T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

Construction of Japanese Search Engine Based on Computational Model of Inductive Reasoning

Hidekazu KAMINISHI[a*], Shusuke SUZUKI[a], Asuka TERAI[b] & Masanori NAKAGAWA[a]
[a]*Graduate School of Decision Science and Technology, Department of Human System Science,Tokyo Institute of Technology, Japan*
[b]*Global Edge Institute,Tokyo Institute of Technology, Japan*
*hideka@mr.hum.titech.ac.jp

Abstract: Associative search systems can be applied to the education system for language learning. In this study, we developed an information search system for language learning based on the results of previous studies. This search system is based on a computational model of inductive reasoning which uses relationships of features and nouns in structures of language knowledge (probabilistically-based language knowledge structure). The system can eliminate unwanted results in accordance with association. In addition, via the system, previous studies results were modified based on psychological experiments. Finally, a survey was conducted in order to assess the validity of the results and confirm whether the output is valid.

Keywords: search engine, inductive reasoning, computational model, language learning

1. Introduction

1.1 Background

Recently, various search engines which do not depend on the match of keywords have been developed.. These search systems are applied to the education system for language learning. For example, with this system, we are able to type a word (or words) and find associated words. However, Search engine system which can get rid of the unnecessary result is required.

• fill-in-the-blank objective reasoning (e.g., Rips, 1975; Osherson et al. 1990)

[Blank sentence]	The person likes (_____).
[Premise sentence]	The person likes wine.
[Conclusion sentence]	The person likes champagne.

• positive premise sentences and negative sentences

[Positive premise sentence]	The person likes wine.
[Negative premise sentence]	The person does not like beer.
[Conclusion sentence]	The person likes champagne.

Figure 1. Examples of the sentence for inductive reasoning

1.2 Inductive Reasoning

To accomplish the aim of developing our search system described in this paper, we apply the inductive reasoning assumption. *Inductive reasoning* is a kind of reasoning that constructs or evaluates inductive arguments. Inductive reasoning is classified into complete induction reasoning and incomplete induction reasoning. In this study, we consider "inductive reasoning" as an incomplete one, and we tried to make new knowledge by inductive reasoning.

In this study, we deal with the fill-in-the-blank objective reasoning shown in upper part of figure 1. In this type of argument, its strength (the likelihood of the objective of conclusion sentences) depends mainly on the entities in each sentence (e.g., "wine", "beer", "champagne") since these sentences share the same basic predicate (e.g., "The person likes

~.” and “The person doesn’t like ~.”). Especially, as premise sentences, we consider the positive premise sentences and negative sentences like shown in the bottom part of figure 1.

1.3 Inductive Reasoning Model

Sakamoto & Nakagawa (2007; 2006) proposed models of inductive reasoning, developed from the internal representation assumption. The assumption explains the way in which the stimuli and the contrasting categories (in this study, corresponding to negative premise entities) are represented. We adopt *feature-based* model to develop our search system, because the feature based model is able to correct the probability values by the psychological experiment, mentioned in chapter 3.

1.4 Problems

There are problems as to this method. A few of the statistic analysis data is included, of which the related strength is low though people will regard it as natural as a feature of the word, and vice versa.

1.5 Purpose of this study

In this study, we aim to develop a search system which allows searching exclusively unnecessary result by association. This search system is based on inductive reasoning.

First, we built a web search system with and without modified result. Second, we modified the probability-based language knowledge structure based on psychological experiments. Finally, we evaluated the system by comparing each search system.

Figure 2: Input interface of positive or negative cases

Figure 3: Output Example of the system

2. Development of the system

We developed the search system based on Sakamoto et al. (2006; 2007)’s inductive reasoning model.

The properties of the corpora are shown in table 1. We used the co-occurrence of noun-adj, ga, wo, ni.

The search system was developed as a PHP web service. An input interface and an output interface are shown in figure 1 and figure 2.

Table.1: the properties of the corpora

Corpora	Mainichi Shimbun (Japanese Newspaper) 1992–2003	
Combination number of nouns		18142
Number of Co-occurrence	noun-adj	3403
	noun-ga(が)	21487
	noun-wo(を)	22832
	noun-ni(に)	10000
	total	58322
Extraction tool	CaboCha (Kudo & Matsumoto 2002)	

3. Modification Experiment of Language Knowledge Structure

3.1 Procedure of the Experiment

We conducted a psychological experiment to modify the probability-based language knowledge structure. The experiment was a hard-copy questionnaire which asked the plausibility of the Japanese relation “noun-adjective”, “noun-‘が (ga)’-verb”, “noun-‘を(wo)’-verb” and “noun-‘に(ni)’-verb”. As the target of the word experiment, 140 general nouns were chosen from words list of 4th grade of Japanese-Language Proficiency Test.

In this experiment, examinees answered the plausibility of the each of the 10 adjectives

Table 2: Modified result of probability language knowledge structure (“荷物”, adjective)

Rank	Non-modified	Prob. value	Modified	Prob. value
1	重い (heavy)	0.665	重い (heavy)	0.124
2	近い (near; close)	0.027	大事な (important; precious;)	0.106
3	必要な (necessary)	0.027	大切な (important)	0.106

which feature the given noun. Where, 10 features had the largest *P(Feature|Noun)* which was calculated by Language Statistical Analysis. Given nouns were 140 general nouns.

Table 3: Combination of the words of the experiments

	1	2	3	4	5
Positive words	勉強 (study)	パーティー (party)	映画 (movie)	手紙 (letter)	スポーツ (sport)
	政治 (politics)	ショー (show program)	ファンタジー (fantasy)	はがき (post card)	スキューバダイビング (scuba diving)
Negative words	テレビ (tele-vision)	儀式 (ceremony)	散歩 (walk; stroll)	通信 (commu-nication)	賭け (bet; gamble)
	遊び (play)	授業 (lecture)	探索 (search; explore)	電話 (tele-phone)	花札 (Japanese playing cards)

3.2 Results of Modification

The example of previous and new modified result, "荷物 (*adective.*)"(荷物 means baggage.) are shown in table 2. The feature "重い (Heavy)", which indicated abnormally higher probability value, was modified to appropriate value compare to other features. In addition, the value of the feature "近い (near; close)", which is hardly associated by human was decreased compared with the value before modification.

4. Comparing Experiment of unmodified/modified system result

We conducted an experiment for comparison among the unmodified and modified system results. We investigated the output result of five cases. We input two positive and two negative words to the system in each case respectively. In the questionnaire, we obtained 15 most plausible output result's appropriateness through 7-point questionnaire. The word list of each of the 5 cases is shown in table 3.

We analyzed the correlation coefficient of modified output and the result of experiment is significantly larger than the one of unmodified output and the result of experiment.

The result of the experiment is shown in table 4. The two cases are significant and larger in $p < .01$ level and one case is significant and larger in $p < .05$ level, while two cases are not significant but show larger correlation. These experimental results indicate that modification makes the system outputting more plausible results for human.

Table 4: The result of the comparing Experiment of unmodified/modified system result

Combination	1	2	3	4	5
Non-modification	-0.532	-0.508	-0.098	-0.139	-0.239
Modification	0.781	0.579	0.579	0.321	0.162
Test between two group's correlation coefficient	p<0.01	p<0.01	p<0.05	n.s.	n.s.

5. Conclusion and Future Study

We developed a search system which can exclude unnecessary things for users from the result by using inductive reasoning. The system enables users to exclude these needless things by inputting key words of positive or negative cases. Then, we conducted psychological experiments to solve these problems, and we corrected the related strength presumed by the language statistically. Comparison experiments showed that the corrected result is more appropriate than uncorrected.

This study is one of the trials that show efforts of language statistical analysis applied to educational systems. Our ultimate goal is to provide practical methods for enhancing education.

References

[1] Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of verbal learning and verbal behavior*, *14*(6), 665-681.

[2] Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological reviews*, *97*(2), 185-200.

[3] Sakamoto, K. & Nakagawa, M. (2007).Risk context effects in inductive reasoning: an experimental and computational modeling study. In B. Kokinov et al. (Eds.), *Context 2007* (pp. 425-438). LNAI 4635, Berlin Heidelberg: Springer-Verlag.

[4] Sakamoto, K. & Nakagawa, M. (2006).The effects of negative premises on inductive reasoning: a psychological experiment and computational modeling study. In *Proc. the CogSci 2009* (pp. 1623-1628).

[5] Kudoh, T., Matsumoto, Y. (2002). Japanese dependency analysis using cascaded chunking. *Transactions of information processing Society of Japan*, *43*(6) 1834-1842. (in Japanese)