Simulated Spoken Dialogue System Based on IOHMM with User History

Changliang Li*, Bo Xu, XiuYing Wang, WenDong Ge, and HongWei Hao

Interactive Digital Media Technology Research Center (IDMTech),
Institute of Automation, Chinese Academy of Sciences, Beijing, China
{Changliang.li,boxu,xiuying.wang,wendong.ge,
hongwei.hao}@ia.ac.cn

Abstract. Expanding corpora is very important in designing a spoken dialogue system (SDS). In this big data era, data is expensive to collect and there are rare annotations. Some researchers make much work to expand corpora, most of which is based on rule. This paper presents a probabilistic method to simulate dialogues between human and machine so as to expand a small corpus with more varied simulated dialogue acts. The method employs Input/output HMM with user history (UH-IOHMM) to learn system and user dialogue behavior. In addition, this paper compares with simulation system based on standard IOHMM. We perform experiments using the WDC-ICA corpus, weather domain corpus with annotation. And the experiment result shows that the method we present in this paper can produce high quality dialogue acts which are similar to real dialogue acts.

Keywords: SDS, Corpora, UH-IOHMM, dialogue acts.

1 Introduction

Recently, SDSs have been developing rapidly. For example, Apple introduced "Siri", an intelligent personal assistant and knowledge navigator which works as an application for Apple Inc.'s iOS. In addition, Android phones have employed speech-activated "Voice Actions", which can be used to call your contacts, get directions, send messages, and perform a number of other common tasks and so on[1]. Cambridge designed CamInfo system, which offers service of travel information to people, based on partially observed Markov decision process (POMDP) [2] [3].

In spite of its fast developing, SDS still remains challenging. Among the changing, insufficient available data with annotation is the biggest bottleneck. There are no corpora big enough to sufficiently explore the vast space of possible dialogue states and strategies [3]. In addition, Data is expensive to collect and annotate. So, it is vital to expand corpora in designing a SDS.

* Corresponding author.

G. Zhou et al. (Eds.): NLPCC 2013, CCIS 400, pp. 83–92, 2013.

Many research efforts have been undertaken in expanding corpora, including rule-based and data-driven approaches.

The point of rule-based intention simulation approach is that the developer can create many kinds of rules, which can generate variant dialogue acts given some certain information (Chung, 2004, López-Cózar et al., 2006 and López-Cózar et al., 2003) [4] [5]. In addition, Schatzmann et al. proposed an agenda-based user simulation technique for bootstrapping a statistical dialog manager, the main feature of which is that it is without access to training data (Schatzmann et al., 2007a). It generates user dialogue acts based on a full representation of the user goal and a stack-like user agenda[6].

The main feature of data-driven approach is to use statistical methods to simulate users' dialogue acts given corpora. The "bigram" model of dialog is employed in earlier studies. Its distinguish feature is the simulated user input is decided only by the previous system utterance (Eckert et al., 1997) [7]. One of the main advantages of the approach is that it is simple and another advantage is that it is independent on domain and language. There also remains improvement, for example, Levin et al. modified the bigram model to make a more realistic degree of conventional structure in dialog (Levin et al., 2000). In order to solve the problem of lack of the goal consistency in the model proposed by Levin, Scheffler and Young introduced the use of a graph-based model (Scheffler and Young, 2000 and Scheffler and Young, 2001). The model is goal directed. The main characteristic of this model is that it defined a goal as a specification of the dialog transaction that the user wants to accomplish[8] [9] [10] [11]. Cuayahuitl, Renals, Lemon, and Shimodaira (Cuayahuitl et al., 2005) presents a method for dialogue simulation based on Hidden Markov Models (HMMs). Their method generates system and user actions [2] [12]. It expands a small corpus of dialogue data with more varied simulated conversations.

Previous works expand the corpora based on rule or probability [2]. However, they mostly focus on the system turns and neglect the user history information. This brings the problem that the expanded dialogues acts are not as similar as real dialogues acts. The simulated user often repeat dialogue acts. This paper presents a probabilistic method to simulate dialogues between human and machine based on UH-IOHMM so as to expand corpora with more varied simulated dialogue acts. The dialogues acts generated through the method presented in this article are more close to real dialogues and in task-orient domain, such as weather information inquiry domain, less turns are needed before the task is satisfied[2] [13].

The rest of this article proceeds as follows. In section 2 we present some related knowledge and technology including SDS structure and typical simulation structure. In Section 3 we present simulation system based on UH-IOHMM. In section 4 we design the experiment on WDC-ICA corpora, which consists 100 dialogue sections with annotation in the domain of weather information, and analyze the experiment result. Finally, we summarize our conclusions and point future work.

2 Related Knowledge and Technology

2.1 SDS Structure

Typically, SDSs are composed of five components: automatic speech recognition (ASR); Natural language understanding (NLU); Dialogue manager (DM), Natural Language generation (NLG); Speech generation, such as text-to-speech (TTS) [1]. Among these units, the central module of any SDS is DM, which is in charge of the course of the interaction with user: it receives the semantically parsed representation of the user input and generates an appropriate high-level representation of the next system action [1]. DM mainly includes three parts: a dialogue model representing state information such as the user's goal, the user's last dialogue act and the dialogue history; a policy which selects the system's responses based on the inferred dialogue state; and a cumulative reward function which specifies the desired behavior of the system[1] [2] [3]. Figure 1 shows the structure of spoken dialogue system.

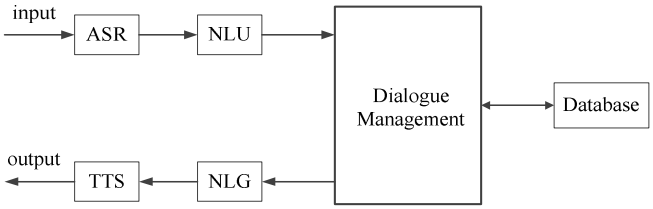

Fig. 1. The structure of spoken dialogue system

2.2 Typical Simulation Structure

A typical probabilistic dialogue simulation model often consist two main modules: system module and user module. The former's role is to control the flow of the conversation; the latter's role is to generate user's dialogue acts based on conditional probabilities. However, a training corpus with annotation is required, which is used to acquire knowledge and train system and user models [2]. The simulation model is shown in figure 2.

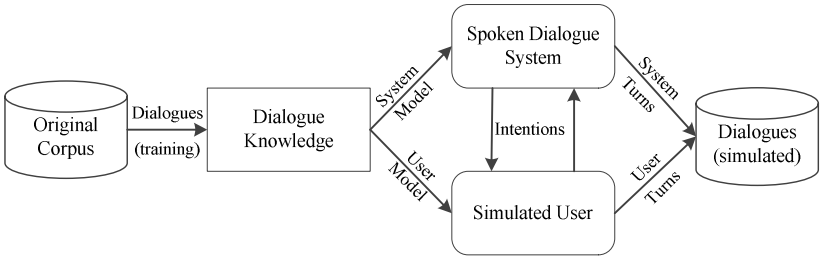

Fig. 2. The structure of simulation system

3 Simulation System

3.1 UH-IOHMM Based Simulation System

Our work in this article is to improve IOHMM based simulation system proposed by Cuayahuitl, Renals, Lemon, and Shimodaira. The proposed model is based on the assumption that a user's action depends on not only the previous system response, but also users' history information [13] [14].

A dialogue simulation system based on UH-IOHMM is shown in Figure 3, where empty circles represent visible states; the lightly shaded circles represent observations; the dark shaded circles represented user responses; A represents transfer matrix; B represents confusion matrix; U represents user action transfer matrix[2] [15]. The model is characterized by a set of visible states $S=\{S_1,S_2,\ldots,S_N\}$ which correspond to system turns, and a set of observations $V=\{v_1,v_2,\ldots,v_M\}$ which represent system action set [3] [16]. We employ q_t to represent the state at time t. The user responses are represented using a set of user intentions $H=\{H_1,H_2,\ldots,H_L\}$ and the user action at time t is denoted using u_t [2] [3].

UH-IOHMM we presented in this paper gathers together the next state transition q_{t+1} on the current state q_t and current user response u_t as conditions. The state transition probability is represented as $P(q_{t+1}\mid q_t,u_t)$, and the user intentions is conditioned not only on the system intention and symbol observed at time t, but also on the user's history acts represented as $P(u_t\mid q_t,c_t,u_{t-1},u_{t-2}\ldots u_{t-m})$, where m represents the steps we trace back the user's history acts. Apparently, the bigger m is, the more history information can be considered. But at the same time the computation complexity becomes bigger too. In this paper, in order to balance the history information and computation complexity, we select m as 1. Figure 3 illustrates the structure of dialogue simulation system based on UH-IOHMM.

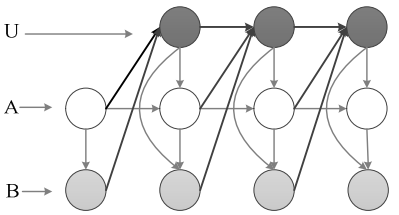

Fig. 3. UH-IOHMM based simulation system

Because there are many dialogue turns in corpora, some of which may repeat, we cut the WDC-ICA corpora into some sub goals. And each sub goal is modeled as a UH-IOHMM. We train each separate model for each sub goal. So we don't need to

train a single giant IOHMM for simulating complete dialogues. The full corpora are not viewed as a whole set, but as a bag of sub goals [2] [3]. The next goal is decided by the conditional probability $P(g_n \mid g_{n-1})$ [2] [17]. Figure 4 shows the language model about sub goals [2].

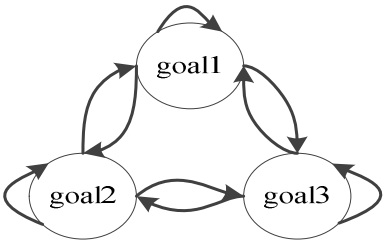

Fig. 4. Language model of sub goals

The dialogue simulation algorithm using UH-IOHMM is shown in figure 5. Simulate UH-IOHMM function generates a sequence of system intentions and user intentions.

```
1. function Simulate UH-IOHMM
2. t ← 0
3. q_t ←random system turn from π
4. c_t ←random system intention from
P(c_t | q_t)
5.   loop
6.     print c_t
7.     u_t ←random user intention from
P(u_t | q_t, c_t, u_{t-1})
8.     print u_t
9.     q_t ←random system turn from
P(q_{t+1} | q_t, u_t)
10.     if q_t = q_n then return
11.     else t ← t +1
12.     c_t ←random system intention from
P(c_t | q_t, u_t)
13.   end
14. end
```

Fig. 5. Simulation algorithm

4 Experimental Design and Result

We employ WDC-ICA corpus, which consists 100 dialogue sections with annotation in the domain of weather information. The annotation consists of the system action and user intention as well as parameters needed in the domain like location and time and so on. We assume that there is no ASR error [2] [18] [19]. We use affinity propagation algorithm [20], considering its simplicity, general applicability, and performance, to classifier system turns and user dialogues acts into states used in UH-IOHMM. For example, the system turns are classified four different states and the user intention are classified as seven different states.

Table 1 shows the result after classifying, where 0 represents that the parameter lacks, and 1 represents that the parameter is filled. For example, the state "s1" represents that the system asked the user to offer time and location information, because both parameters lack.

Table 1. States and action in weather domain corpora

s1	Inquiry (time=0, location=0)
s2	Inquiry (time=0, location=1)
s3	Inquiry (time=1,location=0)
s4	Inquiry weather(time=1, location=1)
s5	Response (time=0, location=0)
s6	Response (time=0, location=1)
s7	Response (time=1,location=0)

a1	Inquiry (time, location)
a2	Inquiry (time)
a3	Inquiry (location)
a4	inquiry weather

Due to the limitation of corpora size, there may be some intentions that can't appear in the corpora. So considering unseen entries, we use Good-Turing algorithm, which provide a simple estimate of the total probability of the objects not seen, to smooth the probability distributions.

It is hard to evaluate the simulated dialogues due to the fact the flexibility of dialogues acts. The quality of the simulated dialogues has been assessed using a variety of different direct or indirect evaluation methods. We use dialogue length to evaluate the result, which computers the average number of turn per dialogue, giving a rough indication of agreement between two sets of dialogues. We train the corpora and generate 10^5 dialogues based on IOHMM and UH-IOHMM proposed in this paper. Fragments of a simulated dialogue generated by IOHMM and UH-IOHMM are shown in figure 6 at left and right side respectively. We can see that the dialogue generated based on UH-IOHMM is more efficient than that based on IOHMM.

```
dialogue 1: sentence Number = 2
请问昌平十七点天气是不是很好啊？ (s1)
调用APP。 (a4)

dialogue 2: sentence Number = 4
请问林芝天气温度怎么样？ (s3)
请问你问的是什么时候的天气？ (a2)
星期四夜间的天气。 (s6)
调用APP。 (a4)

dialogue 3: sentence Number = 4
请问天气是哪样的？ (s4)
请问你问的是什么时候哪里的天气？ (a1)
江门星期天的天气。 (s5)
调用APP。 (a4)

dialogue 4: sentence Number = 4
请问中午十一点天气温度是多少？ (s2)
请问你问的是哪里的天气？ (a3)
淮安的天气。 (s7)
调用APP。 (a4)
```

```
dialogue 5: sentence Number = 10
请问十九点天气冷不冷？ (s2)
请问你问的是哪里的天气？ (a3)
十九点的天气。 (s6)
请问你问的是什么时候哪里的天气？ (a1)
十九点的天气。 (s6)
请问你问的是什么时候的天气？ (a2)
日照的天气。 (s7)
请问你问的是什么时候的天气？ (a2)
日照十九点的天气。 (s5)
调用APP。 (a4)

dialogue 6: sentence Number = 12
请问天气怎么样？ (s4)
请问你问的是什么时候哪里的天气？ (a1)
巴中晚上九点的天气。 (s5)
请问你问的是什么时候的天气？ (a2)
巴中的天气。 (s7)
请问你问的是哪里的天气？ (a3)
巴中晚上九点的天气。 (s5)
请问你问的是什么时候哪里的天气？ (a1)
巴中的天气。 (s7)
请问你问的是什么时候的天气？ (a2)
晚上九点的天气。 (s6)
调用APP。 (a4)
```

Fig. 6. Fragments of a simulated dialogue generated by IOHMM (left) and UH-IOHMM (right)

We select 10000 dialogue sections and add up the dialogue turns in each dialogue sections. From figure 7, where horizontal axis represents the number of dialogue turns in each dialogue section while vertical axis represents the number of dialogue sections, we can make a safe conclusion that: based on WDC-ICA corpus, the method proposed in this paper can satisfy users' need in less turns than the state of the art method which is based on IOHMM.

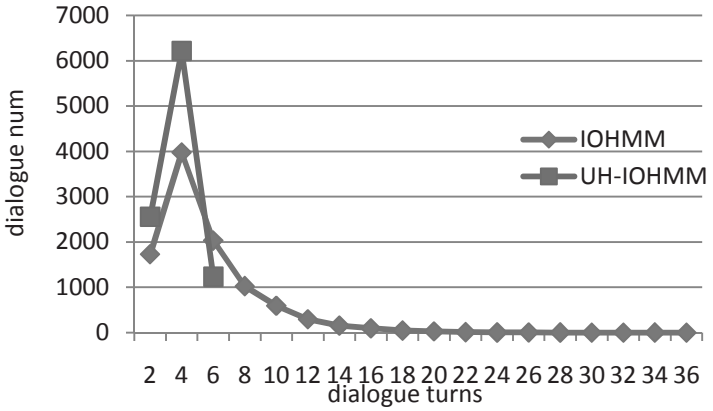

Fig. 7. Dialogue numbers of different dialogue turns

From figure 8, where the vertical axis represents the average turns of dialogues generated through both simulations systems, we can see that the simulated dialogues based on IOHMM take an average of 5.5 turns to complete the weather information inquiry task, while the simulated dialogues based on UH-IOHMM we proposed in this article takes an average of 3.7 turns to complete the task. Apparently, the dialogues generated by simulate dialogue system based on UH-IOHMM is both efficient and more similar to human's real dialogue behavior.

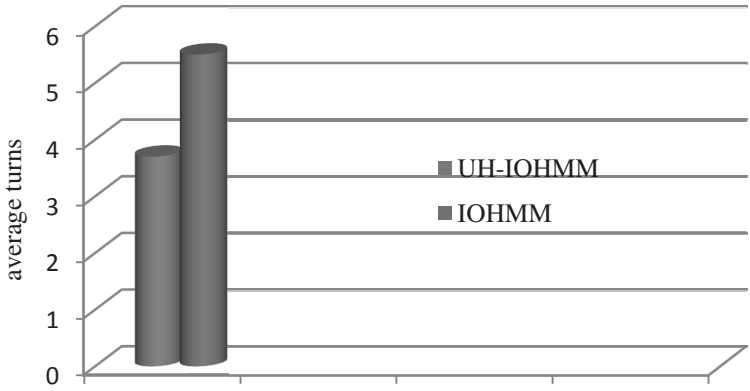

Fig. 8. Average turns of IOHMM and UH-IOHMM

5 Conclusion and Future Work

This paper focused on the problem of limitation of data scale while building SDS. To cope with this problem, we presented a method based on UH-IOHMM to simulate dialogue between human and machine, so as to expand corpora. This method learnt a system model and a user model: the system model is a probabilistic dialogue manager that models the sequence of system intentions, and the user model consists of conditional probabilities of the possible user responses. Due to the fact that all the possible system and user intentions may occur in each state, more exploratory dialogues can be generated than observed in the real data. We compared the proposed model with IOHMM model. Our experiments revealed that the UH-IOHMM models obtained very similar performance, clearly outperforming random dialogues, and are considered to be close to reality.

Although the method expanded the corpora by simulating the dialogue behavior between human and machine, there remained much to improve. First of all, it would be more efficient the problem of computation complexity can be solved. In addition, the cluster algorithm is still not efficient enough, and the result of cluster brings big influence to the simulation result. If we can find a more accurate cluster algorithm to replace the AP algorithm employed in this paper, the simulation will be trained as a more effective way. These issues will be addressed in future work.

Acknowledgement. This work is partly supported by National Program on Key Basic Research Project (973 Program) under Grant 2013CB329302. The work described in this paper represents the combined efforts of many people.

References

1. Lemon, O., Pietquin, O.: Data-Driven Methods for Adaptive Spoken Dialogue System. Springer (2012)
2. Cuayahuitl, H., Renals, S., Lemon, O., Shimodaira, H.: Human-computer dialogue simulation using hidden markov models. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (2005); Chung, G.: Developing a Flexible Spoken Dialog System Using Simulation. In: Proc. ACL, pp. 63–70 (2004)
3. Schatzmann, J., Weilhammer, K., Stuttle, M., Young, S.: A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. The Knowledge Engineering Review 21, 97–126 (2006)
4. López-Cózar, R., De la Torre, A., Segura, J.C., Ru-bio., A.J.: Assessment of dialogue systems by means of a new simulation technique. Speech Communication 40(3), 387–407 (2003)
5. Ramón, L.-C., Callejas, Z., Mctear, M.: Testing the performance of spoken dialogue systems by means of an artificially simulated user. Artif. Intell. Rev. 26(4), 291–323 (2006)
6. Schatzmann, J., Thomson, B., Young, S.: Error simulation for training statistical dialogue systems. In: IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU 2007, pp. 526–531 (2007a)
7. Eckert, W., Levin, E., Pieraccini, R.: User modeling for spoken dialogue system evaluation. In: Proceedings of the 1997 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 80–87 (1997)
8. Scheffler, K., Young, S.: Probabilistic simulation of human-machine dialogues. In: Proc. of ICASSP, vol. 2, pp. 1217–1220 (2000)
9. Scheffler, K., Young, S.: Corpus-based dialogue simulation for automatic strategy learning and evaluation. In: Proc. NAACL Workshop on Adaptation in Dialogue Systems, pp. 64–70 (2001)
10. Levin, E., Pieraccini, R., Eckert, W.: A stochastic model of human-machine interaction for learning dialog strategies. IEEE Trans. on Speech and Audio Processing 8(1), 11–23 (2000)
11. Schatzmann, J., Weilhammer, K., Stuttle, M., Young, S.: A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. The Knowledge Engineering Review 21, 97–126 (2006)
12. Fan, L., Yu, D., Peng, X., Lu, S., Xu, B.: A Spoken Dialogue System Based on FST and DBN. In: Zhou, M., Zhou, G., Zhao, D., Liu, Q., Zou, L. (eds.) NLPCC 2012. CCIS, vol. 333, pp. 34–45. Springer, Heidelberg (2012)
13. Kearns, M., Mansour, Y., Ng, A.Y.: Sparse sampling algorithm for near optimal planning in large markov decision processes. In: Proceedings of the Sixteenth International Joint Conference on Articial Intel ligence Stockholm (1999) (to appear)
14. Chen, K.: Boosting input/output hidden Markov models for sequence classification. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3611, pp. 656–665. Springer, Heidelberg (2005)
15. Litman, D.J., Pan, S.: Designing and Evaluating an Adaptive Spoken Dialogue System. User Modeling and User-Adapted Interaction 12, 111–137 (2002)

16. Chiappa, S., Bengio, S.: HMM and IOHMM Modeling of EEG Rhythms for Asynchronous BCI Systems. In: ESANN 2004 Proceedings-European Symposium on Artificial Neural Networks Bruges, Belgium, April 28-30 (2004)
17. Bengio, Y., Frasconi, P.: Input-Output HM:M' s for Sequence Processing. IEEE Transactions on Neural Networks 7(5) (September 1996)
18. Cuayahuitl, H., Renals, S., Lemon, O., Shimodaira, H.: Human-computer Dialogue Simulation using Hidden Markov Models. In: ASRU (2005)
19. Bengio, Y., Frasconi, P.: An Input Output HMM Architecture. In: Tesauro, G., Touretzky, D., Leen, T. (eds.) Advances in Neural Information Processing Systems, vol. 7, pp. 427–434. MIT Press, Cambridge (1995)
20. Young, S., Gasic, M., Thomson, B., Williams, J.D.: POMDP-based Statistical Spoken Dialogue Systems: A Review. In: Proc. IEEE, vol. X(X) (January 2012)
21. Kumaravelan, G., Sivakumar, R.: Simulation of Dialogue Management for Learning Dialogue Strategy Using Learning Automata. IEEE (2009)
22. Polifroni, J., Chung, G., Seneff, S.: Towards the Automatic Generation of Mixed-Initiative Dialogue Systems fromWeb Content. EUROSPEECH (2003)
23. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science 315, 972–976 (2007)