

BAYESIAN NETWORKS MODEL FOR DIALOGUE ACT RECOGNITION BASED ON INTRA-UTTERANCE FEATURES AND INTER-UTTERANCES CONTEXT

Anwar Ali*, Ramlan Mahmod, Fatimah Ahmad and Md. Nasir Sulaiman

Received: Nov 3, 2005; Revised: Feb 24, 2006; Accepted: Mar 6, 2006

Abstract

Previous dialogue act recognition models assume inter-utterances independency, in which each utterance is independent of the preceding utterance given the preceding utterance dialogue act. Accordingly, in these models, the recognition of the dialogue act of an utterance depends on the linguistic features extracted from the utterance itself and the dialogue act of the preceding utterance. This paper presents a Bayesian Networks model for dialogue act recognition in a dialogue system. In addition to the linguistic features of the user utterance and the previous utterance dialogue act, the presented model employs inter-utterances context which results from relaxing inter-utterance independency assumption. To design the model two sets of linguistic features have been identified, intra-utterance features extracted from the user utterance and context features extracted from the previous utterance. Bayesian networks machine learning has been used to induce the networks from a task oriented dialogue corpus. A series of experimental cases have been conducted to evaluate the Bayesian Networks model. In each case, different features have been used. The results show that the inter-utterance context is an effective factor in the recognition of dialogue act and the model which is based on intra-utterance features and inter-utterances context has the highest recognition accuracy.

Keywords: Dialogue act recognition, dialogue system, bayesian networks, machine learning

Introduction

A dialogue system is a computer system which communicates with human users via natural language to achieve a certain goal such as making travel arrangements, answering questions about weather or sports, routing telephone calls, acting as a general telephone assistant, or performing even more sophisticated tasks. Figure 1 shows the typical architecture of a dialogue system which integrates many

components such as speech recognition, natural language understanding, dialogue manager, natural language generation and text-to-speech synthesis (Jurafsky and Martin, 2003). The role of the natural language understanding component is to convert the user's utterance into semantic representation to figure out his intention so that the dialogue system can generate an appropriate response. In the information-state approach to

Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia. E-mail: anwar_althari@hotmail.com {ramlan, fatimah, nasir}@fsktm.upm.edu.my

** Corresponding author*

dialogue this task is known as Dialogue Act Recognition (DAR). In linguistics, a Dialogue Act (DA) is a concise abstraction of a speaker's intention. Examples of DAs are REQUEST, SUGGEST, ACCEPT or THANK. DAs are derived from speech act theory (Searle, 1969) which interprets utterances as actions and categorises them into speech act categories (Searle, 1975). DAs, however, take into account the context of the utterance and the dialogue state (Bunt, 1994).

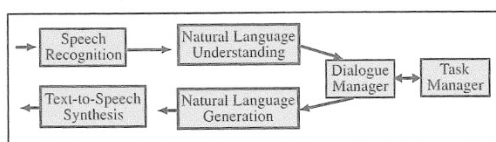


Figure 1. Dialogue system architecture

DAR is formulated as follows; given an utterance with its preceding context, how to determine the DA it realises. It is a challenging task, because most often the DA cannot be directly inferred from a literal meaning of the utterance. A common example is indirect REQUEST DA, in which what looks, literally, as a question is actually a polite form of a directive or a request to perform an action. The utterance *Can you reserve one ticket?* has a literal meaning of a question: Do you have the ability to reserve ticket? But the DA it realises is a REQUEST to reserve one ticket.

The recent efforts in DAR concentrate on employing Machine Learning (ML) techniques (Clark, 2003) as an alternative to the previous efforts which concentrated on logical inference (Perrault *et al.*, 1980). The key idea behind using ML techniques is that each DA in a discourse is reflected in many aspects of its linguistic realisations such as lexical (i.e. words and words sequences) prosodic, syntactic and conversational structure (i.e. DA sequences (Jurafsky *et al.*, 1998)). These features are widely known as DA cues. The key role of ML is the discovery of what Goodwin (1996) called the DAs microgrammars which associate each DA with its cues. Then the discovered micro-

grammars are used to infer DAs. Statistical ML approaches, due to their distinctive properties of modularity and compensability and ability to handle well both rules and exceptions to those rules (Stolcke *et al.*, 2000), have been used extensively in DAR. Different statistical ML models have been experimented (Jurafsky, 2004), N-grams (Reithinger and Klesen, 1997), hidden markov model (Chu-Carroll, 1998; Stolcke *et al.*, 2000), Statistical decision trees (Shriberg *et al.*, 2000), Naïve Bayes (Ivanovic, 2005) and Bayesian Networks (Keizer, 2003)

A common simplifying assumption in previous ML models of DAR is the inter-utterances independency. It has been pointed out by Stolcke *et al.* (2000) and Clark (2003) who recommended further researches to relax this assumption. A review of these models (Jurafsky, 2004) reveals that they depend, essentially, on the features extracted from the utterance itself and the DA of the preceding utterance. Obviously, in these models given the previous utterance dialogue act, two forms of inter-utterances independency are realised. First, the linguistic features of the current utterance are independent of the linguistic features of the preceding utterance. Second, the linguistic features of the current utterance are independent of the preceding utterance DA. The sole form of inter-utterances dependency, which has received attention, is the adjacent DAs dependency which does not always allow one to guess the current utterance DA (Midgley, 2003). It has been argued that the inter-utterances independency assumption deprives DAs microgrammars of a valuable source of contextual cues. The previous utterance might contain lexical or syntactical cues indicating the current utterance DA or indicating other cues in the current utterance. A statistical analysis of the relative frequencies for a number of features from the previous utterance given the DA of the current utterance in the SCHISMA dialogue corpus (English version) furthers this argument. It has been found that some of these features are informative for certain DAs in the current utterance; they have a high relative frequency value. One might argue that the informativeness of these features

for the current utterance DAs are already present in the previous utterance DAs. However, using the statistical analysis of the relative frequencies of the previous utterance features given the current utterance DA, the previous utterance feature given the previous utterance DA and the previous utterance DA given the current utterance DA, it has been found that the informativeness of the previous utterance features are distorted when they are expressed in the previous utterance DA. Extending the source of cues to the preceding utterance and modeling the dependency between them could lead to an accurate DAR model.

On the basis of the above argument, a Bayesian networks models for DAR have been designed. In contrast to the BNs model proposed by Keizer *et al.* (2002) which still preserves the inter-utterance independency assumption, the designed model relaxes this assumption by extending the source of the cues to encompass the previous utterance. More specifically, two sets of features have been identified, intra-utterance features and context features. Bayesian ML has been used to construct the networks and to assess their parameters from the annotated SCHISMA corpus. Several experiments have been conducted to evaluate the role of different sets of features. The results have shown that the model which is based on intra-utterance features set and a previous utterance DA performs better than the model which is based purely on the intra-utterance features. Whereas the performance of the model which is based on intra-utterance and inter-utterances context is the best.

The remainder of this paper is organised as follows. Section 2 introduces the BNs model for DAR. Section 3 describes the corpus, DAs annotation, sets of features and a series of experiments. Section 4 is devoted to the conclusion and future works.

Bayesian Networks and Dialogue Act Recognition

A Bayesian network (BN) is a graphical structure, directed and acyclic, that allows us to represent and reason about uncertain domains

(Korb and Nicholson, 2004). The nodes represent variables of the domain whereas the directed arcs (or links) represent the direct dependencies between them. The strength of the relation between variables is quantified by conditional probability distributions associated with each node. BN requires the assumption of the Markov property in which each variable is conditionally independent of its non-descendents given its parent. The assumption of Markov property in BNs is justified by the concept of d-separation which explains when a change of a certainty in one variable may change the certainty for other variables (Jensen, 2001). The consequence of Markov property is that the joint probability distribution for all the variables can be expressed as the product of the conditional probabilities of each variable. Consider a BN consisting of a set of n variables, $X_n = \{X_1, \dots, X_n\}$, a particular value in the joint probability distribution $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$, or more compactly, $P(x_1, x_2, \dots, x_n)$ can be computed as follows

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i | \text{parent}(x_i)) \quad (1)$$

Four consecutive steps are required to model a certain domain using BNs. First the domain of interest must be represented as random variables. Then, the structure of the network must be specified. That is, to specify the conditional independencies among the variables by means of building a directed acyclic graph. It can be accomplished either by using domain experts to decide on the conditional independencies, or by inducing the network structure from raw, empirical data using ML techniques. After that, the conditional dependencies must be assessed. As is the case with structural learning, domain expert or ML techniques are used. Neapolitan (2003) gives a detailed overview on learning BNs.

Once the domain is represented as a BN, different kinds of reasoning about the domain can be performed, for example diagnosis, prediction, explaining away, etc. Reasoning in BNs consists of computing posterior probability distribution of a set of variables called query

variables X_Q , given the states of another set called evidence variables X_E , where $X_N = X_Q \cup X_E$ as follows

$$P(X_Q | X_E) = \frac{P(X_Q, X_E)}{P(X_E)} \quad (2)$$

The present work is concerned with a kind of reasoning called classification. Given new information about variables in the evidence set X_E , what is the configuration of the class variable Y , that maximises its probability,

$$Y^* = \underset{Y}{\operatorname{argmax}} P(Y | X_E) \quad (3)$$

To apply BNs to DAR domain, the following clarifications and assumptions about the domain should be stated. The dialogue is a task-oriented dialogue between a system and a user. The task is information exchange and transaction. The user is assumed to be rational: there is a dependency between the user utterance and his intentional state. Also the user is assumed to perform one of a finite number of possible DAs in each utterance. There is no information loss between the model that records the utterance and the DAR model.

The domain is represented as random variables, the utterance U produced by the user, Dialog Act DA , represents the intention of the user in his utterance U , and the context C , represent information about the dialogue states until now. The task is to answer the question: what is the most likely DA intended by the user given his utterance U and dialogue context C . That is $\max P(DA | U, C)$. To make this problem tractable, the utterance U is expressed as a set of superficial linguistic features (the features are extracted from the superficial analysis of the utterance such as word occurrence, word count, punctuation symbols ... ect) which represent intra-utterance features. The context C is also expressed as a set of superficial linguistic features from the previous utterance and represents context features. Each variable, from both sets, has a certain value from a finite set of

values. Thus the problem of DAR is to find that value of DA that maximises $P(DA = da | F_1 = f_1, \dots, F_n = f_n, C_1 = c_1, \dots, C_m = c_m)$. Figure 2 shows an example of BNs model of DAR. It consists of one variable which represents the user DA and three variables which represent intra-utterance features (DA_u , *Contains Pos_u*, *Contains Neg_u*, *Contains Do_u*) and four variables represent context features (DA_{u-1} , *StartsWithTh_{u-1}*, *Contains Num_{u-1}*, *StartsWithDoYou_{u-1}*). In this network the goal is to find the value of the variable DA_{u-1} which has the maximum probability given the values of the remaining variables. Applying the concepts of d-separation and causal reasoning to this network can justify the conditional independencies imposed by Markov property. For example in this network, knowing the values of the variable DA_{u-1} blocks the effect which may caused by knowing the values of some variables such as *StartsWithTh_{u-1}* or *Contains Num_{u-1}*.

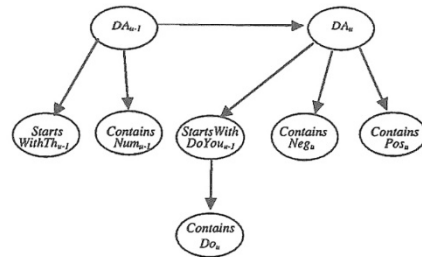


Figure 2. BN model for DAR

The selection of the features should be based on their informativeness for the DAs. The construction of BNs can be achieved either by human expert or ML as mentioned above. In this work ML has been used to construct the networks from annotated SCHISMA corpus. Observations (values of the variables) are extracted from the user utterances and their preceding and presented to the Bayesian ML algorithm to construct the networks.

Method

The methodology of this work consists of four stages: dialogue corpus collection, dialogue corpus annotation, features specification and

Bayesian ML application. The following sections contain a description of the work in each stage.

Dialogue Corpus Collection

The SCHISMA dialogue corpus, a collection of 64 dialogues, has been collected through a series of Wizard of Oz experiments (Dahlback, 1993). The interaction between the wizard (a human simulating the system to be developed) and the human user was established through keyboard-entered utterances, so the dialogues are textual. Furthermore, the dialogues are mixed initiatives: the initiative may switch between the participants within a single dialogue. The task domain concerns information exchange and transaction in a theatre. Users are enabled to make inquiries about theatre performances scheduled and if they wish to make ticket reservations. To get an idea of the corpus, a small fragment is given below (Figure 3).

U: When is Sweeny Todd on?
[query_ref, no_blf]
S: You can see Sweeny Todd in the period December 28 to 30
[influencing_addressee_future_action, positive_answer]
U: What about under a blue roof?
[query_ref, no_blf]
S: You can see 'Under a blue roof' in the 'Grote Zaal' on May
[influencing_addressee_future_action, positive_answer]
U: Can I order a ticket for that
[query_if, no_blf]
S: Do you have a reduction card?
[query_if, hold]
U: I dont have a reduction card.
[statement, negative_answer]
U: Four tickets please
[influencing_addressee_future_action, positive_answer]

Figure 3. Fragment of SCHISMA dialogue corpus with associated FLF and BLF DAs

SCHISMA corpus has been annotated with the SCHISMA-DAMSL annotation scheme developed by Keizer (2003). It was adapted from the DAMSL annotation scheme (Allen and Core, 1997) a standard for annotating task-oriented dialogues. SCHISMA-DAMSL consists of five layers, each of which covers a different aspect of a communicative action that can be performed during a dialogue. The present work is concerned with two layers, Forward-looking function (FLF) and Backward-looking function (BLF). FLF contains acts that characterise the effect an utterance has on the subsequent dialogue, while

acts on the BLF indicate how an utterance relates to the previous dialogue. In the above example, the FLF and BLF of each utterance are given in brackets. In the annotation stage, the layers are assigned to parts of the dialogue text, called segments. Each dialogue in the corpus is first subdivided into turns, which may consist of one or more utterances. Because there may be more than one dialogue act in an utterance which may be performed subsequently, each utterance is subdivided into one or more segments. Segments are then the basic units for DAs labeling. Table 1 and 2 contain FLF and BLF DAs and their frequencies in the SCHISMA.

Table 1. FLF dialogue acts for SCHISMA

FLF	User	System	Total
Statement	123	694	817
Action-Directive	239	11	250
Open-option	0	111	111
Query-if	70	38	108
Query-ref	433	165	598
Committing-speaker-future-action	0	70	70
Conventional	29	31	60
No-flf	0	0	0
Others	26	7	33

Table 2. BLF dialogue acts for SCHISMA

BLF	User	System	Total
Accept	69	54	123
Hold	39	161	200
Reject	13	7	20
Signal-non-understanding	3	20	23
Signal-understanding	7	0	7
Positive-answer	162	399	561
Negative-answer	30	42	72
No-answer-feedback	3	63	66
No-Blf	589	379	968
Other	5	2	7

Features Specification

Two sets of features have been identified. Intra-utterance features and context features. The

selection of these features was based on their informativeness for recognising the FLF and BLF DAs of the user utterance. Measuring the informativeness of a feature is based either on the intuition or on some metric. As it will be explained below, the selection of intra-utterance features were based on intuition, while for the selection of context features a simple metric from Samuel *et al.* (1999) has been used.

Intra-Utterance Features

A set of features proposed previously by Keizer (2003). These features have been redefined to suit the English version of SCHISMA dialogue corpus. The selection of these features were based on the initial intuitions on what features are most informative for identifying the FLF and BLF of an utterance. Table 3 contains a list of intra-utterance features.

Context Features

A set of linguistic features extracted from the utterance that precedes the current user utterance. These features are informative for identifying the FLF and BLF of the current user utterance. Table 4 contains a list of context features.

The selection of the context features was based on a simple metric which takes into account the frequency of features within each DA and the prior distribution of DAs. It is the conditional probability of a feature given a DA. Table 5 and 6 contain $(F_{u-1} | DA_u)$ in FLF and BLF.

$$CP(F) = \max_{DA} P(F | DA) \quad (4)$$

Machine Learning Application

In this stage, Bayesian ML techniques are applied to construct the networks and to assess the conditional probability distribution. To learn the structure of the networks two general approaches can be distinguished: constraint-based (Spirtes *et al.*, 1993) which starts with a fully connected graph, arcs are removed if certain conditional independencies are measured in the

data. search-and-score in which a search is performed through the space of possible DAGs, and either the best one found, or a sample of the models found is returned; because the number of possible DAGs is exponential in the number of nodes, a set of assumptions and a heuristic search method is required, e.g., a local search algorithm like greedy hill climbing or a global search algorithm like Markov Chain Monte Carlo (Heckerman, 1999). The second type of structure learning is the most popular approach. After finding a graph structure for a Bayesian network, whether by using machine learning or not, the required conditional probability distributions for the model must be assessed. The classical method of assessing these values is Maximum Likelihood Estimation (Buntine, 1994)

It is worth mentioning that many BNs software tools have become available. BNs tools vary in aspects such as functionality (support for learning and inference algorithms, graphical interface for presenting and editing networks, export possibilities for re-using networks in other software) and availability (free/share-ware, open-source, programming language). In (Korb and Nicholson, 2004 (Appendix B)) a detailed list of BNs tools is given.

The application of Bayesian ML techniques to design DAR model has been performed in two phases, learning and testing. In the learning phase, the BNs have been constructed and the conditional probabilities have been assessed using the observations extracted from the dialogue corpus whereas in the testing phase the recognition accuracy of the induced BNs is evaluated. A BNs tool called Probabilistic Networks Library (PNL) from Intel, have been used. It is available from <http://www.intel.com/research/mrl/pnl>. PNL supports hill climbing algorithm for learning the structure of the networks and maximum likelihood estimation to assess the conditional probabilities. In (Keizer, 2003) WEKA BNs software tool was used. They applied different combination of Bayesian machine learning algorithms (K2, Maximum A-Posteriori Estimation) to induce BNs for DAR. Figure 4

Table 3. Intra-utterance features

Feature	Type	Description	Intuition
<i>Len_u</i>	Scalar	Number of words in the utterance; one for 1 word, few for 2, 3 or 4 words, and many for 5 or more words	One indicator for short answers like yes or No. few indicator for short sentence like Thank you or good bye
<i>IsContinue_u</i>	Boolean	True if the utterance shows a continuation pattern, i.e, and..., what about...	Indicate that the current utterance has the same FLF as the previous.
<i>StartWithWHExp_u</i>	Boolean	True if the utterance starts with a wh-expression i.e. which, where, when, how ... or proposition followed by wh-word.	This feature is expected to be closely related to the Flf type info-request and query-ref.
<i>EndsWithQuestionMark_u</i>	Boolean	True if the utterance ends with a question mark.	Separate utterances with FLF-type statement from those with FLF-type info-request.
<i>StartsWithIwant_u</i>	Boolean	True if the utterance starts with I want, I would, I prefer.	This feature is also meant for identifying requests.
<i>ContainsPositive_u</i>	Boolean	True if the utterance contains any words from {yes, exactly, certain, indeed, certainly}.	Indicator for positive-answer and accept.
<i>ContainsNegative_u</i>	Boolean	True if the utterance contains words {no, not, none, nothing, don't}.	Indicator for negative-answer and reject
<i>ContainsOkay_u</i>	Boolean	True if the utterance contains word {ok, okay, well, fine, thank, thanks}.	Indicator for accept
<i>ContainsDo_u</i>	Boolean	True if the utterance contains verbs {tell, give or reserve}.	Indicator for action directive

Table 4. Context features

Feature	Type	Description
<i>Flf_{u-1}</i>	Scalar	FLF of previous utterance
<i>ContainsNumber_{u-1}</i>	Boolean	True if the utterance contains numbers
<i>StartsWithDoYou_{u-1}</i>	Boolean	True if the utterance starts With phrase Do You
<i>StartsWithShall_{u-1}</i>	Boolean	True if the utterance starts With phrase Shall I
<i>StartsWithTh_{u-1}</i>	Boolean	True if the utterance starts a word {the, this, there, that, than}

and 5 show examples of BNs induced from annotated SCHISMA corpus. Figure 4 is a BNs for recognising FLF DAs and Figure 5 is a BNs for recognising BLF DAs. In both figures the directed arcs which link the variables subscripted with $u-1$ and the variables subscripted with u represent the inter-utterances context.

To evaluate the recognition accuracy of the induced networks, Junction tree inference algorithm (Lauritzen *et al.*, 1988) supported by PNL has been used.

Results and Discussion

The experiments have been divided into several cases. In each case, there are two recognition tasks, FLF and BLF. Also in each case, different

features are used to represent BNs random variables. 10-fold cross validation is a technique for estimating the accuracy of the learned model on a data set. Due to its popularity in the literature, it has been used in this work. In this technique, the data is split into ten approximately equal partitions, each being used in turn for testing, with the reminder combined for training. The chosen baseline for the experiments is the relative frequency of the most frequent class value. For FLF it is 46.57% while for BLF it is 64.38%

Case 1: Intra-Utterance Features

In these experiments, the intra-utterance features have been used as BNs random variables. The value of these variables come from the user utterance only without considering the

Table 5. $P(F_{u-1}|DA_u)$ in FLF DAs

<i>FLF</i>	<i>StartsWith DoYou_{u-1}</i>	<i>Contains Number_{u-1}</i>	<i>StartsWith ShallI_{u-1}</i>	<i>StartsWith Th_{u-1}</i>
Statement	0.35	0.08	0.02	0.07
Action-directive	0	0.25	0.20	0.21
Open-option	0	0	0	0
Query-if	0.01	0.53	0.09	0.36
Query-ref	0.01	0.38	0.01	0.35
Committing-Speaker- future-action	0	0	0	0
Conventional	0	0.41	0	0.14
Other flf	0.04	0.38	0	0.23

Table 6. $P(F_{u-1}|DA_u)$ in BLF DAs

<i>BLF</i>	<i>StartsWith DoYou_{u-1}</i>	<i>Contains Number_{u-1}</i>	<i>StartsWith ShallI_{u-1}</i>	<i>StartsWith Th_{u-1}</i>
Accept	0.03	0.29	0.55	0.17
Hold	0.03	0.36	0.31	0.26
Reject	0.15	0.08	0.46	0.31
Signal-non-understanding	0	0.67	0	0.33
Signal-understand	0	0.71	0	0
Positive-answer	0.1	0.09	0.01	0.09
Negative-answer	0.09	0	0	0
No-answer-feedback	0	0	0	0

context. The obtained results through 10 runs are summarised in Table 7. The results show that, compared to the baseline, using intra-utterance features result in a significant increase in the recognition accuracy of FLF (18.63%, computed by subtracting the maximum recognition accuracy obtained in both cases). However using intra-utterance features lead to a slight increase in the recognition accuracy of BLF (1.35%). One conclusion to be drawn from the results is that the intra-utterance features are more informative to the recognition of the FLF DAs than the recognition of BLF DAs. It can be interpreted by the high relative frequency of No-blf DA in the user utterances which make training data set abundant with No-blf cases while sparse of the remaining BLF DAs cases. As a consequence the constructed BN has a higher tendency to recognise No-blf than other DAs. The analysis of confusion matrix, like the one given in Table 8, conforms to this interpretation.

Case 2: Intra-Utterance Features and Previous utterance DA

These experiments have been conducted to study the effect of previous DA on the recognition accuracy. Beside intra-utterance features, previous utterance DA, more precisely FLF, has been used as BNs random variables. The observations come from the features extracted from the user utterance and the FLF of its preceding which might be user, system or null utterance. In most cases the utterance that precedes

the user utterance is the system utterance but when the first utterance in the dialogue is the user utterance then the preceding utterance is null. Also when the user turn is divided into multiple segments during the annotation process, the previous utterance is the user utterance. Table 9 contains the results of the experiments. The results are comparable to the results obtained by Keizer (2003) who experimented on the Dutch version of SCHISMA corpus. The results show that using previous utterance FLF leads to a significant increase in the recognition accuracy, particularly in BLF, compared to using intra-utterance features only. This increase is 7.73% for FLF recognition and 21.60% for BLF recognition. It is obvious that the previous utterance FLF is more informative for the recognition of BLF than FLF. It can be interpreted by intuition regarding the definition of FLF and BLF given above in which FLF of the previous utterance is more related to the BLF than the FLF of the current user utterance. One more conclusion which also can be drawn from the analysis of the confusion matrix is that the previous utterance FLF is informative to BLF DAs other than No-blf. The tendency of BLF DAs other than No-blf and the previous utterance FLF to couple themselves with the adjacency pair relation is an intuitive interpretation.

Case 3: Intra-utterance and context features

These series of experiments are an extension to the experiments in Case 2. The intra-utterance features of user utterance along with the previous

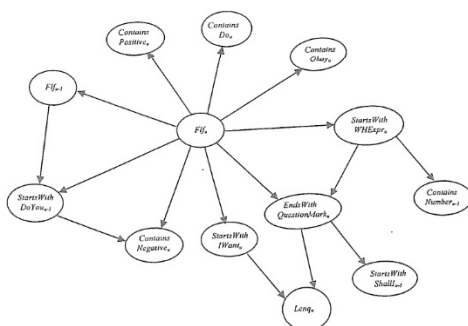


Figure 4. BN for FLF DAs recognition

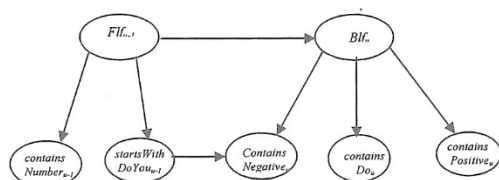


Figure 5. BN for BLF DAs recognition

Table 9. Results of Case 2 experiments

Recognition task	Intra-utterance features	Context features	Accuracy	
			Min%	Max%
FLF	All	Flf_{u-1}	71.6	72.93
BLF	$ContainsPositive_w$ $ContainsNegative_w$ $ContainsOkay_w$ $ContainsDo_u$	Flf_{u-1}	86.27	87.33

Table 10. Results of Case 3 experiments

Task	Intra- Utterance Features	Context Features	Accuracy	
			Min%	Max%
FLF	All	$StartsWithDoYou_{u-1}, Flf_{u-1},$	71.70	75.70
		$StartsWithDoYou_{u-1}, ContainsNumber_{u-1}, Flf_{u-1}$	72.00	77.06
		$StartsWithDoYou_{u-1}, ContainsNumber_{u-1}, StartsWithShallI_{u-1}, Flf_{u-1}$	71.07	73.33
	All except $IsContinue_u$	$StartsWithDoYou_{u-1}, ContainsNumber_{u-1}, StartsWithShallI_{u-1}, Flf_{u-1}$	74.10	75.73
		$StartsWithDoYou_{u-1}, ContainsNumber_{u-1}, StartsWithShallI_{u-1}, StartWithTh_{u-1}, Flf_{u-1}$	71.33	75.20
BLF	All	$StartsWithDoYou_{u-1}, Flf_{u-1}$	86.27	88.00
		$StartsWithDoYou_{u-1}, StartWithTh_{u-1}, Flf_{u-1}$	86.67	87.73
		$StartsWithDoYou_{u-1}, ContainsNumber_{u-1}, StartWithTh_{u-1}, Flf_{u-1}$	86.13	88.00

BNs model of DAR which were constructed using different features extracted from the user utterance (intra-utterance features) and its preceding (context features) in an annotated dialogue corpus. It has been shown that the BNs model that is based on the intra-utterance features and inter-utterances context achieved the highest recognition accuracy, particularly for FLF DAs. For the recognition of BLF DAs, it has been shown that the previous utterance DA, more specifically FLF, is the crucial factor, whereas the intra-utterance features are not relevant as they are for FLF DAs recognition.

In our future research, we plan to pursue issues along three main paths. First, we intend to identify additional linguistic features for both intra-utterance and context features sets which may provide higher DAR accuracies. In this regard it would be interesting to propose a more accurate metric to measure the informativeness of a feature. Second, one factor that could give rise to improved accuracies is the dialogue structure. The dialogue is structured with subdialogues such as subdialogue for opening,

verification or subdialogues related to the task at hand. In this work the dialogue structure has not been taken into account. We intend to include dialogue structure information to improve the recognition accuracy. Third, we intend to test our model on data-sets that are larger, so as to obtain more reliable results with respect to the effects on inter-utterance context features.

Acknowledgments

We would like to thank prof. Anton Nigholt and Simon Keizer from Parlivnk research group, University of Twente, The Netherlands, for their generous sharing of the dialogue corpus, dissertations and ideas.

References

- Allen, J., and Core, M. (1997). Draft of DAMSL: Dialog act markup in several layers. The Multiparty Discourse Group. University of Rochester, Rochester, USA: Available

- from <http://www.cs.rochester.edu/research/cisd/resources/damsl>. Last accessed: Oct 18, 2005.
- Buntine, W. (1994). Operations for learning graphical models. *J. of Artificial Intelligence Research*, 2:159-225.
- Bunt, H. (1994). Context and dialogue control. *Think*, (3):19-31.
- Chu-Carroll, J. (1998). A statistical model for discourse act recognition in dialogue interactions. In: *Applying Machine Learning to Discourse Processing. Papers from the 1998 AAAI Spring Symposium*, p. 12-17.
- Clark, A. (2003). *Machine Learning Approaches to Shallow Discourse Parsing: A Literature Review IM2.MDM/Discourse Parsing WP Project*. Geneva University, Switzerland, 9 p.
- Dahlback, N., Jonsson, A., and Ahrenberg, L. (1993). Wizard of Oz studies-why and how. *Proceedings of the International Workshop on Intelligent User Interfaces (IUI'93)*, Orlando, Florida, USA, p. 193-200.
- Goodwin, C. (1996). Transparent vision. In: *Interaction and Grammar*. Ochs, B., Schegloff, E.A., and Thompson, S.A. (eds). Cambridge University Press, Cambridge, p. 370-404.
- Heckerman, D. (1999). A tutorial on learning with bayesian networks. In: *Learning in Graphical Models*. Cambridge, Jordan, M. (ed). MIT Press, Cambridge, MA, p. 301-354.
- Ivanovic, E. (2005). Dialogue act tagging for instant messaging chat sessions. *Proceedings of ACL Student Workshop 2005*, June, Ann Arbor, Michigan, USA, p. 79-84.
- Jensen, F.V. (2001). *Bayesian Networks and Decision Graphs*. 1st ed. Springer, NY, 268 p.
- Jurafsky, D., and Martin, J.H. (2003). *Speech and language processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. 2nd ed. Prentice Hall series in Artificial Intelligence. Prentice Hall, 960 p.
- Jurafsky, D. (2004). Pragmatics and computational linguistics. In: *The Handbook of Pragmatics*. Horn, L., and Ward, G., (eds). Oxford, Blackwell, p. 578-604.
- Jurafsky, D., Shriberg, E., Fox, B., and Curl, T. (1998). Lexical, prosodic, and syntactic cues for dialog acts. *Proceedings of ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*, p. 114-120.
- Keizer, S. (2003). *Reasoning under uncertainty in natural language dialogue using bayesian networks*, [Ph.D. thesis]. University of Twente, Department of Computer Science, Enschede, The Netherlands, 166 p.
- Keizer, S., op den Akker, R., and Nijholt, A. (2002). Dialogue act recognition with bayesian networks for dutch dialogues. *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*. Jokinen, K., and McRoy, S. (eds); July 2002; Philadelphia, Pennsylvania, USA, p. 88-94.
- Korb, K.B., and Nicholson, A.E. (2004). *Bayesian Artificial Intelligence*. 1st ed. Chapman & Hall/CRC, London, UK, 364 p.
- Lauritzen, S.L., and Spiegelhalter, D.J. (1988). Local computation with probabilities on graphical structures and their application to expert systems. *J. of Royal Statistics B.*, 50:157-224.
- Midgley, T.D. (2003). Discourse chunking: a tool in dialogue act tagging. *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*; July 07-12, 2003; Sapporo, Japan, p. 58-63.
- Neapolitan, R.E. (2003). *Learning Bayesian Networks*. 1st ed. Prentice Hall, 674 p.
- Perrault, C., and Allen, J. (1980). A plan-based analysis of indirect speech acts. *American J. of Computational Linguistics*, 6:167-182.
- Reithinger, N., and Klesen, M. (1997). Dialogue act classification using language models. *Proceedings of the 5th European Conference on Speech Communication and Technology*; September 22-25, 1997; Rhodes, Greece, p. 2,235-2,238.

- Samuel, K., Carberry, S., and Vijay-Shanker, K. (1999). Automatically selecting useful phrases for dialogue act tagging. Proceedings of the 4th Conference of the Pacific Association for Computational Linguistics; August 25-28, 1999; Waterloo, Ontario, Canada, 14 p.
- Searle, J.R. (1969). *Speech Acts*. Cambridge/GB: University Press.
- Searle, J.R. (1975). A taxonomy of illocutionary acts. In: *Language, Mind Knowledge*, Minnesota Studies in the Philosophy of Science. Gunderson, K. (ed). University of Minnesota Press, Amsterdam, (7):344-369.
- Shriberg, E., Bates, R., Stolcke, A., Taylor, P., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., Ess-Dykema, C.V. (2000). Can prosody aid the automatic classification of dialog acts in conversational Speech? *Language and Speech (Special Issue on Prosody and Conversation)*, (41):439-487.
- Spirites, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction and Search*. Number 81 in *Lectures Notes in Statistics*. Springer-Verlag, NY.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C.V., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational Speech. *Computational Linguistics*, 26(3):339-373.