Research Article

A Neural Network Approach to Intention Modeling for User-Adapted Conversational Agents

David Griol[1] and Zoraida Callejas[2]

[1]*Department of Computer Science, Carlos III University of Madrid, Avenida de la Universidad 30, 28911 Leganés, Spain*
[2]*Department of Languages and Computer Systems, University of Granada, CITIC-UGR, C/ Pdta. Daniel Saucedo Aranda s/n, 18071 Granada, Spain*

Correspondence should be addressed to David Griol; dgriol@uc3m.es

Spoken dialogue systems have been proposed to enable a more natural and intuitive interaction with the environment and human-computer interfaces. In this contribution, we present a framework based on neural networks that allows modeling of the user's intention during the dialogue and uses this prediction to dynamically adapt the dialogue model of the system taking into consideration the user's needs and preferences. We have evaluated our proposal to develop a user-adapted spoken dialogue system that facilitates tourist information and services and provide a detailed discussion of the positive influence of our proposal in the success of the interaction, the information and services provided, and the quality perceived by the users.

1. Introduction

Dialogue systems are artificial systems able to hold a conversation with a human user, usually to achieve a certain objective (e.g., providing some information or fulfilling a task) through a natural language dialogue [1, 2]. With a spoken interface, the user is not restricted to traditional interfaces and so the number of contexts in which the system can be used grows, for example, with robots, with small devices, in environments where the user's eyes are busy (e.g., in-car systems), or with older adults or the disabled.

Such dialogues require a sequence of user and system turns until the objective is fulfilled. To organize these exchanges and gradually attain the user's objective, the system must be endowed with complex mechanisms to compute what the best system intervention is given the user's responses up to the current moment of the dialogue, to take the initiative in the conversation when necessary to lead the dialogue and maintain it within the limits of the system's domain of expertise, to request the information missing to achieve the objective, and to solve the possible miscommunication problems that may arise, including asking the user for clarifications.

Although it has been underlined that considering user models is of great importance to enhance the system services, these are usually not considered when designing the dialogue model for conversational agents, which center in solving the previously mentioned challenges for a general user [3, 4].

Adapting a conversational system to the peculiarities of its different users requires learning a user model jointly to the dialogue management process design. The application of statistical approaches to perform both tasks allows using real dialogues to explore a wider space of models (both user and dialogue models) if compared with rule-based systems [5, 6]. Although the definition of the parameters of a statistical model depends on expert knowledge about the application domain, the final objective is to develop more portable, robust, and adaptive conversational systems.

In this paper we present a framework to develop spoken dialogue systems able to adapt dynamically to the requisites and preferences of the user. The user model proposed estimates the user intention during the dialogue, that is, the information that the user is going to provide during the dialogue in order to achieve their objective. The prediction is performed after each system intervention, so that the adaptation can

be achieved dynamically. To complete the adaptation a dialogue manager has been designed following a statistical approach based on neural networks that incorporates the user information generated by the user model along with the history of the dialogue.

Both models can be computed from a dialogue corpus using a format defined to code the information in a data structure that considers only whether each piece of information has been provided or not by the user and with which confidence it has been recognized by the speech recognition and understanding modules. This allows the proposed framework to be used in complex domains as it reduces the search space without losing relevant information.

We also describe an implementation of our proposal for the development of a spoken dialogue system information that provides user-adapted tourist information. We have evaluated the developed system and assessed the influence of the adaptation in the quality of the acquired dialogues and the information provided. The evaluation results show that the users perceive a higher quality and performance in the interaction with the system when our proposal is integrated.

The current paper presents the following important contributions. One of the main ideas of our proposal is that the dialogue system and the user model interact simultaneously, not offline as in our previous works [7, 8], so that the user model is employed in real time by the dialogue manager to decide the best answer dynamically and not in a posteriori laboratory evaluations. In addition, the technology used to build the proposed user model does not replicate that of the dialogue manager. This avoids the dialogues generated to be biased.

The novel user model incorporates numerous information sources to decide the next user action and integrates and orchestrates these heterogeneous sources and uses them significantly to make decisions. To optimize these computations it is very important to estimate the task underlying the current dialogue. Determining the task is one of the main innovations of the paper as it is used not only for the user simulator but also for the practical implementation of the dialogue system. Another important contribution is that such implementation is performed with a neural-network-based classifier trained for each of the tasks considered, so that a better selection of the next system can be attained weighting the outputs of these specialized classifiers.

All these decisions have been carefully designed so that the proposal can be portable across domains and applied to systems with varying complexity. In particular, the current paper shows the application of our proposal to develop a dialogue system that facilitates touristic information. This provides the benefit of showing an application of our scientific proposals in a task designed for a real system that provides a real service to real users. Finally, the overall evaluation of the proposal not only employs measures to compare the answers provided by the manager but also incorporates the evaluation of the system using objective and subjective measures obtained with real users.

The remainder of the paper is as follows. In Section 2 we describe the context and main motivation of our proposal and review main approaches focused on key aspects related to it, such as user modeling techniques when interacting with spoken conversational agents and the application of statistical methodologies for interaction management. Section 3 presents in detail our proposal to develop user-adapted spoken conversational agents. Section 4 shows a practical implementation of our proposal to develop a system providing tourist information and services. Section 5 describes the evaluation results obtained after a comparative evaluation of a baseline version of this system with a user-adapted version that integrates our proposal. Finally, in Section 6 we describe some conclusions and guidelines for future work.

2. Related Work

Usually dialogue models are aimed at a generality of users and are not developed to consider their peculiarities through specific user models [9–11]. However, some authors have addressed before the challenge of building adaptive systems through different adaptation levels [12]. A simple approach is to use different user profiles containing the preferences of individual users or user groups, which the users have expressed prior to using the system; this is frequently used in ambient intelligence environments where the system considers the user preferences to operate the environment.

Other approaches attempt to adapt to the users dynamically by detecting their needs instead of explicitly asking the users to introduce the parameters. This has been used, for example, to adapt to the users by detecting their skills while using the system [13–15] and also to detect the variable intentions of the user during the dialogue and how these are affected by the system interventions [16, 17].

Modeling the dialogue using intentions allows overcoming of the ambiguity introduced by natural language when the communication process considers the words uttered [18], as dialogue acts can be used to describe and represent intentions [19, 20]. Dialogue acts have been used to represent key aspects in dialogue systems, from corpus annotation to the design of dialogue models or the representation of the behavior of interlocutors during the conversation. The semantic representation of information can also be achieved through other models such as Hybrid Deep Belief Networks (HDBNs) [21], time-dependent semantic similarities [22], fuzzy set theories [20], or ontologies [23].

The use of n-grams to predict the user intention was introduced in [24]. The model predicts the most probable action according to the actions defined for the user and the previous dialogue history. As n-grams limit the dialogue history to the previous n user actions (usually 2), sometimes important information is lost and the response selected by the model makes sense in the context of the previous user intervention but not in the whole dialogue.

In [17] a user model is proposed based on Hidden Markov Models to represent the previous states and history of the dialogue in more detail. The proposed model is based on the theory of Information States [16], feature vectors that include the current state, the previous history of the dialogue, and the possible user actions. In [25] an extension is proposed to represent both the user and the system behavior by training a specific model for each of the objectives of the dialogue.

In [26] a method is proposed to simulate the user intention by integrating different levels of discourse knowledge with Markov models and regression. In [27] the model proposed also incorporates acoustic features to increase the accuracy of the intention prediction, which are extracted during the speech recognition process from the words uttered and the confidence measures computed by the recognizer.

The approach described in [28] uses other features for the prediction of the user intention, a mixture of morpheme, discourse, and domain level characteristics that are integrated by means of a maximum entropy model. In [29] a technique is proposed that uses information about the activities performed by the users as a solution to disambiguate their inputs. This information is achieved with a reinforcement learning algorithm and then used by the dialogue manager to decide the next system action.

The user simulator technique in [30] is based on an agenda that stores the possible objectives of the user during the dialogue and represents the interaction as a sequence of states and dialogue acts that define the user plan and that may vary during the dialogue. This structure is implemented as a stack that stores the dialogue acts which are pending in order for the user to provide all the information required to achieve their objective within the dialogue. The agenda and the objective are updated dynamically during the conversation.

As described in Section 3, the technique that we propose to model the user intention combines different information sources and heuristics to enhance the dialogue model. It can easily be integrated with the dialogue manager and can be used to automatically simulate dialogues by means of its interaction with the system, thus decreasing the effort necessary to enhance the dialogue model.

Once the user model has been learnt, it is necessary to define how it is going to be used by the dialogue manager to adapt the system. In [31–33] the main architectures and methodologies for dialogue management are described. The use of statistical techniques to automatize the process reduces the time and effort needed for the development of new managers and allows covering new requisites and makes the system portable to new tasks and users [9], then increasing the system scalability.

The main statistical methodology for dialogue management relies on decision Markov models and reinforcement learning to model the dialogue as an optimization problem [34, 35]. One of the main drawbacks of this proposal is the difficulty to apply it to real tasks with a high number of states [36]. Partially Observable Markov Models (POMDPs), though also limited in real tasks [36], offer a possibility to explicitly represent uncertainty in human-machine communication [37]. Other statistical proposals for dialogue management combine POMDPs with rule-based methodologies [18, 38] or create partitions in the search space to make it more tractable [39, 40]. Also they have been merged with Bayesian models to favour learning the best features [41, 42]. Other recent works are merged with Hidden Markov Models [25] and Finite-State Transducers [43, 44].

The statistical methodology that we propose for dialogue management is based on using a data corpus to learn the next system action considering the history of the dialogue up to the current moment and the user model. This is a benefit of our proposal with respect to the state of the art, enhanced by the fact that the data structure used makes it possible to store the information efficiently to adapt to different tasks.

In our previous work we have evaluated different classifiers widely used in the field of natural language processing and speech technologies [6–8]. Neural networks have outperformed the other classifiers in different application domains [6, 45–48].

3. Proposed Framework to Develop User-Adapted Spoken Dialogue Systems

Figure 1 shows the architecture that integrates our proposed framework to generate adaptive spoken conversational agents. A user modeling module considers the previous dialogue interactions and specific features of the user (defined by means of user profiles) to predict the user intention, defined as the next user action, which we represent by one or more dialogue acts as described in the previous section.

The dialogue manager takes as input this prediction, the current user utterance, and the sequence of user and system dialogue acts until the current moment. Using this information it selects the next system action (next system dialogue act). The following subsections describe the statistical methodologies proposed for the development of the two modules.

3.1. User Modeling. Our proposed technique for user modeling simulates the user intention providing the next user dialogue act in the same representation defined for the spoken language understanding module. If we denote the user and system interventions at a certain moment i as U_i and A_i, respectively, a dialogue can be modeled as a sequence of pairs (A_i, U_i) where A_0 is the initial system prompt and U_n is the system response to the system prompt A_n. Each of these interventions is modeled as dialogue acts.

The lexical, syntactic, and semantic information associated with the speaker u's ith turn (U_i) is denoted as c_i^u. This information is usually represented by the following:

(i) the words uttered;

(ii) part of speech tags, also called word classes or lexical categories; common linguistic categories include noun, adjective, and verb;

(iii) predicate-argument structures, used by SLU modules in various contexts to represent relations within a sentence structure; they are usually represented as triples (subject-verb-object);

(iv) named entities or groups of words that identify objects, for example, proper names, dates, and numerical expressions.

The model proposed to simulate the user intention is based on enhanced version of the one presented in [49]. Each user action is represented by means of the subtask in the dialogue where it belongs, the named entities spotted, and the dialogue acts that intervene in the user utterance.
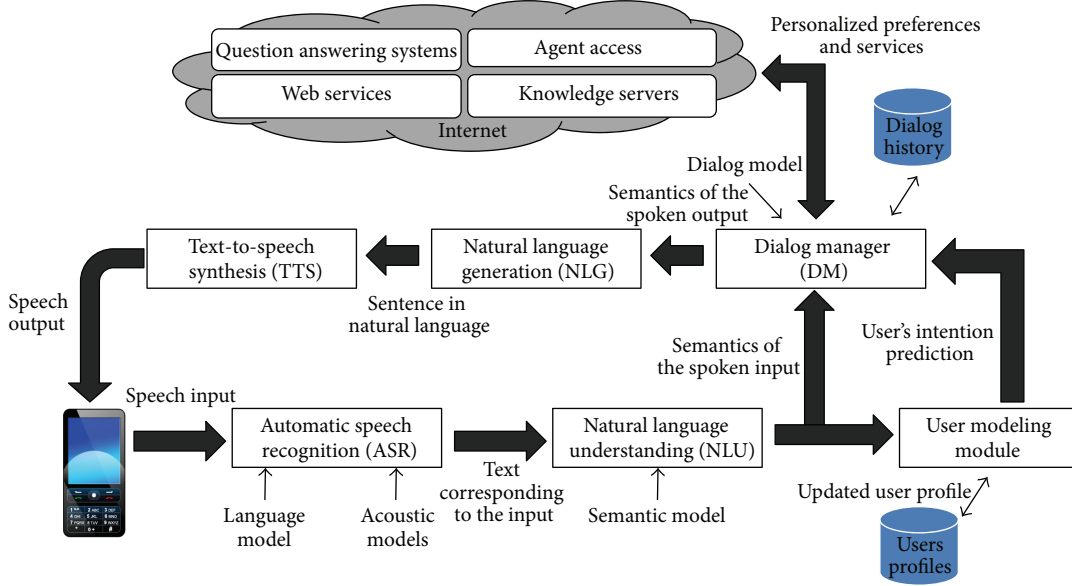
FIGURE 1: Architecture to develop user-adapted spoken conversational agents.

For instance, for the user action corresponding to conveying the origin of a trip, the subtask may be timetable query, the named entity could be the name of the origin city, and the dialogue act could be coded as *Provide-Origin*.

For speaker u, DA_i^u denotes the dialogue label of the ith turn, and ST_i^u denotes the subtask label to which the ith turn contributes. The next dialogue act in the user model is computed as indicated in (1) from the k previous utterances (dialogue history) and the lexical, syntactic, and semantic features extracted from the words uttered by the user:

$$DA_i^u = \underset{d^u \in \mathscr{D}}{\operatorname{argmax}} P\left(d^u \mid c_i^u, ST_{i-1}^{i-k}, DA_{i-1}^{i-k}, c_{i-1}^{i-k}\right). \quad (1)$$

The dialogue act computed is then used together with the dialogue history (k previous interventions) and the characteristics of the user utterance to determine the subtask to which the user intervention contributes as indicated in the following:

$$ST_i^u = \underset{s^u \in \mathscr{S}}{\operatorname{argmax}} P\left(s^u \mid DA_i^u, c_i^u, ST_{i-1}^{i-k}, DA_{i-1}^{i-k}, c_{i-1}^{i-k}\right). \quad (2)$$

In our proposal, we consider static and dynamic features to estimate the conditional distributions shown in (1) and (2). Dynamic features include the dialogue act and the task/subtask. Static features include the words in each utterance, the dialogue acts in each utterance, predicate-arguments in each utterance, and also a set of features included in a user profile. All pieces of information are computed from corpora using n-grams, that is, computing the frequency of the combination of the n previous words, dialogue acts, or predicate-arguments in the user turn.

The user profile includes a unique identifier for the user, their expertise level, gender, most frequent queries, and the device that they usually employ to communicate with the system.

The previous equations are solved in [49, 50] using the maximum entropy technique, which allows obtaining the average parameters from a corpus. Equation (3) shows how this estimation can be computed by means of a Gibbs distribution of weighted (λ) parameters, where X refers to the distributions in (1) and (2) (DA_i^u or ST_i^u) and ϕ corresponds to our feature vector:

$$P\left(X = st_i \mid \phi\right) = \frac{e^{\lambda_{st_i} \cdot \phi}}{\sum_{st=1}^{V} e^{\lambda_{st_i} \cdot \phi}}. \quad (3)$$

This calculation fastens the calculation of the distributions, which in turn allows using bigger corpora if compared to state of the art approaches [49–51]. The binary classifiers for each task are computed using

$$P\left(y \mid \phi\right) = 1 - P\left(\overline{y} \mid \phi\right) = \frac{e^{\lambda_y \cdot \phi}}{e^{\lambda_y \cdot \phi} + e^{\lambda_{\overline{y}} \cdot \phi}} = \frac{1}{1 + e^{-\lambda'_{\overline{y}} \cdot \phi}}, \quad (4)$$

where $\lambda_{\overline{y}}$ is the parameter vector for the antilabel $\overline{y}$ and $\lambda'_{\overline{y}} = \lambda_y - \lambda_{\overline{y}}$.

3.2. Interaction Management.
In the dialogue management process a dialogue sequence S_i represents the pair (A_i, U_i). The aim is to select the best system response according to the dialogue sequences up to the current moment. This is processed as shown in the following:

$$\widehat{A}_i = \underset{A_i \in \mathscr{A}}{\operatorname{argmax}} P\left(A_i \mid S_1, \ldots, S_{i-1}\right), \quad (5)$$

where set $\mathscr{A}$ contains all the possible system answers.

The main challenge for this calculation is the high number of dialogue sequences that it may have to deal with, even in simple application domains. We tackle this problem using a data structure, the interaction register (IR), in which we store
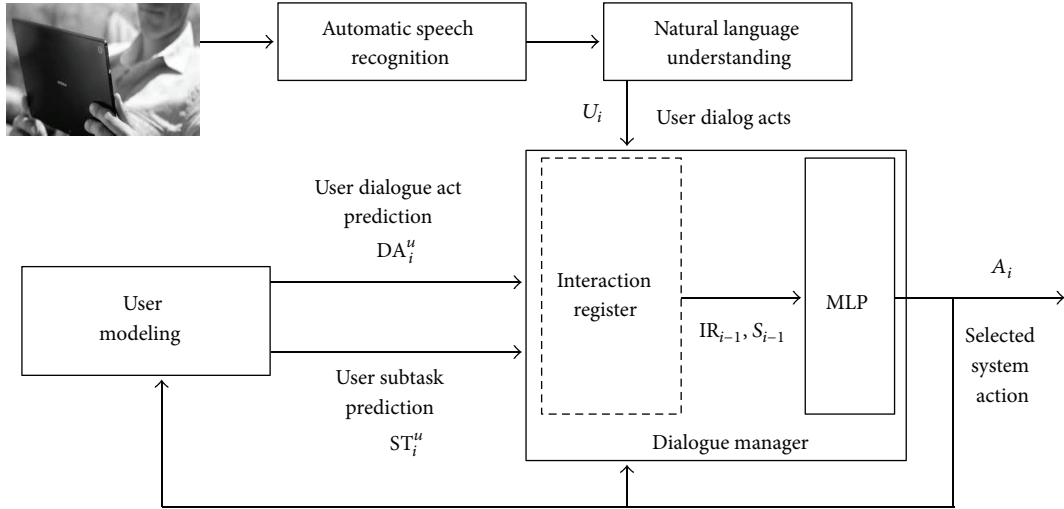
FIGURE 2: Integration of the proposed user model and dialogue management methodology for the development of user-adapted spoken dialogue agents.

the information provided by the user during the dialogue in terms of dialogue acts, the act prediction generated by (1), and the subtask prediction provided by (2).

At each moment i, the IR contains an abstract representation of the information that has been provided independently of the order in which it has been communicated. This allows reducing the number of possible dialogue histories to consider, so that the selection of the system response can be computed even in complex tasks as shown in (6). As can be observed, the previous dialogue sequence (S_{i-1}) is also considered in the calculation. This is due to the task-independent dialogue acts, such as confirmations or negations, that demand considering the previous turn:

$$\widehat{A_i} = \underset{A_i \in \mathcal{A}}{\mathrm{argmax}}\, P\left(A_i \mid \mathrm{IR}_{i-1}, S_{i-1}\right). \quad (6)$$

The IR only considers the dialogue acts and not the actual values provided. This way, each position of the IR can only take one out of three possible values: 0 when the user has not conveyed the information, 1 when it has been conveyed and processed with high confidence [52], or 2 when it has been conveyed but the recognition confidence is low and thus it probably requires confirmation. As it has been previously described, (6) is solved by means of a neural network whose input is the pair (IR_{i-1}, A_i) and outputs the probability of selecting each of the possible system answers.

Figure 2 summarizes the combination of the proposed user modeling and dialogue management methodologies. As can be observed, the user modeling module provides a prediction of the next user dialogue act and the current subtask of the dialogue. The set of user dialogue acts and predicted values for the current user's dialogue act and subtask are used to update the interaction register. The dialogue manager considers this register and the last system response for the selection of the next system action.

3.3. MLP Classifier. The codification of the input pair (IR_{i-1}, S_{i-1}) for the MLP includes the following variables:

(i) The last system response (A_{i-1}) is a variable with the same number of bits compared to possible system responses (C)

$$\vec{x}_1 = \left(x_{1_1}, x_{1_2}, x_{1_3}, \ldots, x_{1_C}\right) \in \{0, 1\}^C. \quad (7)$$

(ii) The interaction register (IR_{i-1}) is modeled by the variables required to model each of the user dialogue acts that are dependent on the task (each one modeled using a variable with three bits, which are related to the previously described $\{0, 1, 2\}$ values), the predicted current user dialogue act obtained with (1) (variable with the same number of bits compared to possible user responses, N), and the predicted dialogue subtask obtained with (1) (variable with the same number of bits compared to possible subtasks, T):

$$\vec{x}_i = \left(x_{i_1}, x_{i_2}, x_{i_3}\right) \in \{0, 1\}^3 \quad i = 2, \ldots, N+1,$$

$$\vec{x}_{N+2} = \left(x_{1_1}, x_{1_2}, x_{1_3}, \ldots, x_{1_N}\right) \in \{0, 1\}^N, \quad (8)$$

$$\vec{x}_{N+3} = \left(x_{1_1}, x_{1_2}, x_{1_3}, \ldots, x_{1_T}\right) \in \{0, 1\}^T.$$

(iii) The task-independent information is modeled using three variables, which are, respectively, related to the dialogue acts used by the user to provide an affirmation or a rejection or to indicate that they have not understood the previous system's action. Each one of these dialogue acts is modeled using a variable with the same number of bits compared to possible values ($\{0, 1, 2\}$). Consider

$$\vec{x}_i = \left(x_{i_1}, x_{i_2}, x_{i_3}\right) \in \{0, 1\}^3 \quad i = N+4, \ldots, N+6. \quad (9)$$

User's utterance:
[Spanish] *Quisiera saber el horario de visita de la Madraza de Granada para manana.*
[English] *I would like to know the visit hours of the Madraza in Granada for tomorrow.*
Semantic interpretation:
(*Places-Interest*)
Query-type: timetables
Place: Madraza Palace
Date: tomorrow

FIGURE 3: An example of the labeling of a user turn in the tourist information system.

4. Case Application: Development of a Dialogue System Providing Tourist Information and Services

We have developed a tourist information dialogue system that integrates our proposal for user adaptation. The system provides information about navigation, hotels, interesting spots and monuments, restaurants and bars, shopping, cultural and sport activities, and public transportation taking into account the preferences and physical situation of the user in a city. The information provided is retrieved from different web repositories, most of them updated daily.

We have defined 10 types of queries with 115 user dialogue acts corresponding to the pieces of data that the user must provide (*Places-Interest*, *Weather-Forecast*, *Hotel-Booking*, *Restaurants-Bars*, *Shopping*, *Street-how-to-get*, *Cultural*, *Sport*, *Festivities*, and *Public-Transport*) and 3 task-independent dialogue acts (*Affirmation*, *Negation*, and *Not-Understood*). The IR defined contains 129 fields corresponding to the dialogue acts described. Figure 3 illustrates how a user utterance can be represented using our proposal.

5. Evaluation of Our Proposal for the Tourist Information System

We have evaluated the system with respect to a baseline that employs the same approach for dialogue management but does not incorporate the user model [53].

To assess the benefits of our proposal to include user adaptation, we have evaluated the developed system for the specific scenario of the city of Granada (Spain) and compared it to the baseline for the same city. In order to do so, 150 recruited users have followed a set of scenarios that specifies a set of objectives that must be fulfilled by the user at the end of the dialogue and are designed to include the complete set of functionalities previously described for the system. The number of recorded dialogues was 600: 300 dialogues were acquired by 75 users using the baseline version of the system and 300 dialogues were acquired by 75 users using the user-adapted system. An example of the defined scenarios is as follows:

```
User name: Jose Madrigal

Location: Periodistas Street

Date and Time: 2015-04-03, 19:45

Device: MOT-TVS-2-GP 01-00-34-7B-06-A7
```

```
Objective: Cultural activities for today.

Listings for next weekend
```

For illustrative purposes, Figure 4 shows a dialogue corresponding to the previous scenario acquired with the baseline and user-adapted systems (U: user turns; S: system turns). As can be observed, the user-adapted system shows a tendency of providing the required services with higher agility and using more natural answers than the baseline.

We have carried out an objective and subjective assessment to compare the two versions of the system. The objective evaluation was aimed mainly at detecting the nature of the information exchanged and the quality and duration of the exchanges: dialogue success, error correction, dialogue and turn duration, frequency of dialogues and number of different dialogues generated, and proportion of user versus system actions during the dialogue. The measures defined to complete this evaluation are summarized as follows [54, 55].

Evaluation Measures Based on the Interaction Parameters Gathered from the Dialogues Acquired with the Two Versions of the Enjoy Your City System

Dialogue Success

Dialogue success rate (%success): the percentage of successfully completed tasks. In each scenario, the user has to obtain one or several pieces of information, and the dialogue success depends on whether the system provides the correct data (according to the aims of the scenario) or incorrect data to the user.

Average number of corrected errors per dialogue (nCE).

The average of errors detected and corrected by the dialogue manager. We have considered only the errors that modify the values of the attributes and that could cause dialogue failure.

Average number of uncorrected errors per dialogue (nNCE): the average of errors not corrected by the dialogue manager. Again, only errors that modify the values of the attributes are considered.

Error correction rate (%ECR) which is the percentage of corrected errors, computed as nCE/(nCE + nNCE).

High-Level Dialogue Features

Average number of turns per dialogue (avgturns/dial).

Percentage of different dialogues (%diff).

S1: Welcome to Enjoy Your City Granada. How can I help you?	*No personalization.*
U1: I want to know about cultural activities for today.	
S2: There are more than 10 activities. To hear the complete list say list. To hear about a specific kind of activity say cinema, theater, music, exhibition, literature, or science.	*No information about location or interests.*
U2: Science.	
S3: There is one activity: conference about transgenic food in Palacio de la Madraza at 20:00. Do you want to know about any other activities?	*No information about location or interests.*
U3: Yes, literature.	
S4: There are two activities: conference on Scandinavian black novels in Biblioteca de Andalucía at 19:00 and conference on the Gold Century in Faculty of Philosophy at 19:00. Do you want to know about any other activities?	*No information about location or interests.*
U4: No, thank you.	
S5: Have a nice day!	

(a)

S1: Good afternoon Javier. How can I help you?	*Personalized prompt.*
U1: I want to know about cultural activities for today.	
S2: Provided your current position, there are five activities that you can attend today. Two of them are from the science and literature categories. Do you want to know more about them?	*Only the activities which are reachable in time from the current location of the user are shown, and the categories preferred by the user are highlighted.*
U2: Yes.	
S3: There is a conference on transgenic food in Palacio de la Madraza at 20:00 and a conference on Scandinavian black novels in Biblioteca de Andalucía at 19:00. Do you want to know about any other activities?	
U3: No, thank you.	
S4: Have a nice day!	

(b)

FIGURE 4: An example of a dialogue for the tourist domain using the initial system (a) or the user-adapted system (b).

Number of repetitions of the most seen dialogue (#repMS).

Number of turns of the most seen dialogue (#turnsMS).

Number of turns of the shortest dialogue (#turnsSh).

Number of turns of the longest dialogue (#turnsLo).

Ratio users versus system actions (us/sysAct).

Dialogue Style/Cooperativeness Measures

System dialogue acts: confirmation of concepts and attributes, questions to require information, and answers generated after a database query.

Confirmation rate (%confirm) which was computed as the ratio between the number of explicit confirmations turns (nCT) and the number of turns in the dialogue (nCT/nT).

User dialogue acts: request to the system, providing information, confirmation, yes/no answers, and other answers.

Goal-directed actions versus grounding actions: goal-directed actions are requesting and providing information, while grounding actions are explicit and implicit confirmations, dialogue formalities (greetings, instructions, etc.), and unrecognized actions.

We then used two-tailed t-tests to compare the means across the different types of scenarios and users as described in [54]. The significance of the results was computed using the SPSS software [56] with a significance level of 95%. (The degrees of freedom employed for t-tests are $N - 1$ in case the compared groups have the same number of samples (N) and $N1 + N2 - 1$ when they differ in the number of samples ($N1$ and $N2$).)

To account for the users' perception, we conducted a subjective evaluation by means of the following questionnaire with responses in a 5-point Likert scale: (Q1) the system understood me; (Q2) I understood the system messages; (Q3) it was easy to obtain the information I was looking for; (Q4) the interaction pace was adequate; (Q5) if the system made recognition mistakes, I could solve them easily. The answers for the different questions were the same: never/not at all, seldom/in some measure, sometimes/acceptably, usually/well, and always/very well.

The objective evaluation was aimed mainly at detecting the nature of the information exchanged and the quality and duration of the exchanges: dialogue success, error correction, dialogue and turn duration, frequency of dialogues and

TABLE 1: Results of the high-level dialogue features defined for the comparison of the baseline and user-adapted systems.

	Baseline system	User-adapted system
%success	85%	97%
nCE	0.83	0.90
nNCE	0.18	0.09
%ECR	81%	92%
avgturn/dial	12.2	8.1
%diff	85%	83%
#repMS	5	8
#turnsMS	9	7
#turnsSh	7	7
#turnsLo	17	15

TABLE 2: Distribution of the user dialogue acts in both systems.

	Baseline system	User-adapted system
U_query	31.74%	33.43%
$U\_provide$	21.72%	19.98%
$U\_confirm$	10.81%	9.34%
U_yesno	33.47%	35.77%
U_other	2.26%	1.48%

TABLE 3: Distribution of the system dialogue acts in both systems.

	Baseline system	User-adapted system
$S\_confirm$	13.31%	11.16%
$S\_request$	18.14%	16.27%
S_inform	67.63%	71.84%
S_other	0.92%	0.73%

TABLE 4: Dialogue spent on goal-directed actions, ground actions, and other possible actions.

	Baseline system	User-adapted system
Goal-directed actions	66.37%	74.16%
Grounding actions	32.34%	24.39%
Rest of actions	1.29%	1.45%

number of different dialogues generated, and proportion of user versus system actions during the dialogue.

Table 1 shows the results; as can be observed both systems attained good success rates, though the system with user model outperforms the baseline (85% versus 97%). The user-adapted dialogue obtained a better user error detection and correction rate, though this improvement was not statistically significant using a t-test. This may be because we did not incorporate any improvement on the recognition and understanding modules with respect to the baseline. Though this was not an objective for incorporating the user model, it makes it possible to correct errors faster when they are detected during the dialogue.

Also with the adapted system the duration of the successful dialogues was shorter (8.1 turns in average instead of the average 12.2 turns of the baseline), as the users are less often required to provide additional information because the system is already adapted to them. This reduction showed to be significant with the t-test. The adaptation also makes it possible for the users to choose between a wider range of alternatives, as the number of turns of successful dialogues presents a higher deviation than in the baseline. Also the number of user turns is higher than the number of system turns, and the nature of such turns is more in providing new values than in confirming information that was already provided.

Regarding the dialogue participant activity, the dialogues acquired with the user-adapted version of the system have a higher proportion of system actions, as less information requires confirmation in the user-adapted system. There is also a slight reduction in the mean values of the turn length; these dialogues are statistically shorter, as they provide 1.41 actions per user turn instead of the 1.67 actions provided by the baseline dialogues. This is again because the users have to explicitly provide and confirm more information in the baseline system. The results of the t-test in a comparative analysis of this measure showed a significant difference (significance value of 0.029).

Regarding the measures used to evaluate the style of the dialogues, we counted the number of each of the possible actions of the system and the user. To do this, the system's

actions have been grouped into confirmations of queries and user-supplied values ($S\_confirm$), questions to prompt the user ($S\_request$), and system responses generated after consultation system information repositories (S_inform). User actions have been grouped into queries to the system (U_query), actions in which the user provides information required by the system ($U\_provide$), confirmation of queries and values previously provided ($U\_confirm$), affirmations and rejections (U_yesno), and other actions not included in the previously described categories (U_other).

Table 2 shows that users need to provide less information using the user-adapted system. This explains the higher proportion for the rest of user actions with regard to the baseline system (both differences significant over 98%). There is also a higher proportion of yes/no actions for the user-adapted dialogues, which was not significant in the t-test. Table 3 shows a higher proportion of the inform and confirmation system actions when this system is used in comparison with the baseline system (both differences significant over 98%). A higher proportion of yes/no actions for the user-adapted dialogues can also be observed. These actions are mainly used to confirm that the specific queries have been correctly provided.

We also measured the ratio of actions to achieve the objective(s) of the dialogue (actions in which queries are performed and/or information is provided), actions that increase the number of turns (especially confirmations and rejections), and remaining actions. Table 4 shows a comparison between these categories. As can be observed, the dialogues provided by the user-adapted system have a better quality, as the proportion of goal-directed actions is higher.

Table 5: Results of the subjective evaluation of the baseline and user-adapted systems with real users (1 = worst and 5 = best evaluation).

	Baseline system	User-adapted system
Q1	M = 4.61, SD = 0.37	M = 4.83, SD = 0.34
Q2	M = 3.63, SD = 0.24	M = 3.97, SD = 0.26
Q3	M = 3.84, SD = 0.56	M = 4.35, SD = 0.35
Q4	M = 3.41, SD = 0.28	M = 4.24, SD = 0.29
Q5	M = 3.25, SD = 0.59	M = 3.36, SD = 0.57

This difference showed a significance value of 0.029 in the two-tailed t-test.

Finally, Table 5 shows the results of the subjective evaluation. The values obtained show that both systems were successful, though the adapted system was perceived as having a more appropriate pace, as it is faster, and thus that it was easier to find the information required.

6. Conclusions and Future Work

In this paper, we present a framework to build user adaptive dialogue systems based on statistical dialogue management enhanced with a dynamically computed user model. In our proposal, the user model computes the most probable next user turn and merges it with information about the user preferences and needs, so that it is considered by the dialogue manager along with the complete history of the interaction in order to adapt the interaction. These calculations are performed automatically using an abstract representation of the interaction structure based on dialogue acts that simplifies the search space and makes our proposal appropriate even for complex interactions and portable across domains.

To show the benefits of our proposal we have implemented a dialogue system that provides touristic information employing the framework and compared it with a baseline system that uses the same dialogue management process but does not incorporate the user model. In order to perform the evaluation, we have performed a series of experiments with recruited users recording the interaction parameters as well as assessing their subjective opinion.

The results show that the system that incorporates our framework obtained a higher success rate in the provision of the adapted services. By means of the user-adapted system, the time required to provide the information can be reduced. In addition, the interaction was more fluid and the dialogues with the adapted system present a better ratio of goal-directed actions selected by the system to successfully provide the different services, thus decreasing the actions that might discourage users (e.g., confirmations or re-request of information).

For future work we plan to build other systems in different application domains in order to demonstrate empirically that our proposal can be used with different user and dialogue models of varying complexity. We also plan to extend our proposal to dialogues that cannot be translated into a form-filling structure, such as companions or health coaches.
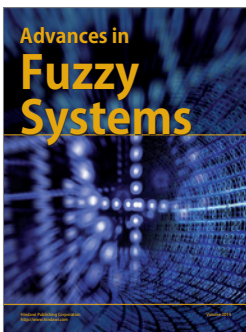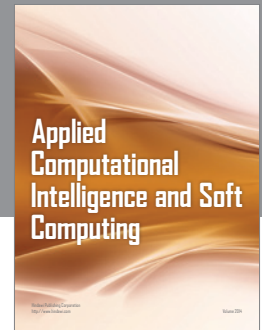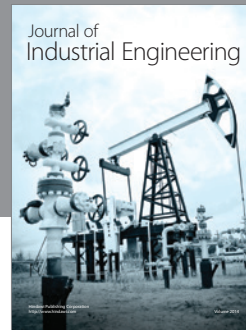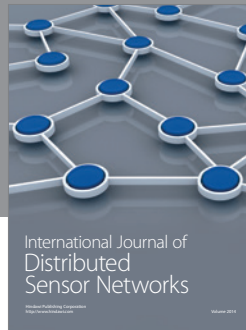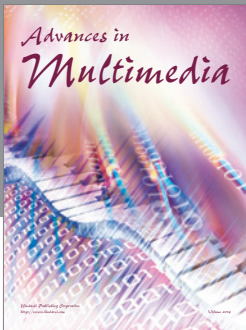
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] R. Pieraccini, *The Voice in the Machine: Building Computers That Understand Speech*, MIT Press, 2012.

[2] M. F. McTear, *Spoken Dialogue Technology: Towards the Conversational User Interface*, Springer, 2004.

[3] S. Seneff, M. Adler, J. Glass et al., "Exploiting context information in spoken dialogue interaction with mobile devices," in *Proceedings of the International Workshop on Improved Mobile User Experience (IMUx '07)*, pp. 1–11, Toronto, Canada, May 2007.

[4] S. Kartakis and C. Stephanidis, "A design-and-play approach to accessible user interface development in Ambient Intelligence environments," *Computers in Industry*, vol. 61, no. 4, pp. 318–328, 2010.

[5] S. Young, "The statistical approach to the design of spoken dialogue systems," Tech. Rep. CUED/F-INFENG/TR.433, Cambridge University Engineering Department, Cambridge, UK, 2002.

[6] D. Griol, L. F. Hurtado, E. Segarra, and E. Sanchis, "A statistical approach to spoken dialog systems design and evaluation," *Speech Communication*, vol. 50, no. 8-9, pp. 666–682, 2008.

[7] D. Griol, L. F. Hurtado, E. Segarra, and E. Sanchis, "Managing unseen situations in a stochastic dialog model," in *Proceedings of the AAAI Workshop Stadistical and Empirical Approaches for Spoken Dialogue Systems*, pp. 25–30, 2006.

[8] J. Iglesias, D. Griol, A. Ledezma, and A. Sanchis, "Influence of the data codification when applying evolving classifiers to develop spoken dialog systems," in *Proceedings of the IEEE Symposium on Evolving and Adaptive Intelligent Systems (EALS '14)*, pp. 58–64, Linz, Austria, June 2014.

[9] J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young, "A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies," *Knowledge Engineering Review*, vol. 21, no. 2, pp. 97–126, 2006.

[10] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, "Web usage mining as a tool for personalization: a survey," *User Modelling and User-Adapted Interaction*, vol. 13, no. 4, pp. 311–372, 2003.

[11] M. Eirinaki and M. Vazirgiannis, "Web mining for Web personalization," *ACM Transactions on Internet Technology*, vol. 3, no. 1, pp. 1–27, 2003.

[12] K. Jokinen, "Natural interaction in spoken dialogue systems," in *Proceedings of the Workshop Ontologies and Multilinguality in User Interfaces*, pp. 730–734, Crete, Greece, 2003.

[13] K. Evanini, P. Hunter, J. Liscombe, D. Suendermann, K. Dayanidhi, and R. Pieraccini, "Caller experience: a method for evaluating dialog systems and its automatic prediction," in *Proceedings of the IEEE Workshop on Spoken Language Technology (SLT '08)*, pp. 129–132, Goa, India, December 2008.

[14] K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds., *Computers Helping People with Special Needs. Proceedings of the 12th International Conference on Computers Helping People with Special Needs (ICCHP 2010)*, vol. 4061 of *Lecture Notes in Computer Science*, Springer, 2010.

[15] J. Ginzburg, "Dynamics and the semantics of dialogue," in *Logic, Language and Computation*, J. Seligman and D. Westerstahl, Eds., vol. 1, CSLI Publications, 1996.

[16] J. Bos, E. Klein, O. Lemon, and T. Oka, "DIPPER: description and formalisation of an information-state update dialogue system architecture," in *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*, pp. 115–124, Sapporo, Japan, July 2003.

[17] K. Georgila, J. Henderson, and O. Lemon, "Learning user simulations for information state update dialogue systems," in *Proceedings of the 9th European Conference on Speech Communication and Technology (Eurospeech '05)*, pp. 893–896, September 2005.

[18] S. Singh, D. L. Man, M. Kearns, and M. Walker, "Optimizing dialogue management with reinforcement learning: experiments with the NJFun system," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 105–133, 2002.

[19] J. Searle, *Speech Acts. An Essay on the Philosophy of Language*, Cambridge University Press, Cambridge, UK, 1969.

[20] D. R. Traum, "Speech acts for dialogue agents," in *Foundations of Rational Agency*, vol. 14 of *Applied Logic Series*, pp. 169–201, Springer, Dordrecht, The Netherlands, 1999.

[21] Y. Yan, X. Yin, S. Li, M. Yang, and H. Hao, "Learning document semantic representation with hybrid deep belief network," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 650527, 9 pages, 2015.

[22] F. Carrillo, G. A. Cecchi, M. Sigman, and D. Fernández Slezak, "Fast distributed dynamics of semantic networks via social media," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 712835, 9 pages, 2015.

[23] W. Gao and L. Zhu, "Gradient learning algorithms for ontology computing," *Computational Intelligence and Neuroscience*, vol. 2014, Article ID 438291, 12 pages, 2014.

[24] W. Eckert, E. Levin, and R. Pieraccini, "User modeling for spoken dialogue system evaluation," in *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '97)*, pp. 80–87, Santa Barbara, Calif, USA, December 1997.

[25] H. Cuayáhuitl, S. Renals, O. Lemon, and H. Shimodaira, "Human-computer dialogue simulation using hidden Markov models," in *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop (ASRU '05)*, pp. 290–295, IEEE, San Juan, Puerto Rico, December 2005.

[26] S. Jung, C. Lee, K. Kim, D. Lee, and G. G. Lee, "Hybrid user intention modeling to diversify dialog simulations," *Computer Speech and Language*, vol. 25, no. 2, pp. 307–326, 2011.

[27] R. Higashinaka, K. Sudoh, and M. Nakano, "Incorporating discourse features into confidence scoring of intention recognition results in spoken dialogue systems," *Speech Communication*, vol. 48, no. 3-4, pp. 417–436, 2006.

[28] C.-N. Seon, H. Kim, and J. Seo, "A statistical prediction model of speakers' intentions using multi-level features in a goal-oriented dialog system," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1397–1404, 2012.

[29] F. Wang and K. Swegles, "Modeling user behavior online for disambiguating user input in a spoken dialogue system," *Speech Communication*, vol. 55, no. 1, pp. 84–98, 2013.

[30] J. Schatzmann, B. Thomson, K. Weilhammer, H. Ye, and S. Young, "Agenda-based user simulation for bootstrapping a POMDP dialogue system," in *Proceedings of the 8th Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL '07)*, pp. 149–152, April 2007.

[31] M. Ahmed, R. Riyaz, and S. Afzal, "A comparative study of various approaches for dialogue management," *International Journal of Advanced Computer Technology*, vol. 2, no. 4, pp. 89–96, 2013.

[32] Y. Wilks, R. Catizone, S. Worgan, and M. Turunen, "Some background on dialogue management and conversational speech for dialogue systems," *Computer Speech and Language*, vol. 25, no. 2, pp. 128–139, 2011.

[33] C. Lee, S. Jung, K. Kim, D. Lee, and G. Lee, "Recent approaches to dialog management for spoken dialog systems," *Journal of Computing Science and Engineering*, vol. 4, no. 1, pp. 1–22, 2010.

[34] E. Levin, R. Pieraccini, and W. Eckert, "A stochastic model of human-machine interaction for learning dialog strategies," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 1, pp. 11–23, 2000.

[35] S. Singh, M. Kearns, D. Litman, and M. Walker, "Reinforcement learning for spoken dialogue systems," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS '99)*, pp. 956–962, 1999.

[36] S. Young, J. Schatzmann, K. Weilhammer, and H. Ye, "The hidden information state approach to dialog management," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. 149–152, IEEE, Honolulu, Hawaii, USA, April 2007.

[37] N. Roy, J. Pineau, and S. Thrun, "Spoken dialogue management using probabilistic reasoning," in *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL '00)*, pp. 93–100, Hong Kong, October 2000.

[38] P. A. Heeman, "Combining reinforcement learning with information-state update rules," in *Proceedings of the 8th Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL '07)*, pp. 268–275, April 2007.

[39] P. Lison, "Model-based bayesian reinforcement learning for dialogue management," in *Proceedings of the International Conference on Spoken Language Processing (INTERSPEECH '13)*, pp. 457–461, Lyon, France, August 2013.

[40] S. Png, J. Pineau, and B. Chaib-draa, "Building adaptive dialogue systems via bayes-adaptive POMDPs," *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 8, pp. 917–927, 2012.

[41] T. Paek and E. Horvitz, "Conversation as action under uncertainty," in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI '00)*, pp. 455–464, Stanford, Calif, USA, June-July 2000.

[42] H. M. Meng, C. Wai, and R. Pieraccini, "The use of belief networks for mixed-initiative dialog modeling," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 757–773, 2003.

[43] C. Hori, K. Ohtake, T. Misu, H. Kashioka, and S. Nakamura, "Recent advances in WFST-based dialog system," in *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH '09)*, pp. 268–271, September 2009.

[44] J. Planells, L. Hurtado, E. Sanchis, and E. Segarra, "An online generated transducer to increase dialog manager coverage," in *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH '12)*, Portland, Ore, USA, September 2012.

[45] S. M. Siniscalchi, T. S. Svendsen, and C.-H. Lee, "An artificial neural network approach to automatic speech processing," *Neurocomputing*, vol. 140, pp. 326–338, 2014.

[46] F. Zamora-Martínez, V. Frinken, S. España-Boquera, M. J. Castro-Bleda, A. Fischer, and H. Bunke, "Neural network language models for off-line handwriting recognition," *Pattern Recognition*, vol. 47, no. 4, pp. 1642–1652, 2014.

[47] L. Gajecki, "Architectures of neural networks applied for LVCSR language modeling," *Neurocomputing Journal*, vol. 133, pp. 46–53, 2014.

[48] T. Sagara and M. Hagiwara, "Natural language neural network and its application to question-answering system," *Neurocomputing*, vol. 142, pp. 201–208, 2014.

[49] S. Bangalore, G. Di Fabbrizio, and A. Stent, "Learning the structure of task-driven human-human dialogs," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 7, pp. 1249–1259, 2008.

[50] A. L. Berger, V. J. Pietra, and S. A. Pietra, "A maximum entropy approach to natural language processing," *Computational Linguistics*, vol. 22, no. 1, pp. 39–71, 1996.

[51] P. Haffner, "Scaling large margin classifiers for spoken language understanding," *Speech Communication*, vol. 48, no. 4, pp. 239–261, 2006.

[52] F. Torres, L. F. Hurtado, F. García, E. Sanchis, and E. Segarra, "Error handling in a stochastic dialog system through confidence measures," *Speech Communication*, vol. 45, no. 3, pp. 211–229, 2005.

[53] D. Griol and M. García-Jiménez, "Development of interactive virtual voice portals to provide municipal information," *Advances in Intelligent and Soft Computing*, vol. 151, pp. 161–172, 2012.

[54] H. Ai, A. Raux, D. Bohus, M. Eskenazi, and D. Litman, "Comparing spoken dialog corpora collected with recruited subjects versus real users," in *Proceedings of teh 8th SIGdial Workshop on Discourseand Dialogue*, pp. 124–131, Antwerp, Belgium, September 2007.

[55] J. Schatzmann, K. Georgila, and S. Young, "Quantitative evaluation of user simulation techniques for spoken dialogue systems," in *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pp. 45–54, Lisbon, Portugal, September 2005.

[56] L. Kirkpatrick, *A Simple Guide to IBM SPSS: For Version 20.0*, Wadsworth Publishing, 2012.

Advances in
Multimedia

The Scientific World Journal

International Journal of
Distributed Sensor Networks

Journal of
Industrial Engineering

Applied Computational Intelligence and Soft Computing

Advances in
Fuzzy Systems

Modelling & Simulation in Engineering

Journal of
Computer Networks and Communications

Advances in
Artificial Intelligence

![Hindawi]

Submit your manuscripts at
http://www.hindawi.com

Advances in
Computer Engineering

International Journal of
Computer Games Technology

International Journal of
Biomedical Imaging

Advances in
Artificial Neural Systems

Advances in
Software Engineering

Journal of
Robotics

Advances in
Human-Computer Interaction

Computational Intelligence and Neuroscience

International Journal of
Reconfigurable Computing

Journal of
Electrical and Computer Engineering