Training Bayesian networks for realistic man-machine spoken dialogue simulation

Olivier Pietquin, Stéphane Rossignol, and Michel Ianotto
IMS research team, Supélec Campus de Metz, F-57070 Metz, France
olivier.pietquin@supelec.fr
Home page: http://ims.metz.supelec.fr/~pietquin

Abstract. Data collection and annotation are generally required to design or assess spoken dialogue systems. Yet, this is a very time consuming and expensive process. For these reasons, user simulation has become an important trend of research in the field of spoken dialogue systems. The general problem of user simulation is thus to produce as many as necessary natural, various and consistent interactions from as few data as possible. In this paper, we propose a user simulation method based on Bayesian networks (BN) that is able to produce consistent interactions in terms of user goal and dialogue history. The model as been introduced in previous work but parameters were hand-tuned and it was assessed in the framework of automatic learning of optimal dialogue strategies. In this paper, the BN is trained on a database of 1234 human-machine dialogues in the TownInfo domain (a tourist information application). Experiments with a state-of-the-art dialogue system (REALL-DUDE/DIPPER/OAA) have been realized and results in terms of dialog statistics are presented.

1 Introduction

Tests and validation of spoken dialogue systems (SDS) require interactions between released systems and real human users which are very expensive and time consuming. For this reason, user simulation have become an important trend of research during the last decade. User simulation can be used for performance assessment [Eckert et al., 1997] [López-Cózar et al., 2006] or for optimisation purpose [Levin et al., 2000] [Pietquin and Dutoit, 2006a] [Schatzmann et al., 2007]. Dialogue simulation can occur at several levels of description. In this work, simulated interactions will take place at the intention level (see [Eckert et al., 1997] [Levin et al., 2000] [Pietquin and Dutoit, 2006a] [Schatzmann et al., 2007]) and not at the signal level as proposed in [López-Cózar et al., 2006]. Simulation will be modeled as dialogue acts exchanges between a Simulated User (SU) and a Dialogue Manager (DM). This allows error modelling of all the parts of the system, including speech recognition [Pietquin and Renals, 2002] and understanding [Pietquin and Dutoit, 2006b].

This paper relies on previous work [Pietquin and Dutoit, 2006a]. Bayesian networks were used but parameters were hand-tuned and assessment was done
in the framework of automatic learning of optimal dialogue strategies. This short paper reports results obtained on a trained BN in terms of statistics measurements. The BN is trained on a database of 1234 human-machine dialogues. Experiments with a state-of-the-art dialogue system are reported.

The considered domain is the TownInfo domain, a tourist information task. The task is described by attribute-value pairs and consists in retrieving information about restaurants in a given city. This can be considered as a slot filling task where we consider three different slots: “food”, “price range” and “area”. Possible values for these slots are provided in Table 1. The possible dialogue acts are “inform”, “confirm” and “close”.

Table 1. Slots in the task, and corresponding possible values

<table>
<thead>
<tr>
<th>Food</th>
<th>Range price</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>“italian”</td>
<td>“moderate”</td>
<td>“central”</td>
</tr>
<tr>
<td>“indian”</td>
<td>“expensive”</td>
<td>“north”</td>
</tr>
<tr>
<td>“chinese”</td>
<td>“cheap”</td>
<td>“south”</td>
</tr>
</tbody>
</table>

2 Description of the Model

The BN in use is as described in [Pietquin and Dutoit, 2006a]. Figure 1 shows the SU for this application. The node AS corresponds to the system act. The nodes SYSTEM correspond to the slots concerned by the system act. The nodes KNW correspond to the knowledge the SU has concerning the state of the dialogue. There is a knowledge value for each of the slots in the task. Three levels of knowledge are considered: low, medium and high given that the concerned slot was already provided 0, 1 or several times by the SU. The nodes GOAL correspond
to the user goal. The **goal values** indicate the value given to each slot in the goal. The nodes **inform** and **inform values** correspond to the user act “inform” and its attributes. **Inform** correspond to the slots transmitted in the SU utterance, and **inform values** correspond to the set of values associated to the transmitted slots. The nodes **confirm** and **confirm values** correspond to the user act “confirm” or “negate”. **Confirm** correspond to the slots transmitted in the SU utterance, and **confirm values** correspond to the set of values (that is to say answer yes or answer no) associated to the transmitted slots. The node close indicates whether the Simulated User decides or not to close the dialogue.

## 3 Experiments

Two versions of the BN are compared hereafter. A heuristic BN where the parameters are set by an expert and a trained BN. The parameters of this last BN are learnt from a database described in details in [Williams and Young, 2007] and containing 1234 dialogues. Simulation are obtained by interaction between one of the SU and a state-of-the-art dialogue system (REALL-DUDE/DIPPER/OAA) [Bos et al., 2003]. 1000 dialogues are simulated. In table 2, are presented the results obtained using the trained BN. 1000 dialogues have been simulated. The dialogues last quite more turns than when using the heuristic BN. This can not be seen considering the mean number of turns, but considering the percentage of dialogues which needed exactly four turns to reach their end: this percentage drops from 93.2 % to 58.0 %. However, very promisingly, the mean number of turns and the percentage of dialogues which needed less than nine turns to reach their end are quite better using the trained BN, and furthermore it can be noticed that very long dialogues (more than nine turns), indicating some deep misunderstanding between the DM and the SU, have completely disappeared. This allows thinking that the dialogues obtained with the trained BN are more natural, at least from a DM point of view, than the dialogues obtained with the heuristic BN. When looking at the average number of turns per slots (Table 3), it appears that the trained BN reaches performances that are closer to those measured in the database.

### Table 2. Statistics obtained after simulation of 1000 dialogues

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>max</th>
<th>min</th>
<th>4 turns</th>
<th>&lt; 9 turns</th>
</tr>
</thead>
<tbody>
<tr>
<td>heuristic BN</td>
<td>4.969</td>
<td>21</td>
<td>4</td>
<td>93.20 %</td>
<td>93.40 %</td>
</tr>
<tr>
<td>trained BN</td>
<td>4.577</td>
<td>9</td>
<td>4</td>
<td>58.00 %</td>
<td>99.90 %</td>
</tr>
</tbody>
</table>

### Table 3. Number of turns needed per slot

<table>
<thead>
<tr>
<th></th>
<th>heuristic BN with longest dialogues</th>
<th>heuristic BN only shortest dialogues</th>
<th>trained BN</th>
<th>database</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean number of turns per slot</td>
<td>1.656333</td>
<td>1.398672</td>
<td>1.525667</td>
<td>1.566179</td>
</tr>
</tbody>
</table>
4 Conclusion

This paper shows that the BN-based user simulator described in previous work can be trained on actual dialogue examples. Measures of simple statistics on simulated dialogues are encouraging. More work is required to analyse the behaviour of the simulated user and more complex training methods (missing data) have to be experimented.

Acknowledgement

This work is supported by the CLASSIC (Computational Learning in Adaptive Systems for Spoken Conversation) European project, Project Number 216594 funded under EC FP7, Call 1 (http://www.classic-project.org/).

References


