Adaptive Spoken Dialogue Systems

Anna Hjalmarsson
Centre for Speech Technology
Department of Speech, Music and Hearing, KTH
annah@speech.kth.se
January 2005

Adaptive systems cover a broad range of interactive systems which adjust to new tasks, situations, users or expressions. These systems identify and classify relevant features to develop over time, adjusting their behaviour to different users and situations. The topic of this paper is adaptation in spoken dialogue systems based on features in the dialogue. These systems automatically extract dialogue features to make intelligible inferences about its user in order to adapt dynamically over time. The paper discusses the utility of adaptation in spoken dialogue system including user modelling, basic techniques for user modelling and evaluation of adaptive systems.

1 Introduction

There are several demands that can reasonably be put on dialogue systems in terms of usability, efficiency, and accuracy. Like user interfaces in general, spoken dialogue systems are becoming increasingly complex. At the same time the number of users increases. It is difficult to design one single system that considers all the requirements and individual characteristics from all various users. Usability in system design is generally approached by designing for a representative subset of its users without considering their individual differences (Fischer, 2000). Unfortunately, the stereotyped user that these systems are designed for rarely exists in reality. The intention of allowing users to interact through natural language is to make the interaction as natural and efficient as possible. Using a language we already master will likely allow us to interact instantly with the system without any other specific skills. Unfortunately, this has not been realised even in systems using natural language. Our language behaviour is influenced by our individual goals, experiences, skills and voice characterises. The quality of the interaction in spoken dialogue systems differs significantly between different users and even for the same user between different occasions. A system with the ability to extract information about its users and intelligently use this knowledge to adapt its behaviour dynamically could be both more efficient and pleasant to use (Kass & Finin, 1988). This purpose of this paper is to establish the research area of adaptive spoken dialogue systems in a general discussion about their utility and function. This paper further discusses the basis for adaptive dialogue systems; user modelling, techniques for user modelling and evaluation of adaptive systems.

2 Adaptive spoken dialogue systems

Spoken dialogue systems include a wide range of systems. Regardless of their differences these systems are generally based on components for language understanding, dialogue management, and language generation. These components have been thoroughly described elsewhere (McTear, 2002) and they will only be mentioned briefly in order to set the theme for this paper. *Language understanding* generally includes components for extracting and computing input. However, in adaptive systems these components also need to extract relevant features from the subject's language use in order to create an accurate representation of the user. Extraction of user information will be discussed more thoroughly under Section 3. The main purpose of the *dialogue manager* is to control the flow of the dialogue. This includes determining if the system has elicited adequate information from the user, contextual understanding, information retrieval and response generation. In relation to adaptive dialogue systems dialogue management is crucial for making intelligible inferences from the extracted features and based on these determine what actions to take. The adaptive characteristics of the system will be realised in components for *language generation*. Issues in language generation are what information to include and how it should be presented. Challenges related to dialogue systems and how to approach these using adaptive methods will be the central topic of the rest of this paper.

2.1 Adaptive spoken dialogue systems

This paper focus on *adaptive* spoken dialogue systems, in particular systems that adjust to its users in a flexible and automatic manner based on features in the dialogue. Systems that allow the user to manually change certain system parameters, and thereby adjust the behavior of these systems, are in general called *adaptable systems* (Fink, Kobsa, Nill, 1999). However, adaptable systems will only be mentioned briefly. *Dynamic*[1] adaptation refers to systems that are dynamically updated during runtime and adapts subsequently. Since our language behaviour is most likely influenced by our individual goals, experiences and skills this behaviour can preferably also be used to make intelligible inferences about these characteristics. Dialogue features which are often used to make inferences about user behaviour in spoken dialogue systems are confidence scores for automatic speech recognition, input rate, and time per utterance (Komatani et. al, 2003, Walker et. al, 2000).

Despite the fact that much research in the area of adaptive systems has been done within artificial intelligence (Kass & Finin, 1998) and human computer interaction in general there has also been studies that have focused on spoken dialogue systems. The research area of adaptive dialogue systems can be approached from various perspectives such as levels of communication, system architectural or which dialogue parameters to consider. This paper is an initial approach to address the area of interest from a functional point of view. This perspective was chosen in order to

---

[1] A static model is created initially and is not updated during the interaction with the user (McTear, 1993).

stress the importance of a user-centred approach in which the motivation is the utility of the adaptation. User adaptation can be used for a wide range of purposes.

Jameson (2003) discusses nine different functions in two categories:

(1) **Supporting system use**: taking over parts of routine tasks, adapting the interface, giving advice about system use, controlling a dialogue

(2) **Supporting information acquisition:** helping users to find information, tailoring information presentation, recommending products, supporting collaboration, supporting learning

Supporting information acquisition is often concerned with more domain specific adaptation such as recommending products to users (Johansson, 2004). This paper will, however focus on dialogue systems which adapt to *support system use* in terms of two challenges for spoken dialogue systems: (1) Error handling and (2) Matching expectations with capabilities. These challenges will hopefully also illustrate the need, area of research, and expectations of adaptive dialogue systems.

2.1.1 Error handling

One challenging area of research in spoken dialogue systems is recognition of errors and how to recover from them. Errors and misunderstandings occur at various levels and have several origins. Work up to date has focused on errors at the level of speech recognition. Insufficient speech recognition can cause systems to make false interpretations of the users' intentions and for that reason make incorrect dialogue decisions. However, as speech recognition improves, research at all levels, syntactic, semantic, discourse, and pragmatic, gain attention (Skantze, 2003). According to Turunen and Hakulinen (2001) error handling can be viewed in terms of *error prevention*, *error detection*, and *error recovery*. In non-adaptive spoken dialogue systems error prevention is approached at the design stage. The designer tries to predict potential errors in the dialogue and design the system in a way that prevents these errors from occurring. Adaptive methods may also be valuable to prevent errors since this to some extent can be done during run-time. If the user's future actions can be calculated from previous actions the system might be able to adapt its dialogue behaviour dynamically to prevent errors from occurring.

How May I Help You?

In an experimental study Walker et. al (2000) try to predict and identify problematic dialogues with the spoken dialogue system *How May I Help You?* (HMIHY). HMIHY is an automated customer service using natural spoken dialogue at AT&T Labs. The system tries to detect problematic dialogues early in the interaction based on automatically extracted dialogue features. The machine learning program RIPPER was used to generate a model for classifying problematic dialogues. The model is trained using a number of predefined classes and the dialogue features, which are assumed to be predictors of these classes. The two classes were TASK SUCCESS and PROBLAMTIC. The predictors were for example the number of recognized words per utterance, confidence measure for automatic speech recognition (ASR)

and time per utterance. In an evaluation the model generated by RIPPER performed 8 % better than the random classifier.

A frequently used approach to avoid and recover from errors and misunderstandings is to shift dialogue management strategy (Van Zanten, 1999). If the system classifies the dialogue as problematic for some reason it may be better off using a different dialogue strategy. Adaptive methods can possibly also be valuable to detect and recover from errors. If the system somehow can 'sense' error prone dialogues, a shift of dialogue strategy can be used to detect what type of error has occurred and develop a strategy to recover from it.

2.1.2 Matching expectations with capabilities

Another challenging issue in spoken dialogue systems is how to match user expectations with system capabilities. In the ideal system the user knows immediately how to use the system. However, this is far from reality. Our style of interaction is strongly influenced by our goals, experiences and individual characteristics. We even change style of interaction from one occasion to another depending on contextual circumstances such as stress, purpose and mood. System experience is one dimension that is relatively often considered in adaptive systems (e.g. Van Zanten, 1999, Komatani, Ueno, Kawahara, Okuna, 2003). A user with no system experience will likely need extra guiding. An experienced user, on the other hand, will probably experience such an introduction as irritating and a waste of time. Moreover, too specific questions from the system can be misleading since users tend to believe that the system handles less than it is actually capable of. A system which only gives guidance when needed, that is when lack of guidance leads to errors, can improve the system's performance and the usability.

Van Zanten

Van Zanten (1999) tries to adapt the level of initiative to overcome the problem of various levels of system expertise. The system, which provides train timetable information, adapts initiative strategies based on a hierarchical slot structure. The structure enables questions at various levels of abstraction to provide the user with the accurate level of guidance.

Komatani

Komatani et al. (2003) try to generate cooperative system utterances in a dialogue system that provides timetable information for inner city busses in Kyoto. A user model is generated based on classifications of three different dimensions of user behaviour: (1) *skill level to the system*, (2) *knowledge level on the target domain* and (3) *degree of hastiness*. Skill level to the system is trying to model whether the user is a novice or expert user. Knowledge level on the target domain models the user's domain knowledge, i.e. the geography in Kyoto and its bus routes. Degree of hastiness models the user's level of stress. The adaptation is based on the dialogue features such as rate of input and presence of barge ins. The system adapts initiative, response strategy and the semantics of the responses. For example, if the user is considered to have a high skill level to the system, the system only use open prompts, i.e. allows user initiative. However, if the skill level to the system is

considered as low the system takes the initiative. The level of stress influences the system's confirmation strategy. If the user is not considered as stressed the system confirms input explicitly. However, if the user is stressed input is only confirmed implicitly. The user model based on these three dimensions is automatically derived through decision tree learning. The decision tree is created from a learning algorithm and trained on dialogues that had been collected with the system. The results from an empirical evaluation of the system are presented in Section 4. Decision tree learning is discussed in section 2.2.2.

2.2 User modelling

In the field of natural language processing one of the main objectives is to endow a computer with natural language capabilities. In human-human dialogue we acquire and use knowledge about our conversational partners. For machines to interact in the same natural way it is likely that they too would benefit from acquiring knowledge about its conversational partners. Consequently, the system needs to extract user dialogue characteristics that can be used to make intelligible assumptions about the quality of the interaction. This leads us to the research area of user modelling. According to Kass and Finin (p. 6, 1998):

> *"A user model is the knowledge about the user, either explicitly or implicitly encoded, that is used by the system to improve the interaction"*

The expected utility of user modelling is to acquire a detailed representation of the user to serve as a basis for adaptation. According to Kass and Finin (1998):

2.2.1 Approaches to user modelling

For the purposes of this paper it is not relevant to discuss all various dimensions of user behaviour that have been modelled in detail. A few general aspects of user modelling will be mentioned and then the development of user modelling within the field of spoken dialogue systems will be discussed. For more detailed classifications see McTear (1993) or Kass & Finin (1988).

The emulation approach vs. the complementing approach

There are two different approaches related to user modelling: *the emulation approach* and *the complementing approach* (Fischer, 2001). The emulation approach is based on the metaphor that to improve human-machine interaction it is necessary to implement features of human communication. Following the complementing approach there is a difference between human-human interaction and human-machine interaction. Consequently human-machine interaction has to be studied in its own specific context. It is, however, possible to combine both approaches. Human interaction is an important influence, but in order to develop useful systems, the interaction still has to be studied in its own specific context.

Generic and individual models

It is reasonable to say that all systems have at least one user model. A single model for all users in a system is called a *generic model* (McTear, 1993). A generic model describes the user in general terms and does not change during the interaction. A

model which takes the individual user characteristics in consideration is called an *individual model.*

Long and short term models

A user model can be *long* or *short term.* Long term models are stored and reused in the user's next session with the system. In short term modelling a new model is created for each session.

Explicit and implicit models

From a user perspective information can be provided either *explicitly* or *implicitly.* Consequently, the system designer either chooses to make the user aware that (s)he is contributing to the user model or not. In general, explicit extraction of information is done through questionnaires before or during the interaction. This is a straight forward method which is fairly easy for the system to act upon. However, filling out questionnaires is time consuming and the user might not want to spend time doing this even if it will improve the rest of the interaction. Moreover, the user might not be willing to provide the system with information that will be stored on a long term basis. An implicit approach is to passively observe and extract features from the user's behaviour. Nevertheless, this means that the system has to deal with a larger extent of uncertainty. Since the user is not deliberately contributing with the information it will be less obvious what her/his specific characteristics and preferences are.

Stereotypes

Models based on stereotypes makes predictions about the users based on a small number of observations. Stereotypes can be viewed as a "bridge" between generic and individual models (Elaine Rich, 1979). The stereotype consists of a *body* and *triggers.* A trigger is an observable value or attribute which can be associated with a certain group of users. When a trigger is observed the model associated with that specific trigger is activated. The task is to group a number of assertions that are normally true for a certain group of users in a specific context. To avoid contradictions some sort of conflict resolution is needed. A common approach is to include models of novice and expert users.

2.2.2 Methods for User modelling

Early user models were representations of the users' goals, competence, attitudes, domain knowledge and knowledge about other agents (Zukerman & Litman, 2001). However, most systems focused on one of these dimensions.

Rule based models

Early user models were based on the designer's intuition rather than collected data. These models used hand made inference rules to make assumptions about the users. The rules were based on careful behavioural analyses of the problem at hand. Unfortunately the development of these models was time consuming and could not be applied outside the specific case of study. They may appear intuitive, but their actual benefits were rarely evaluated. Moreover, in this early work there was no clear distinction made between the modelling component and other system components

(Kobsa, 2001). This resulted in system specific models that were difficult to reuse in other systems. The problems related to the rule based models have led to an extended use of statistical models that are 'shallow' and robust. The models in these systems are automatically generated based on empirical data and separated from the other system components. The features used as predictors are various measures of which most can be automatically extracted from the corpus. Statistically based knowledge may be less expressive, but it is often more practical and easy to handle. Statistical models are trained using various methods for machine learning, which is an extensive area of research all on its own.

Statistical models

Statistical models are based on samples, observable values of a number of various parameters, to make inferences about an unknown dependent variable. In predictive models the unknown variable is usually a feature of the user's future behaviour. Artificial Intelligence research on machine learning and reasoning with uncertainty has led to a number of techniques to create predictive statistical models such as *Bayesian networks*, *Reinforcement learning*, and *Decision trees*. A brief, non exhaustive, overview of techniques that has been used for user modelling will be discussed below.

Bayesian networks
Bayesian networks have been a popular technique within user modelling because they inherently propose a theory on how to deal with uncertainty (see Zukerman et. al, 1998). The Bayesian network has representations of both beliefs and logical reasoning. Each node is associated with a table with conditional probabilities that represents the effect of the other nodes on the probability for each possible value for this node. Bayesian networks can be used either to make predictions or to analyze results. In the context of user modelling for discourse planning Bayesian networks are used to predict the user's belief from planned discourse. Assumptions about the user can also be updated based on the user's responses. However, as the network grows the computational effort increases quickly (Müller, 2003).

Reinforcement learning
Reinforcement learning is a machine learning approach where the system learns from interaction (Sutton & Barto, 1998). Rather than trying to theorize how we learn, idealized learning situations are explored and the effectiveness of various learning methods is evaluated. The idea behind the technique is to choose an action that will maximize performance, which is based on feed-back from the surroundings. In adaptive spoken dialogue systems Reinforcement learning has been used to improve the probability to achieve task success in a conversation. In Singh, Kearns, Litman and Walker (2000) this approach is used to determine the system's level of initiative and how often the user's utterances are confirmed. The goal is to optimize the number of dialogues that leads to task success for a certain task. In adaptive systems Reinforcement learning could be used for selecting the best dialogue strategy given a specific user model.

Decision trees

Decision trees are classifiers with a tree-like structure. Decision trees with a top-down algorithm have widely been used to construct classifiers from a set of sample cases. The inner nodes test a condition and the leaf nodes represent a class. The tree input is classified by starting at the root of the tree and executing the instructions in the inner nodes until a leaf is reached. The leaf states the class. In Komatani et. al (2003) a user model is automatically derived through decision tree learning. The tree can contain approximately 30 different features with both semantic information and efficiency measures. Despite the promising results of machine learning user modelling is still problematic. One critical aspect is the observation and interpretation of the user's behaviour. First, the observation obs(x) itself may be noise, and second the interpretations int(obs(x)) is also prone to delivering uncertain results (Müller, 2003).

3 Evaluation

The extra time, effort and money spent on designing adaptive spoken dialogue system has to be rewarding. This is where evaluation comes handy. Unfortunately, evaluation has come in to the development of adaptive systems relatively late. Cross validation is often used to evaluate the classification accuracy of user models generated by machine learning (Komatani et al, 2003, Litman and Pan, 1999). In a cross validation one part of the data is used for training and the rest for testing. The classification accuracy of the model is evaluated by comparing the classifications of the model to classifications made by hand. The problem of using cross-validation based on hand-tagged dialogues is that the actual benefits of user modelling are not addressed. The annotators are often the system designers and the user's subjective judgements are not considered. Nevertheless, cross-validation may be a first helpful step to reveal if the dialogue features extracted in the machine learning are expressive by some means.

Whether the dialogue is classified as problematic or not is not relevant if this information is not used. Empirical evaluations in which the users interact with the system are used to evaluate the actual benefits of user models and user adaptivity. For user models and adaptive methods to be of benefit to the system, the adaptive version has to be superior in some sense, compared to a non adaptive version in a given situation. Moreover, the system should, if possible, be able to adapt effectively and quickly based on features that can be automatically extracted from the dialogues since users might not be willing to spend time providing the system with additional information. Depending variables that are often measured in dialogue systems is the frequency of a certain user behaviour, what kind of behaviour that is used in a certain situation, number of errors, the time to perform a certain task, the time to learn how to use the system, interaction patterns and the users' subjective attitude towards the system. In Komatani et al. (2003) an adaptive system version was evaluated with data collected from 20 novice users. The data suggests that (1) the user model can be efficiently used for novice users, (2) the user model keeps the dialogue from becoming redundant for more experienced users, and that (3) users who received more detailed instructions got faster used to the system. Other

evaluations that support a positive effect of adaptive spoken dialogue systems are Litman and Pan (1999), Chu-Carroll (2000).

Evaluation of dialogue systems is not trivial in the first place and bringing adaptive features to the equation does not make it any easier. It is difficult to rule out that adaptivity yields the best result in a given situation. Litman and Pan acknowledge that it is still possible that some other non-adaptive version of the system can outperform the adaptive version. Moreover, the critical aspects of adaptive user interfaces might be acknowledged in an empirical evaluation. Johansson (2004) points out that adaptive systems violate well established design principles such as the principle of making things visible and the risk of leaving the user without sense of control.

4 Discussion

The first generation of spoken dialogue systems made several important contributions to the area of research that resulted in more sophisticated systems with more far reaching goals. Besides new and more challenging functionalities the major goal has been to create systems that are natural and easy to use. However, there is no such thing as the most efficient way to achieve a goal, because what needs to be done in a specific dialogue depends to a great deal on individual user characteristics such as particular user habits, expectations and preferences. User adaptive systems is an approach to create systems that are natural and easy to use for a new generation of complex systems with an increased number of users. Adaptive systems use sophisticated modelling methods to tailor the system's behaviour to suit the individual user. The systems adapt themselves to the user by reasoning about user characteristics such as preferences, experiences, knowledge and plans to predict the user's goals, correct errors and prevent misunderstandings. However, the effects of adaptation are not all optimistic. Except for the extra time, effort and money adaptive system may be perceived as more unpredictable and error prone than static systems. It is reasonable to question if adaptive methods actually will improve a specific system.

An expressive representation of the user, a user model, is the basis of successful adaptation. The first generation of user models used hand made inference rules based on careful analyzes of user behaviour. Unfortunately, these models were rarely valid outside the system at issue, or even outside the specific dialogue example they were designed to adapt for. Maybe, these models should rather be viewed as cognitive models of human interaction than as functional models for adaptation. A more recent approach to user modelling is statistical user models based on empirically collected data. This is a welcome approach that will hopefully lead to modelling components that are generic and functional. Nevertheless, there are still a number of unsolved issues related to user modelling and adaptive systems. What parameters are relevant to model, what are the implications of these parameters, and how can these observations and implications be made robust to noise? Furthermore, how can we use these implications to adapt systems appropriately and how can the utility of these components be evaluated?

In order to extract relevant features from the user's language behaviour it is reasonable to question which dialogue features are actually available and which are relevant? Are we expressive enough in our interaction with spoken dialogue systems? To automatically detect dialogue features and to make intelligent assumptions about the user is essential in user modelling. However, the predictive models in How May I Help you only predict whether the dialogue is going to be "problematic" or not. It does not say anything about the origin of the problems. To be able to adapt intelligently it would be useful to have more detailed and expressive classifications, preferably classifications that reveal something about the causes of the problems. The next step for the system is to deal with these problems. Which dialogue strategies can be successfully used to deal with poor speech recognition, novice users, experienced users, misunderstandings or other types of errors?

The contributions of evaluation are to pinpoint weak or missing functionalities and to evaluate the performance of individual components or utterances as well as the overall system performance. Unfortunately evaluation has come in fairly late in the development of adaptive dialogue systems. Evaluations of user models have mainly been concerned with their classification accuracy compared to a human annotator. The annotation of the dialogues has often been done from a system design perspective. It would be interesting to try a different approach in which the model is trained and tested based on the user's subjective rating of the dialogues.

Finally, despite the potential problems, adaptive spoken dialogue systems is a promising approach to deal with problems related to individual differences in user behaviour.

5 References

Fink, J., Kobsa, A., Nill, A. (1999): Adaptable and Adaptive Information Provision for All Users, Including Disabled and Elderly People. Available from http://citeseer.ist.psu.edu/fink99adaptable.html

Fischer, G. (2001). User Modelling in Human-Computer Interaction, In: *User Modelling and User-Adapted Interaction* 11 (1-2), 65-86.

Jameson, A., (2003). Adaptive Interfaces and Agents. In: *Human-Computer Interface Handbook*, eds J.A. Jacko and A. Sears, pp 305-330

Johansson, P. (2004). Design and Development of Recommender Dialogue Systems. Licentiate Thesis. *Linköping Studies in Science and Technology*, Thesis no. 1079, LiU-TEK-LIC-2004:08. ISBN 91-7373918-9.

Kass, R., Finin, T. (1988) Modelling the User in Natural Language Systems, In: *Computational Linguistics*, 14, No. 3, 5-22.

Kobsa, A., (2000). Generic User Modelling Systems, In: *User Modelling and User-Adapted Interaction*, 11(1-2), pp 49-63.

Komatani, K., Ueno, S., Kawahara, T., Okun, H, G,. (2003). User Modeling in Spoken Dialogue Systems for Flexible Guidance Generation, In: *Eurospeech*, pp. 745–748.

Litman, D. J., Pan, S., Empirically evaluating an adaptable spoken dialogue system, In: *Proceedings of the seventh international conference on User modeling*, p.55-64, June 1999, Banff, Canada

Litman, D. J., (2000). Towards Adaptive Spoken Dialogue Agents, Available at: http://citeseer.ist.psu.edu/litman00towards.html

McTear, Michael F (1993), User modelling for adaptive computer systems: a survey of recent developments, In: *Artificial Intelligence Review*, 7 p. 157-184.

McTear, M., F., (2002) Spoken Dialogue Technology: Enabling the Conversational User Interface. pp.1-85 ACM Computing Surveys (CSUR)**,** ACM Press New York, NY, USA
(ISSN:0360-0300)

Müller, M. E,. (2003) Learning for User Adaptive Systems: Likely Pitfalls and Daring Rescue. In; *11th GI-Workshop: Adaptivity and User Modelling in Interactive Software Systems in Proc. LLWA 2003,* pages 323-326.

Rich, E., (1979) Stereotypes and user modelling**,** In: *Cognitive Science*, 3(4):329--354.

Singh, S. P., Kearns, M. J., Litman, D. J., Walker, M. A., (2000) Empirical Evaluation of a Reinforcement Learning Spoken Dialogue System, In: *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, p.645-651

Skantze, G., (2003). Error detection in spoken dialogue systems, Available at: http://www.speech.kth.se/~gabriel/courses/DialogueSystems5p/ErrorDetection.pdf

Sutton, R. S., Barto A.G., (1998), Reinforcement Learning: An Introduction, MIT Press

Turunen, M., Hakulinen, J., (2001) Agent-based Error Handling in Spoken Dialogue Systems, In: *Proceedings of the Eurospeech* 2189-2192.

Van Zanten, G., (1999) User Modelling in Adaptive Dialogue Management, In: G. Olaszy, G. Németh and K. Erdöhegyi (eds), *Proceedings of the 6th European Conference on Speech Communication and Technology, Eurospeech-99*, volume 3, pp. 1183–1186.

Walker, M. A., Langkilde, I., Wright, J., Gorin, A. & Litman, D. J. (2000). Learning to Predict Problematic Situations in a Spoken Dialogue System: Experiments with How May I Help You? In: *North American Meeting of the Association of Computational Linguistics.*

Zukerman I., Litman D. (2001) Natural language processing and user modelling: synergies and limitations, In: *User modelling and user adapted interaction*, 11, 129-158.