Machines Having Other Ideas than Ours? – – Evaluation of a System Based on Commonsensical Knowledge Retrieved from the Web

Yali Ge, Rafal Rzepka and Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
{geyali,kabura,araki}@media.eng.hokudai.ac.jp

In our paper we will present results of evaluation made by users who faced output of our system designed to discover uncommon actions. By using web-mining techniques and simple statistics, our system is able (to some extent) to find unusual acts and suggest users more usual ones. In the first step, the program decides if it is natural or not. Although the suggestions do not come from a programmer, different users react in different way to the output which might be seen as unnatural or even cheeky while suggesting that someone's behavior is strange. After we introduce our system and the algorithms of using natural language to retrieve knowledge from the WWW resources, we show the results of retrievals which are followed by analyzing reactions of users who experienced output from a machine pretending to have its own common sense. In the end we would like to conclude with analysis of the users' expectations while using machines with talking abilities as they differ depending on situation.

Keywords: Commonsense Processing, Web-mining, Evaluation Problems

1 INTRODUCTION

1.1 State of Art

Machines would be regarded smarter if they had common sense background, although the amount of data that should be inputted for achieving this knowledge is enormous. The beginning of common sense studies in AI was made by - among others - Minskian frames [1], Schankian scripts [2] or casual theories of Fillmore [14], however an universal program that would behave naturally in any environment with any user is still technically impossible. Today, the same researchers are alarming that it is much more difficult task than they presumed [4][5] and it became obvious that we are not so close to achieve this goal as many thought. However nowadays bring also the Internet - the biggest text repository in the world, which we decided to use as a source for common sense retrieval. Now there are many researchers using the WWW as a corpus [6][7] but our approach is different - they retrieve information to make humans richer or smarter and we retrieve knowledge which should make the machines smarter. Similar, though different goals and resources approach to ours is represented by Inui et al. [8], but most of current researchers concentrate on manual gathering of common sense like Lenat's CyC [9] or Mueller's "Thought Treasure"[10]. As the CyC project does not give satisfying results for years of absorbing millions of US dollars, a MIT group decided to cooperate with Thought Treasure's author and started its own project based on the idea where the hundreds of Internet users are inputting the sentences which are supposed to be categorized within the proposed categories[11] [12].

1.2 Existing Problems

Common sense logic is obviously different from mathematical logic, because its rules cannot be proved in every situation, with everyone or in every culture, maybe except some physical phenomena. Although the first steps for achieving commonsensical knowledge for a given situation were made [13] there are still problems with evaluation of common sense output. While broad knowledge as "fire burns" would be regarded as natural and useful for a housework robot [1], more specific common sense as "fire can destroy the woods" tends to be evaluated lower as too specific for mechanical actor working in the kitchen. This paper will introduce the problem of subjective evaluation by a questionnaire where the judgment is much more influenced by personal experiences and the attitude toward such system than in a case of specific knowledge.

2 OUR SYSTEM

2.1 Description

Our system uses Japanese, and analyses one pattern sentences describing simple actions in form of one noun, one particle and one verb which were freely entered through WWW interface and tries to calculate previous and following actions using an idea of Schankian scripts[2].

2.2 Algorithm

Below is the step-by-step procedure:

1. A sentence (Place)-Noun *de* Noun *wo* Verb is inputted.
2. Is such combination possible?
   if YES then NEXT
   if NO then SEARCH A CASE WHEN YES (e.g.: if place A is impossible suggest up to 3 most likely places B)

[1] We set a "Housework Robot" as a default actor and beholder of the knowledge to simplify the judgment.

3. ChaSen[15] extracts nouns, verbs, particles and adjectives

a) creating Noun Dictionary
b) creating Noun-Particle Dictionary [2]
c) creating Verb Dictionary I[3]
d) creating Verb Dictionary II[4]
e) creating Verb-Verb Dictionary
f) creating Noun-Particle-Verb Dictionary
g) creating Adjective Dictionary
h) creating Places Dictionary

4. By using Verb-Verb Dictionary and Noun-Particle-Verb Dictionary, our system can get the possible actions which might have taken place before and after the "object+verb" combination inputted by user.

This algorithm is to retrieve actions following the one which is inputted to create a script. As a rule, we counted repeated actions as natural if only they were not repeated one after another, for example "watch - think - watch (again)" or "drink - go back home - drink (again)" were accepted while "look - look - look" or "come back home - watch - watch" were not. Such repetitions were due to the fact that our system still does not retrieve noun-particle units with verbs retrieved for actions following input action. This was also the reason we are still not able to fairly evaluate scripts retrieval. If the noun was common topic for actions or it was not needed, they were natural, for example "Give birth to a child - take care of (it) - do best", "wait for friends - have fun - go back home", "wait for friends - go back home - sleep" or "take out the trash - look through (it) - find out (something)". We also approved grammatical forms meaning movement or giving/receiving (*-te morau/ageru, -te kuru/kaeru/iku*) but we did not approved continuous *-te iru* and *-te miru* (try to...) as static and not associated with actions in a satisfactory level. When evaluating naturalness of retrievals we were labeling natural sentences with high frequency and unnatural sentences with low frequency as natural and opposite.

3 EXPERIMENT AND ITS RESULTS

We picked results of 34 inputs made by users through a cgi-based WWW interface caring to choose a wide range of different examples. As can be seen in the appendix, we chose actions not only associated with housework, there were strange actions and inputs with mistaken language[5]. We presented 10 judges an evaluation list where every input was also divided into "place + verb", "object + verb" and possible previous and following actions were proposed. Every such output was labeled by the system as "natural" or "unnatural" and user was to mark this label in a scale from 0 to 5 in two levels: how precise the label is and how useful it would be for a housework robot as a

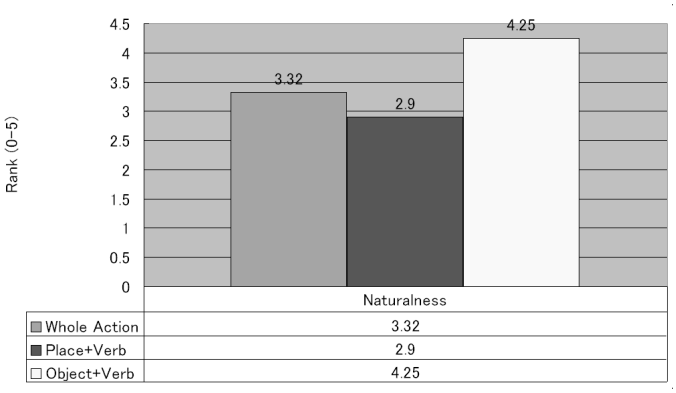

Figure 1: Preciseness evaluation of commonsensical knowledge (0-5 rank)

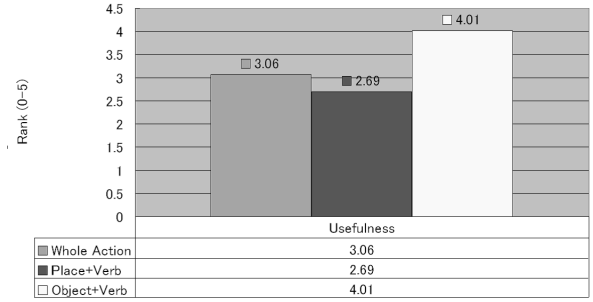

Figure 2: Usefulness evaluation of commonsensical knowledge (0-5 rank)

knowledge about human being. Previous and next actions were evaluated for naturalness instead of preciseness and not the whole input but object-verb form was the base for retrieving the scripts. For example:
Buying a lunch-box at convenient store IS NATURAL (preciseness:0-5)(usefulness:0-5)
Buying a lunch-box IS NATURAL (preciseness:0-5)(usefulness:0-5)
Buying at convenient store IS NATURAL (preciseness:0-5)(usefulness:0-5)
Possible previous action: choosing a product (naturalness:0-5)(usefulness:0-5)
Possible next action: eating a lunch-box (naturalness:0-5)(usefulness:0-5)

As can be seen in the Figure 1., Figure 2. and Table 1., the average marks given by ten judges for the preciseness of naturalness labeling were 66.4% for the whole sentence, 85% for object-verb form and surprisingly only 57% for "place+verb" form. The usefulness for these three forms was evaluated similarly : 61.2%, 80.2% and 54%. Previous actions were evaluated natural in 67.8% and useful in 61.2% cases while next actions were evaluated natural in 55% and useful in only 45.5% cases.

4 CONCLUSIONS AND FUTURE WORK

From our experiments we understood several problems of common sense evaluation. First, we discovered that rela-

[2] *o, wa, ga, no, ni, de, he, mo, to, kara*

[3] in dictionary form *jisho-kei*

[4] in form which was actually used

[5] This might be the case when voice recognition fails and input becomes strange - we wanted to see if our system is able to recognize properly these abnormalities.

Table 1: Average marks (rank 0-5) of possible previous and following actions

	Previous Action	Next Action
Naturalness	3.39	2.75
Usefulness	3.06	2.27

tively low marks for preciseness and usefulness were due to the high differences within the marks of the same output. It showed how different points of view our judges had. For example "Buying fruits IS NATURAL" was judged 0 by one and 5 by another judge while usefulness was marked in opposite way. It suggests than every judge has his or her own idea of housework robot and what kind of knowledge it should have and which common sense is useless. It was also the case in the previous and next actions but interesting thing is that whole sentence was marked higher than place-verb form what proves that the more specific description of situation helps a person to imagine the usage of common sense knowledge. On the other hand places are much more flexible parameter in such knowledge and everyone, depending on his/her background, perceives actions as natural and useful in different places. Oppositely, verbs are strongly associated with objects what was proved with the highest marks in overall evaluation. It appeared that in many cases a judge did not like the machine's idea and evaluated it low even the other judges gave the highest ranks. From these evaluation we understood that common sense is too much depending on personal judgments to be fairly evaluated and the best way to escape this problem is to evaluate machines using common sense knowledge instead of such knowledge itself. Our next step will be a system using commonsensical knowledge during simulation of particular place with particular task.

A Appendix: Input Sentences

Sentences (alphabetical order) inputted for analysis (tools instead places and wrong grammar were allowed as nouns and verbs were free to choose):

1.*Chuusha-jou-de tomodachi-wo matsu*: Waiting friends at a parking lot (PREVIOUS: Not Extracted; NEXT: Not Extracted)
2. *Daidokoro-de kudamono-wo kau*: Buying fruits at the kitchen (PREVIOUS: Choosing goods; NEXT: Coming back home)
3. *Daidokoro-de reizouko-wo arau*: Washing a refrigerator at the kitchen
(PREVIOUS: Openning the key; NEXT: Looking out the sea)
4. *Daidokoro-de sentaku-wo hosu*: Drying laundry at the kitchen
(PREVIOUS: Not Extracted; NEXT: Not Extracted)
5. *Daidokoro-de tabako-wo suu*: Smoking cigarettes at the kitchen
(PREVIOUS: Lighting the fire; NEXT: Having smell)
6. *Doko-ka-de gamu-wo kamu*: Chewing a gum somewhere
(PREVIOUS: Openning the mouth; NEXT: eatting something)
7. *Doko-ka-de gomi-wo taberu*: Eating trash somewhere
(PREVIOUS: Putting in the mouth; NEXT: fish is dying)
8. *Eigakan-de osake-wo kobosu*: Spilling alcohol at the movies
(PREVIOUS: Not Extracted; NEXT: Not Extracted)
9. *Eigakan-de osake-wo nomu*: Drinking alcohol at the movies
(PREVIOUS: Adding the water; NEXT: Driving the car)
10. *Heya-de gohan-wo taberu*: Eating a meal at the room
(PREVIOUS: Doing by myself; NEXT: providing the dessert)
11. *Heya-de mado-wo hirakeru*: Opening window at the room
(PREVIOUS: Attracting attention; NEXT: Changing the air)
12. *Heya-de mado-wo shimeru*: Closing window at the room
(PREVIOUS: Outting door; NEXT: sleeping at home)
13. *Heya-de o-mizu-wo taberu*: Eating water at the room
(PREVIOUS: Putting in hot water; NEXT: Giving food up)
14. *Hikouki-de kodomo-wo umu*: Giving birth inside an aeroplane
(PREVIOUS: Choosing partner; NEXT: Bring up the child)
15. *Hoteru-de tabako-wo suu*: Smoking cigarettes in a hotel
(PREVIOUS: Lighting the fire; NEXT: Having smell)
16. *Izakaya-de eiga-wo miru*: Watching a film at the Japanese-style bar
(PREVIOUS: Renting a video; NEXT: Not Extracted)
17. *Izakaya-de o-sake-wo kobosu*: Spilling alcohol at the Japanese-style bar
(PREVIOUS: Limited old man; NEXT: Coming back home)
18. *Konbini-de o-bentou-wo kau*: Buying a lunch-box at convenient store
(PREVIOUS: Choosing the goods; NEXT: eating the lunch-box)
19. *Michi-de gomi-wo hirou*: Picking up trash on the street. (PREVIOUS: Looking like somebody; NEXT: Taking back home)
20. *Mise-de ranchi-wo taberu*: Eating lunch in a bar
(PREVIOUS: Adding sauce; NEXT: Eating food)
21. *Mise-de tamago-wo kowasu*: Breaking an egg at a store
(PREVIOUS: Not Extracted; NEXT: Not Extracted)
22. *Pasokon-de shinbun-wo yomu*: Reading a newspaper with computer
(PREVIOUS: Choosing news; NEXT: Watching Television)
23. *Reizouko-de hon-wo arau*: Washing a book in a refrigerator
(PREVIOUS: Using machine; NEXT: Not Extracted)
24. *Resutoran-de o-sake-wo nomu*: Drinking alcohol at

restaurant
(PREVIOUS: Eating food; NEXT: Talking the situation)
25. *Ribingu-de kome-wo taberu*: Eating raw rice at living room
(PREVIOUS: start cooking; NEXT: Eating the food)
26. *Ribingu-de kome-wo taku*: Boiling raw rice at living room
(PREVIOUS: putting water in; NEXT: Eatting the food)
27. *Ribingu-de gohan-wo taberu*: Eating boiled rice at living room
(PREVIOUS: adding the soup; NEXT: getting some dressing)
28. *Sukii-jou-de shawaa-wo abiru*: Taking shower at the skiing ground
(PREVIOUS: Not Extracted; NEXT: Not Extracted)
29. *Suupaa-de yasai-wo toru*: Picking up vegetables at a store
(PREVIOUS: Washing the dishes; NEXT: Eating the food)
30. *Toire-de gamu-wo kamu*: Chewing a gum at the toilet
(PREVIOUS: Waxing; NEXT: Eating something)
31. *Toire-de gamu-wo suteru*: Throwing away a gum at the toilet
(PREVIOUS: Wrapping in paper; NEXT: Not Extracted)
32. *Toire-de gomi-wo suteru*: Throwing away trash at the toilet
(PREVIOUS: putting into a bag; NEXT: Having gracious)
33. *Toire-de mado-wo shimeru*: Closing a window at the toilet
(PREVIOUS: Catching a cold; NEXT: Sleeping at home)
34. *Uchuu-de hana-wo sodatsu*: Growing flowers in space
(PREVIOUS: Not Extracted; NEXT: Not Extracted)

Bibliography

[1] Minsky, M. **A Framework for Representing Knowledge**, The Psychology of Computer Vision, P. H. Winston (ed.), McGraw-Hill, 1975.

[2] Schank R.C. and Abelson R.P. **Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures (Chap. 1-3)**, L. Erlbaum, Hillsdale, NJ, 1977.

[3] 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

[4] Fillmore, C. J. **Frames and the Semantica of Understanding**, In: V. Raskin (ed.), Round Table Discussion on Frame/Script Semantics, Part I, Quadernidi Semantics VI: 2, (1985) pp. 222-254

[5] Minsky, M. **Society of Mind**, Simon and Schuster, New York (1988)

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

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

[8] Inui T., Inui K. and Y. Matsumoto **What Kinds and Amounts of Casual Knowledge Can Be Acquired from Text by Using Connective Markers as Clues?**, 6th International Conference on Discovery Science (DS 2003),Sapporo, Japan. Lecture Notes in Artificial Intelligence, vol 2843, 180-193, 2003, Springer-Verlag, Berlin, Heidelberg, New York.

[9] D. Lenat et al.,**Common Sense Knowledge Database CYC**, 1995, http://www.opencyc.org/, http://www.cyc.com/

[10] E. Mueller, **Natural language processing with ThoughtTreasure**, Signiform, New York, 1998.

[11] P. 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.

[12] D.G. Stork, **"Open Mind Initiative"**, 1999, http://openmind.media.mit.edu/

[13] R. Rzepka, K. Araki, K. Tochinai **Bacterium Lingualis - the Web-based Commonsensical Knowledge Discovery Method**, 6th International Conference on Discovery Science (DS 2003),Sapporo, Japan. Lecture Notes in Artificial Intelligence, vol 2843, 453-460, Springer-Verlag, Berlin, Heidelberg, New York, 2003.

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

[15] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, and M. Asahara, **Japanese morphological analysis system ChaSen version 2.0 manual 2nd edition**, 1999.