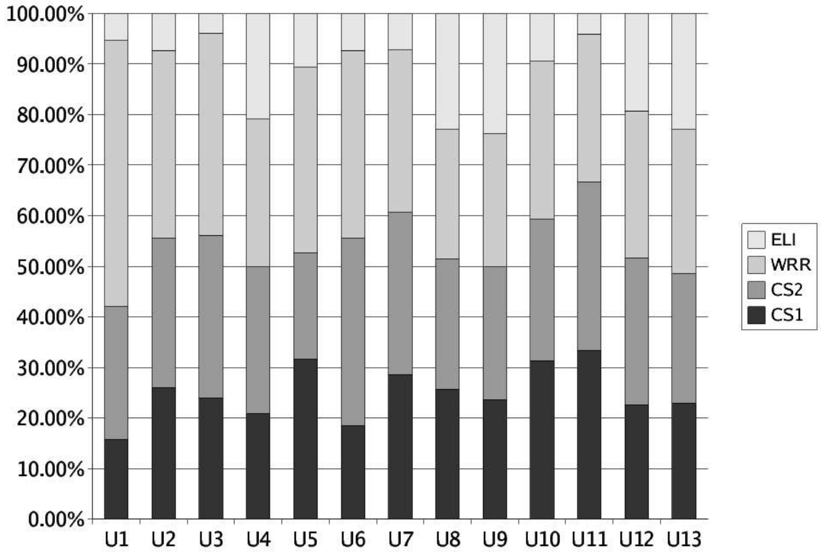

Fig. 3. Interest Level Evaluation

There is remaining problem of time consumption but it is mostly due to the netiquette which does not allow for very fast retrieval within the search engine results. However, the commonsense processing in our future plans is supposed to work with an algorithm reducing causations by the context which will simplify query formation by increasing numbers of query keywords and making the search incomparably faster. It should also help to get rid of causation units' ambiguity, as the Internet brings also often contradictory statements like "drinking water makes you healthier" and "drinking water makes you sick". We do not have to assume that one of these claims is wrong - by discovering the contextual information we will become able to distinguish in which cases above mentioned statements are correct and in which, by contradiction, are not. In the application experiment we proved that this retrieved data can make a Human-Computer Interfaces sound more natural and interesting if we use opposite weights of commonsense expressions. For retrieving the causations we use several Japanese "if" forms which are specific to this language helping to divide the causal knowledge into categories at the start of processing. Our method for creating utterances appeared to be more natural than classic fully automatic methods as ELIZA which remains popular even if such approach requires laborious rules creation. We must underline here that for learning language, the relevancy of the dialog turns was not important on this stage. In a perspective of one utterance quality we achieved higher naturalness without almost any labor and it is obvious that users prefer keep talking to systems based on the WWW that to these limited to their internal databases and programmer's imagination.

References

1. Yasutomo Kimura, Kenji Araki and Koji Tochinai. Evaluation of Spoken Dialogue Processing Method Using Inductive Learning with Genetic Algorithm Proceedings of the IASTED International Conference "Artificial Intelligence and Soft Computing", pp.295-300, 2001.
2. Joseph Weizenbaum. *Computer Power and Human Reason*. W.H. Freeman and Co., New York, 1976.
3. Kenneth Colby. *Artificial Paranoia: A Computer Simulation of Paranoid Process* Pergamon Press., New York, 1975.
4. Kenji Araki and Koji Tochinai. *Effectiveness of Natural Language Processing Method Using Inductive Learning,* Proceedings of the IASTED International Conference "Artificial Intelligence and Soft Computing", pp.295-300, 2001.
5. 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.
6. 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.