Analysis, design and application of flexible, contextual and dynamic dialogue management solutions based on Bayesian Networks

Fernando Fernández-Martínez, Javier Ferreiros

Grupo de Tecnología del Habla,
Universidad Politécnica de Madrid, Madrid, Spain.
ffm@die.upm.es, jfl@die.upm.es

Abstract

In this thesis we tackle the problem of identifying the best practices when designing and evaluating a spoken dialogue system. With the purpose of demonstrating that a more natural, flexible and robust dialogue is possible, and introducing a spoken dialogue system for controlling a Hi-Fi audio system as the selected prototype, we propose a Bayesian Networks (BNs) based solution for dialogue modelling combined with carefully designed contextual information handling strategies. Dynamic capabilities are also provided to keep the dialogue context permanently updated according to the evolution of the dialogue. All the thesis contributions have been evaluated finding an experimental support enough to demonstrate their relevance.

Index Terms: spoken dialogue systems, mixed initiative, Bayesian Networks, contextual information, usability, real users evaluation, electronic devices control

1. Introduction

Speech is the most widely used natural means of communication between people. Speech also is of increasing importance as a user-machine interface. As a result of the knowledge and the experience accumulated during almost half a century of speech technology research, now the time has come to design automated dialogue systems that make use of the communicative aspects of speech. In particular, it is essential to incorporate to the design of such systems some ideas related to the concept of "ambient intelligence" (AmI), for providing intelligent interfaces that are able to conduct a natural dialogue, including negotiations in order to achieve the goals required by users.

A dialogue system can be seen as a computer application that enables interaction and communication between users and machines as naturally as possible. Besides the typical recognition and text-to-speech conversion modules and other components, dialogue systems usually contain a module called Dialogue Manager (DM). This module is responsible for a dual task: to interpret the intention of the user and to decide how to continue the dialogue.

To successfully provide users with answers resembling a human-human interaction as much as possible, we believe that the design of a dialogue system should be approached from both a theoretical and practical point of view [1]. Thus, we must pay attention not only to dialogue management and modelling, but also to the enhancement of such models with knowledge about the specific tasks of the dialogue and the application domain (i.e. task and domain models). That way, it is feasible to develop procedures that support the user-machine interaction by useful elements of communication for realizing a collaborative and cooperative dialogue.

2. Dialogue management based on BNs

2.1. On the spoken dialogue system

Our conversational interface allows users to drive a Hi-Fi system from natural language sentences, differentially from other typical control systems based on simple commands. A detailed description of this system, its architecture and the implemented dialogue strategies can be found in [3][5][7].

2.2. On the dialogue management solution

As an alternative to classical dialogue systems (finite state automata or FSMs, script based systems or dialogue plans, etc.), we are presenting a dialogue solution based on BNs, that allows a greater flexibility and naturalness by appropriately defining dialogue as the interaction with an inference system [2].

The first task of the Dialogue Manager (DM) module is to identify the intention of the user (i.e. dialogue goals) considering the relevant information extracted by a semantic parser from the last utterance (i.e. available concepts) [6][13], together with the dialogue context. Then, according to the inferred goals the DM has to make a decision regarding how the dialogue should continue. Both tasks can be accomplished using BNs.

2.2.1. "Forward Inference"

As can be observed in Figure 1, BNs can be adopted to model the existing causal relation between the goals and the concepts [2][3][7]. Typically, both of them are assumed to be binary (i.e. a concept is true or "present" only when it is observed in the sentence). Thus, from the whole set of available evidences, e.g. $E = \{C_1 = 0, C_2 = 1, ..., C_N = 1\}$ for N defined concepts, a posterior probability $P(G_i = 1|E)$ can be obtained for each goal using the "Forward Inference" (FI) technique [2].

Subsequently, a decision is made for each goal on the comparison of the posterior with a defined threshold, θ. As a result of that comparison, one goal is "active" or "present" if the corresponding posterior is over the threshold ("absent" if not).

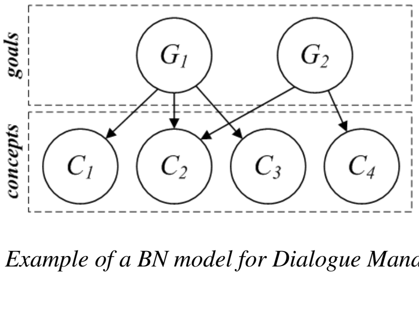

Figure 1: *Example of a BN model for Dialogue Management.*

Table 1: *Concept analysis used to drive the dialogue.*

	$P(C_j = 1\mid E^*) < \theta$	$P(C_j = 1\mid E^*)\geq\theta$
C_j absent $(C_j = 0)$	C_j **unnecessary** (No action)	C_j **missing** (Prompt to request C_j)
C_j present $(C_j = 1)$	C_j **wrong** (Prompt to clarify or notify about C_j)	C_j **required** (C_j is stored in the dialogue memory)

2.2.2. "Backward Inference"

After the FI process, and assuming the inferred results (i.e. those goals which were decided to be "present", $G_i = 1$) as new evidences, Bayesian inference can be applied again but this time aimed at the estimation of $P(C_j = 1|E^*)$, the probability that each concept should be present where E^* refers to the updated set of evidences (i.e. E also including goal evidences obtained through the FI process but removing the evidence corresponding to the target concept, C_j). This process is known as the "Backward Inference" (BI) technique [2]. Making a similar binary decision on the value of $P(C_j = 1|E^*)$, it is possible to check whether that concept should be present or not.

2.2.3. Concept analysis

The BI result can be compared with the actual occurrence of the concept enabling the classification presented in Table 1. As a result of that analysis [2] every concept can be properly classified allowing the DM to perform a suitable action. A possible dialogue proceeding strategy has been suggested for each possible result. For example, the system can drive the dialog prompting for the "missing" concepts.

3. Main thesis contributions

In this thesis, we have completed a thorough and comprehensive study of the BNs. In particular, we have focused on their possible application as new dialogue management solutions.

In the following subsections we will sucessively highlight the main contributions with regard to each explored solution.

3.1. Regarding the BN approach

The main advantages that we can highlight in this regard are:

- The BNs based inference system enables, through the FI process, **a better identification of the dialogue goals according to the intention of the user** (i.e. actions that the user may request the system to perform) **from the available concepts consistently with the context of the ongoing dialogue**. FI evaluation results at [9] showed a F-measure of $92,29\%$ regarding goal identification.
- BN models are defined at a semantic level. This allows their design **with independence of the language used**.
- **BNs can be automatically learnt from training data**. This favors portability and scalability across domains.
- BNs allow **a simple way of incorporating human knowledge**, e.g. refining the topology by hand.
- BNs allow to conduct an **analysis of congruence** between the goals assumed by the system to have been requested by the user, and all data collected during the interaction. Based on this analysis, the system can **determine the flow of interaction** and react according to the semantics of the application domain. In particular, through the BI process, it is possible to automatically detect **which concepts are needed** (available or not), **erroneous** or **optional** with regard to the inferred goals. Thus, the dialogue could go toward the generation of messages requesting the missing items, clarifying the erroneous ones and ignoring the optional ones. The performance of the BI process and its derived concept classification showed an $81,00\%$ F-measure at [9].
- The BNs enable **a true mixed initiative dialogue modelling**. Flexibility is probably the main asset of the proposed solution, and the most significant difference with regard to conventional approaches. In particular, **the user is not constrained to any predetermined goal or data sequence**. This flexibility is twofold, since it not only allows the user to decide the goals at the beginning of interaction, but also lets him/her jump to other goals without having completed the previous ones. Moreover, the user can respond with more data than those requested in a query, or even respond to a fact not asked by the system with regard to the inferred dialogue goals.
- Thanks to the **negotiation process** enabled between the users and the system, based on the FI and BI procedures, the system is capable of responding to complex issues (e.g. when the users provide inaccurate or insufficient information to meet the required dialogue goals) and **to assist or guide the users** toward the achievement of their dialogue goals driving the dialogue in an efficient manner, minimizing the number of questions and making maximum use of the context of dialogue.

3.2. Regarding the dynamic response of the system

We have laid the basis for enabling a dynamic response [10]:

- We have introduced the notion of **"relevance"** as the **remaining evidence** of a concept in the dialogue history.
- **Attenuation mechanisms** have been introduced that lower the relevance or the latency of information stored in past phases of the evolution of dialogue. Hence, the relevance of those elements can evolve to a level below a predefined threshold, so that they finally disappear definitively from the dialogue history. Due to this mechanism, it is possible to **maintain the dialogue history permanently updated** by assigning higher weight to more recent information, and lower to the older.

3.3. Regarding the use of a new BN based inference engine

In relation to the inference engine used by the DM [9]:

- We have presented a new alternative to traditional solutions based on multiple BN models (i.e. individually developed for each specific goal). In particular, we have proposed to rethink the inference problem from a **single global BN model** including all the defined concepts and goals. A "fusion" algorithm has been defined to obtain that BN model from the baseline multiple BN models.
- Unlike the baseline strategy, the proposed fusion method provides a single BN that ensures that both the FI and the BI processes are consistent with the dialogue context. Moreover, **the result of the analysis of congruence is also unique for each concept** and is obtained by considering a **whole goal evidence context**, thus avoiding potential mixed results for the same concept derived from analyzing each goal separately.

- This new solution offers **greatly improved performance in terms of BI**. By contrast, it offers a slightly lower performance in terms of FI. However, the fusion BN provides **a better overall performance** (i.e. combined F-measure is approximately 13% better [9]).
- We have designed solutions to **optimize the computational cost** of the models resulting from the fusion process (e.g. an information gain study from which we can select the most indicative concepts of each goal).

3.4. Regarding the use of contextual information

The DM is also provided with a set of contextual information handling strategies [3][11]. Regarding the benefits of applying those strategies we emphasize:

- **The robustness and the consistency of the system responses are improved** as the system is able to deal with dialogue phenomena such as "anaphora" (i.e. elements that refer to other previous parts of the dialogue) and "ellipsis"(i.e. omission of certain essential elements of the dialogue that may be derived from given context), the main dialogue phenomena that can mean a **loss of crucial information**. These strategies are based on:
  - the available confidence measures (from speech recognition and language understanding modules),
  - the history of the ongoing dialogue (∼short term),
  - the history of dialogue (∼long term, i.e. the dialogue concepts referred so far during the dialogue).
  - the status of the system (i.e. current values of the different parameters of the system, e.g. volume),
  - the task model (e.g. a semantic frame containing all the information needed to meet a specific goal),
  - and the application domain model (e.g. information on the number of tracks of a particular CD).
- To assess the relevance and appropriateness of the designed strategies, [1][8][12] we measured the **percentage of contextual turns** as the fraction of dialogue turns in which some of the strategies were successfully applied. In connection with that metric, we also measured the **percentage of system requests** which should be limited by the contextual capabilities of the system. The results for both metrics confirmed the valuable role of the contextual information handling strategies (i.e. more than half of the turns relied on this type of information) improving both dialogue efficiency and fluency.

3.5. Regarding the use of concept confidence measures

Regarding the proper consideration of concept confidence measures in terms of dialogue [4]:

- We have proposed the use of the confidence measures to weigh the evidence of the concepts. Thus, it is possible to incorporate these measures directly to both FI and BI processes, adopting the available confidences as the evidences from which to conduct the inference. As a consequence, both inference results are **directly weighted by the available concept confidences**.
- We have also expanded the analysis aimed at the correct classification of the concepts. This extension is based on the **incorporation of the confidence measures to the referred analysis** to exploit them by defining specific dialogue proceeding strategies for each confidence level.

3.6. Regarding the evaluations with real users

Most important results derived from these evaluations are:

- The results obtained from the defined set of metrics, collected automatically during the evaluation of different scenarios [1][8][12] (i.e. "basic", "advanced" and "free" scenarios, designed according to different initiative styles and task complexity levels), showed that a **suitable turn-taking algorithm** is essential to ensure a lively and effective dialogue.
- Those results also clearly showed the **learning process** that the user experiences while interacting with the system. Indeed, **"experience" proved to be a key factor regarding dialogue performance**. As the learning stage proceeds, the user is able to exploit the acquired experience leading to more fluent and efficient dialogues. **The user-system interaction improves as the user "learns" how to address the system**. This was supported by the fact that "free" scenarios, though allowing the user the highest degree of initiative and, therefore, favouring much more open and complex expressions, were precisely those that, objectively, performed the best.
- **Expert users are more efficient than novices** (i.e. need less dialogue turns to achieve the same thing). At the same time, novices rely more on contextual information resources. However, both types of users were able to establish productive dialogues with the system since the beginning. The negotiation that the system is able to establish with the users plays a key role in that regard. This negotiation allows users not only to achieve their goals, but also to accelerate the development of their dialogue skills, thereby improving the performance and the quality of the interaction with the system.
- Users tend to need significantly less feedback as they become more familiar with the system. Therefore, the behavior and **the response of the system must be tailored to different skill or experience levels**.
- We have defined a new actuation algorithm that provides **the proper sequence of execution** for those actions corresponding to the positively inferred goals, by combining the **prevalence relations** between those goals (i.e. priority information), and **the order that they appear in the sentence** (i.e. position information). This new solution ensures optimal usability since:
  - the system is acting as soon as it is possible resulting in a much more natural interaction,
  - it is acting even in the case that there are incompleted goals, thus resulting in a flexible interaction,
  - system's actuation is suitable and tidy since it respects both priority and position information, thus resulting in a more robust interaction as well.
- The new actuation algorithm allows optimal usability **solving the problem of potential "blockings"**. Blockings are produced by the observation of active but incomplete goals. This could prevent the execution of every other action corresponding to any goal that, though ready to be executed, either has a lower priority or appears later in the sentence with the same priority. As blockings are avoided, **dialogue performance (i.e. turn efficiency) improves**.

- To better understand how dialogue systems work, we also made **a correlation analysis** between different dialogue metrics. Main results are summarized below:
  - **High values of contextual turns** tend to be associated with **low values of system requests**.
  - **High values of null-efficiency turns** (i.e. out-of-domain sentences or recognition rejections) tend to be associated with a **poor contextuality**.
  - **High values of system requests** tend to be associated with **high values of null-efficiency turns**.
  - **A high turn efficiency** is usually associated with **high contextuality levels**. Yet a high contextuality helps, it does not guarantee a good turn efficiency.
  - **Low values of system requests** mean **good turn efficiencies**.
  - Logically, the lower the null-efficiency turns, the greater the turn efficiency.
- We almost reached the number of **two goals satisfied per turn**. This is a good outcome, especially bearing in mind that users were not given any specification regarding the number of turns in which they had to try to overcome the different scenarios. Therefore, the possibilities of the system in this regard are yet to be fully exploited.
- A better efficiency has led to a more flexible and fluent dialogue which, in turn, has **improved the system's response**. This improvement has been assessed very positively by users. Particularly, **"free" scenarios were ranked as the highest-rated**. This is a result of particular importance since free scenarios lacked of any restriction (i.e. complexity was maximum) and, indeed, were the nearest scenarios to the actual use of the interface.

3.7. Regarding the design methodology

Finally, this thesis delves into the analysis and implementation of efficient mechanisms and techniques that minimize the effort required to generate a new dialogue system (change of semantic context). We proposed the use of strategies for characterizing the application domain and that enable the automatic learning of dialogue models. This methodology allows to obtain a full dialogue model for any application based on the analysis of suitably labeled real situations and a description of the data model along with a semantic description of the application (ontology).

4. General conclusions

The intention of this doctoral thesis was to introduce new ideas whose application to dialogue modelling could prove to be useful. The scientific and technological results obtained enable the design of better devices and intelligent interfaces that fully integrate features that facilitate portability across domains and languages, and improve all aspects of interaction with the end user.

Generally, user satisfaction in relation to a particular system crucially depends on its "usability" and "functionality". To be "useful", a system must be "usable" first (i.e. providing services for which it is designed efficiently) and also "functional" (i.e. the services provided are of interest to users).

One of the keys for the usability of a system, and by extension for its usefulness, is its simplicity of use. The greater or lesser ease of use that a system is able to offer (and also the offered functionality), definitely conditions the final acceptance by users. Easiness of use was the best appreciated feature by users, so it can be considered one of the most important results.

In order to get a fluent and efficient dialogue, the user-sytem interaction should be: **natural, flexible and robust**. It is difficult to attribute each of the above features to a single aspect of the various dialogue solutions proposed. Rather, it is thanks to the synergy of these solutions, to the joint operation of all of them, how those characteristics become true.

In short, a more natural, flexible and robust dialogue is possible thanks to the solutions for dialogue modelling based on BNs that have been suggested This is supported by a good user satisfaction rate and by the results corresponding to the metrics that were automatically collected [1][8][12], which have shown the usefulness and benefits provided by the proposed solutions.

5. Acknowledgements

The authors would especially like to thank all the members of the EDECAN project consortium (TIN2005-08660-C04) who kindly collaborated in the evaluations.

6. References

[1] F. Fernández, PhD Thesis: "Análisis, diseño y aplicación de modelos de diálogo flexibles, contextuales y dinámicos basados en Redes Bayesianas", E.T.S.I.T., Universidad Politécnica de Madrid, Spain, 2008, "http://oa.upm.es/1810/".

[2] H.M. Meng, C.Wai and R.Pieraccini, "The use of belief networks for mixed-initiative dialog modelling", IEEE Trans. on Speech and Audio Processing, 2003, vol.11, n.6, pp.757-773.

[3] F. Fernández et al., "Speech interface for controlling an Hi-fi audio system based on a bayesian belief networks approach for dialog modelling", Eurospeech 2005, Lisboa (Portugal).

[4] J. Ferreiros, F. Fernández et al., "New Word-Level and Sentence-Level Confidence Scoring Using Graph Theory Calculus and its Evaluation on Speech Understanding", Eurospeech 2005, Lisboa.

[5] F. Fernández et al., "Demostración de una interfaz vocal para el control de un sistema de alta fidelidad", Revista Procesamiento del Lenguaje Natural (ISSN 1135-5948), Ed. SEPLN, N 35, Septiembre 2005, pp. 451-452.

[6] F. Fernández et al., "Human spontaneity and linguistic coverage: two related factors relevant to the performance of automatic understanding of ATC speech", IEEE Aerospace & Electronic Systems Magazine (ISSN 0885-8985, JCR 2006: $0,423$), Ed. IEEE-INST (USA), Vol. 21, No. 10, pp. 12-17, October 2006.

[7] R. San-Segundo, F. Fernández et al., "Speech technology at home: enhanced interfaces for people with disabilities", Journal of Intelligent Automation & Soft Computing (ISSN 1079-8587, JCR 2009: $0,349$), Ed. TSI (USA), Vol. 15, No. 4, pp. 645-664, 2009.

[8] F. Fernández et al., "Evaluation of a spoken dialogue system for controlling a Hifi audio system", IEEE SLT 2008, Goa (India).

[9] F. Fernández et al., "A Bayesian Networks approach for dialog modelling: The fusion BN", IEEE ICASSP09, Taipei (Taiwan).

[10] J.M.Lucas, F. Fernández, J.Ferreiros, "Using Dialogue-Based Dynamic Language Models for Improving Speech Recognition", Interspeech 2009 (ISSN 1990-9772), Brighton (UK).

[11] F. Fernández et al., "Flexible, Robust and Dynamic Dialogue Modeling with a Speech Dialogue Interface for Controlling a Hi-Fi Audio System", IEEE DEXA 2010, (ISBN 978-3-642-03572-2, ISSN 1529-4188), Bilbao (Spain).

[12] F. Fernández et al., "HIFI-AV: An Audio-visual Corpus for Spoken Language Human-Machine Dialog Research in Spanish", LREC-ELRA 2010 (ISBN 2-9517408-4-0), Valletta (Malta).

[13] J.M.Pardo, J.Ferreiros, F. Fernández et al., "Automatic Understanding of ATC Speech: Study of Prospectives and Field Experiments for Several Controller Positions", IEEE Transactions on Aerospace & Electronic Systems (ISSN 0018-9251, JCR 2009: $1,230$), In press, 2010.