Dialog Input Ranking in a Multi-Domain Environment Using Transferable Belief Model

Hong-I Ng
Department of Computer Science
School of Computing
National University of Singapore
nghi@comp.nus.edu.sg

Kim-Teng Lua
Department of Computer Science
School of Computing
National University of Singapore
luakt@comp.nus.edu.sg

Abstract

This paper presents results of using belief functions to rank the list of candidate information provided in a noisy dialogue input. The information under consideration is the intended task to be performed and the information provided for the completion of the task. As an example, we use the task of information access in a multi-domain dialogue system. Currently, the system contains knowledge of ten different domains. Callers calling in are greeted with an open-ended "How may I help you?" prompt (Thomson and Wisowaty, 1999; Chu-Carroll and Carpenter, 1999; Gorin et al., 1997). After receiving a reply from the caller, we extract word evidences from the recognized utterances. By using transferable belief model (TBM), we in turn determine the task that the caller intends to perform as well as any information provided.

1 Introduction

Touch-tone menus are prevalent in call centers for accessing personal records and pre-recorded information. However, it can sometimes be very frustrating when we need to listen to a long list of options. Moreover, the information that we are looking for may not seem to be relevant to any of the given options. Recently, systems that allow people to access information based on spoken inputs have been built. They require a speech recognizer that is trained on a specific set of key words and speech grammars to understand the spoken inputs. Callers are guided through a series of prompts. At each prompt, the callers are supposed to speak out their choices in a way that is easy for the systems to understand. However, new callers may not know what should they say at different prompts and how should they say it. They might have spoken their choices too early, or the way they say it is not encoded in the systems grammar. Thus, we are motivated to work on the problem of accessing information using naturally spoken dialogue. We allow callers to speak in a natural way. Our ultimate aim is to provide the caller with the exact piece of information that s/he is looking for through a series of dialogue interaction. The work reported in this paper is our first attempt toward our ultimate aim, i.e., to determine what the callers want and find out the information the callers have provided that are useful for the task. To achieve that, we use Smets' (1988) TBM.

TBM is the concept used to justify the use of belief functions (BFs), Dempster's rule of conditioning and Dempster's rule of combination to model someone's belief (Smets, 1988). Since early 1980's, BFs have generated considerable interest in the Artificial Intelligence community. In Smets (1999), Denœux (2000) and Zouhal and Denœux (1998), BFs are used to provide sound and elegant solutions to real life problems where some information is missing. As in Bayesian model, given the available evidences, parts of the amount of belief are allocated to each object in our problem domain. However, some evidences might support something other than only one of the various domain objects. In this case, Principle of Insufficient Reason (Smets, 1988) is invoked to decide that the belief mass must be split equally among the domain objects involved. TBM does not evoke this principle and leaves the belief mass allocated to the disjunct of the domain objects involved. Examples of the use of BFs include discriminant analysis using a learning set where classes are only partially known; determine the number of sources in a multi-sensor environment by studying the inter-sensors contradiction and pattern classification. As far as we know, nobody has used BFs to solve problems related to human-computer conversational dialogue. However, we belief that BFs can be applied on problems related to human-computer conversational dialogue, where the recognized utterances contain insertion, deletion and substitution errors. Currently, our multi-domain dialogue system contains knowledge of ten different domains. They are phone directory service (T_1), train schedule inquiry (T_2), flight status inquiry (T_3), travel booking (T_4), Bus Service inquiry (T_5), financial planning (T_6), phone banking (T_7), checking of the employee's account (T_8), concert ticket booking (T_9) and course registration (T_{10}).

Similar works have been reported is the past. However, their main aim is to do call routing instead of information access. Their approaches include the use of a vector-based information retrieval technique (Lee et al., 2000; Chu-Carroll and Carpenter, 1999) /bin/bash: line 1: a: command not found Our domains are more varied, which may results in more recognition errors. In addition, we do not have a training corpus. However, we have a knowledge base that provides partial information based on word evidences. For examples, the occurrence of word evidence *account* indicates that the user wants to access her/his employee's account or bank account, the occurrence of a person name indicates that the user is not checking for a flight status or bus service, the occurrence of word evidence *time* indicates that the user probably wants to check the train schedules or flight status.

Due to space limitation, readers are advised to refer to Smets (1988; 1997; 1989) for more detailed discussions on BFs, combination of BFs, decision making using BFs and TBM.

2 Ranking Information from the Recognized Utterance of Naturally Spoken Input

Our aim is to use TBM in dialogue management. First, TBM is used to rank the information identified from the recognized input. Then, the rank list is used in clarification dialogues if necessary. Otherwise, the best result is treated as the user input. Our experiments are done using Sphinx II speech recognition system (Huang et al., 1992). Using a test corpus of 1977 words, we find that the word recognition accuracy is 54.5%. In our experiments, we use 139 naturally spoken responses to an open-ended "How may I help you prompt" prompt. The callers are told in advance the list of tasks that the system can perform. As notations, let *U* denotes a recognized utterance, *n* the length of *U* in number of words and $w_i, i = 1..n$ the word evidences from *U*.

2.1 Identifying the Intended Task

In this experiment, we show whether TBM can be used to identify the caller's intended tasks. First, we need to identify our problem domain or *frame of discernment*, $\triangle$ (Smets, 1988). For task identification, $\triangle = \{T_j | j = 1..10\}$, i.e., the list of tasks presented in Section 1. $m_{w_i}(T)$, i.e., the basic belief mass (*bbm*) of w_i given to T where $T \in 2^{\triangle}$ is calculated based on the occurrence frequency of word evidence w_i in the knowledge-bases of $T_j, j = 1..10$.

Currently the knowledge base K_j of each $T_j, j = 1..10$ consists of (a) a task specification; (b) information schemas for T_j; and (c) information for task, i.e., the database records, facts and rules, and remote tables and databases used in T_j. A task specification specifies the goal of the task and the list of steps required to attain the goal. Each step is linked to either a basic operation, for examples, fetch some records from a database and ask the caller for information, or a sub-task. Information schemas specify the high-level formats of the information used in T_j. They include database schemas, XML schemas of facts and rules, and format descriptions of some remote tables and databases used in T_j.

We do indexing for each $K_j, j = 1..10$ so that it is easy to calculate the bbm's $m_{w_i}(T)$ where $i = 1..n$ and $T \in 2^{\triangle}$. We then do the following adjustments to make sure that $\sum_{A:A \subseteq \triangle} m_{w_i}(A) = 1$: if

$\sum_{A:A\subseteq\triangle} m_{w_i}(A) > 1$, then the BF m_{w_i} is scaled to one; otherwise, $m_{w_i}(\mathbf{1}_\triangle) = 1 - \sum_{A:A\subseteq\triangle} m_{w_i}(A)$ where $\mathbf{1}_\triangle = \bigvee_{T\in\triangle} T$. $m_{w_i}(\mathbf{1}_\triangle)$ is also called the ignorance value relative to $\triangle$ (Smets, 1988). Larger $m_{w_i}(\mathbf{1}_\triangle)$ implies that it is harder to decide which is the intended task of the caller by looking at evidence w_i. The BF's $m_{w_i}, i = 1..n$ are then combined using Dempster's rule of combination, $m_{w_i w_j}(A) = \sum_{X\wedge Y=A} m_{w_i}(X) m_{w_j}(Y)$ where $X, Y, A \subseteq \triangle$ and $i \neq j$. $m_{w_i w_j w_k}$ is computed by combining $m_{w_i w_j}$ and m_{w_k}. Lastly, $T_j, j = 1..10$ are ranked in descending order according to their pignistic probability measure $betP(T_j) = \sum_{B\subseteq\triangle, T_j\in B} \frac{m(B)}{(1-m(\mathbf{0}_\triangle))|B|}$ with the top of the rank being the most probable target task. Experiment results will be presented in Section 3.1.

2.2 Identifying the Provided Information

In this experiment, we show whether TBM can be used to identify the information provided by the caller in U. Here, the *frame of discernment* $\triangle_{info}$ consists of the objects in the information schemas for a specific task. As in Section 2.1, we use the indices of $K_j, j \in 1..10$ to compute the bbm's of $w_i, i = 1..n$ given to each object disjunct $O_d \in 2^{\triangle_{info}}$. Lastly, we combine the BFs $m_{w_i}, i = 1..n$ and compute the pignistic probability measures of each object $O \in \triangle_{info}$. Experiment results will be presented in Section 3.2.

3 Experiment Results

3.1 Identifying the Intended Task

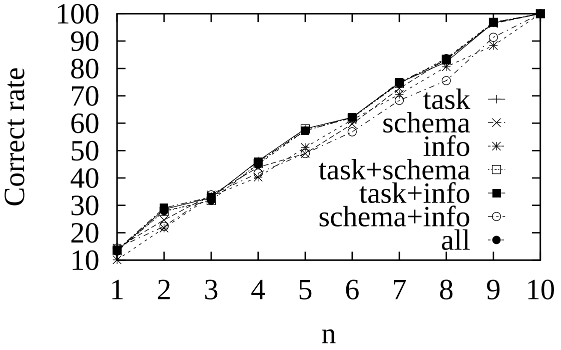

Figure 1: Percentage of time the correct task is included when considering the top n ranked tasks.

Figure 1 shows the results of selecting top-*n*-tasks in the ranked list of $T_j, j = 1..10$. The labels *task*, *schema* and *info* denote that only knowledge in the task specifications, information schemas and basic information respectively are included in the calculation of bbm's. '+' denotes a combination of some and *all* denotes the combination of all. The graphs show that we obtain the best ranking of candidate tasks when knowledge from task specifications and information schemas are used to calculate the BF's. This is intuitive because callers will often say her/his goal and mention the name of the piece of information s/he's looking for, e.g., "I want to buy a *movie ticket* please."

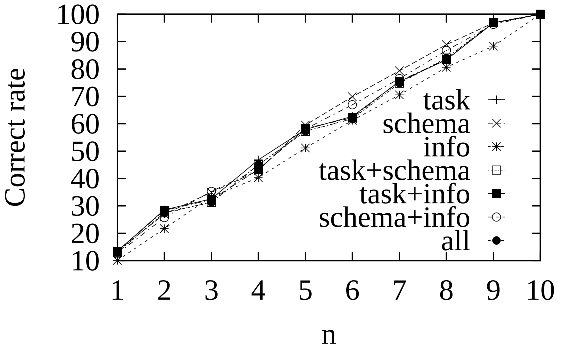

Figure 2: Percentage of time the correct task is included when considering the top n ranked tasks, taking similar words into considerations.

Next, we examine the result of taking similar words into considerations. This is because callers may use words different from those occurring in our knowledge base. Thus, for each word evidence w_i in U, we use WordNet (Fellbaum, 1997) to look for similar words $w'_{ij}, j = 1..p$ in our knowledge base. For each $w'_{ij}, j = 1..p$, we calculate the BF $m_{w'_{ij}}$ as discussed in Section 2.1. This time, we also multiply the bbm's in $m_{w'_{ij}}$ by the distance measure between w'_{ij} and w_i. The distance measures fall in the range [0:1]. These results are shown in Figure 2. Again, the results show that we obtain the best ranking of candidate tasks when knowledge from task specifications and information schemas are used to calculate the BF's. However, there is a decrease in correct rate when only the best (-6.25%) and 2-best (-1.58%) tasks in the ranked list are used to allow

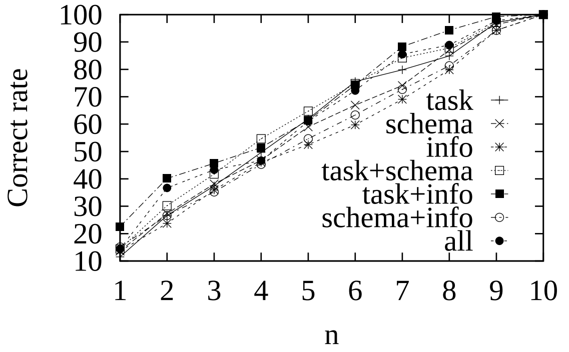

Figure 3: Percentage of time the correct task is included when considering the top n ranked tasks, taking similar words and correlation measures into consideration.

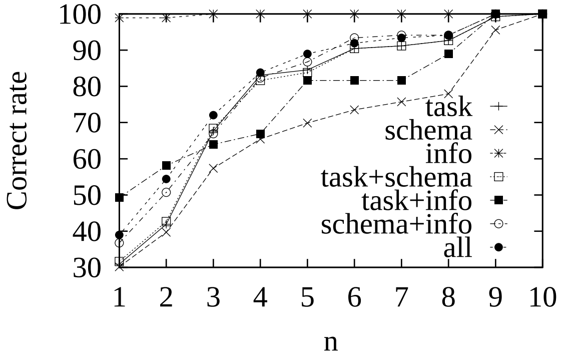

Figure 4: Percentage of time the correct task is included when considering the top n ranked tasks using dialogue transcripts, similar words and correlation measures.

the callers to select. The correct rate is increased only when more than 2 top-ranking tasks are used for callers' selection, i.e., 4.38%, 1.32%, 2.66% and 12.32% when *n* = 3, 4, 5 and 6 respectively.

From the results, we found that some words occur commonly across multiple domains. This phenomena is common in problems related to natural language processing. To alleviate the problem, we have used words that only occur commonly in few domains. We use correlation coefficient (Ng et al., 1997) to measure the correlations of all words to all domains. After that, we scale the correlation measures to 1. In calculating the bbm's, we multiply the original bbm's with the corresponding correlation measures. Figure 3 shows the results when similar words and correlation measures are considered in the calculation of BF's. This time, the results show that we obtain the best ranking of candidate tasks when knowledge from task specifications and basic information are used to calculate the BF's. In addition, there is a 67.31% improvement when the top task in the ranked list is taken as the caller's intended tasks. When top-*n* tasks are used for callers' selection, the improvements are 40.5%, 29.53% and 16.76% for *n* = 2, 3 and 4 respectively.

For the purpose of comparison, we show the results of task identification based on dialogue transcripts, similar words and correlation measures in Figure 4. The results show that with the use of only basic information in the calculation of BF's, a result of 99.1% can be achieved by select the top task in the ranked list. Thus, when the word accuracy of the speech recognizer is high, basic information is sufficient to identify the callers' intended tasks. Otherwise, knowledge from task specifications and information schema are required in target task identifications. We have shown that TBM can be used for task identification in a noisy and multi-domain environment. It would be interesting to compare these results when we have enough corpus to train a vector-based task identifier.

3.2 Identifying the Provided Information

Figure 5 shows the percentage of time the correct information is included in the top-*n* selected information after they have been sorted according to their pignistic probability measures. *SR-best-1* (*SR-best-2*) indicates that the best (respectively, two best) speech recognition results are used for information identification. The results show that there is a 14.25% (10.54%) improvement when the best (respectively, two best) speech recognition results are used for information identification. '*Transcript*' indicates that the dialogue transcripts are used for information identification. The results show that there is an average of 63.79% information lost between '*transcript*' and '*SR-best-2*'.

4 Conclusion

A new naturally spoken dialogue processing based on the TBM has been presented. This approach

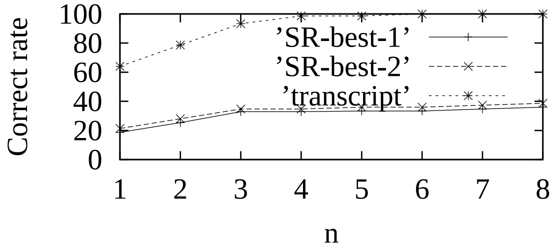

Figure 5: Correct identification rate using the top *n* information in the rank.

can be viewed as looking for evidences from noisy speech inputs to identify the tasks that the callers want to perform and the information that they have provided. Our experiments are tested on a multi-domain environment. The speech recognizer that we use has a word accuracy of around 55%. The experiment results show that there is some initial success in using TBM to aid in task and information identification when the recognized input is noisy.

In order to improve users' satisfaction, we are looking into dialogue processing methods that are able to improve the results of task and information identification. In particular, instead of using word evidences from the recognized inputs, we are looking into the use other evidences such as phonemes. We are also looking into dialogue strategies that are able to collaborate with the callers to correct the identified information. In particular, if the ignorance value $m(\mathbf{1}_\triangle)$ is high, our system should employ system initiative strategies to disambiguate the identified information. If $m(\mathbf{0}_\triangle)$ is high, which means that the evidences do not point strongly to any object in $\triangle$, then our system should employ system initiative strategies to learn new task-related information. If both $m(\mathbf{0}_\triangle)$ and $m(\mathbf{1}_\triangle)$ are low, out system can employ a mixed initiative dialogue strategy.

Acknowledgments

Our thanks go to the undergraduate students who have contributed their valuable time to help us in the recordings without asking for any rewards.

References

Chu-Carroll, Jeniffer and Bob Carpenter. 1999. Vector-based natural language call routing. *Computational Linguistics*, 25(3):361–388.

Denœux, Thierry. 2000. A neural network classifier based on Dempster-Shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(2):131–150, March.

Fellbaum, Christiane (Ed). 1997. *WordNet: An Electronic Lexical Database*. Imprint Cambridge, Mass: MIT Press.

Gorin, Allen L., Giuseppe Riccardi and Jeremy H. Wright. 1997. How may I help you? *Speech Communication*, 23:113–127.

Huang, Xuedong, Fileno Alleva, Hsiao-Wuen Hon, Mei-Yuh Hwang, Ronald Rosenfeld. 1992. The SPHINX-II speech recognition system: an overview. *Computer Speech and Language*, 7(2):137–148.

Lee, Chin-Hui, Bob Carpenter, Wu Chou, Jennifer Chu-Carroll, Wolfgang Reichl, Antoine Saad and Qiru Zhou. 2000. On natural language call routing. *Speech Communication*, 31(4):309-320, Aug.

Ng, Hwee Tou, Goh Wei Boon and Low Kok Leong. 1997. Feature selection, perceptron learning, and a usability case study for text categorization. In *Proceedings of the 20th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 67-73. Philadelphia, Pennsylvania, USA.

Smets, Philippe. 1999. Practical uses of belief functions. *Uncertainty in Artificial Intelligence: Proceedings of the Fifteenth Conference (UAI-1999)*, Morgan Kaufmann Publishers, San Francisco, CA, 612–621.

Smets, Philippe. 1989. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence 5*. Henrion M., Shachter R. D., Kanal L. N. and Lemmer J. F. (Eds). North Holland, Amsterdam, 29–40.

Smets, Philippe. 1988. Belief functions. *Non-standard Logic for Automated Reasoning*. P. Smets, A. Mamdani, D. Dubois, and H. Prade (Eds). New York: Academic, 252–286.

Smets, Philippe. 1997. The axiomatic justification of the transferable belief model. *Artificial Intelligence*, 92:229–242.

Thomson, David L. and Jack J. Wisowaty. 1999. User confusion in natural language services. In *Proc. ESCA Workshop on Interactive Dialogue in Multi-Modal Systems*, Kloster Irsee, Germany, June, 189–196, keynote address.

Zouhal, Lalla Merieme and Thierry Denœux. 1998. An evidence-theoretic *k*-NN rule with parameter optimization. *IEEE Transactions on Systems, Man and Cybernetics – Part C*, 28(2):263-271.