Dialogue Act Modelling Using Bayesian Networks

Simon Keizer
Parlevink Language Engineering Group
University of Twente
The Netherlands
skeizer@cs.utwente.nl

Abstract. A probabilistic approach to interpretation of natural language utterances in terms of dialogue acts is proposed. It is illustrated how using Bayesian Networks, partial information obtained from an NLP component can be combined with knowledge the agent has about the state of the dialogue and about the user, in order to find the most probable dialogue act made.

1 Introduction

In this paper, we are dealing with a conversational agent (the 'server'), which participates in a dialogue with another agent (the 'client'). This conversational agent perceives utterances of the client and tries to react to these utterances in an appropriate way. In Figure 1.1, a possible architecture for the conversational agent is given.

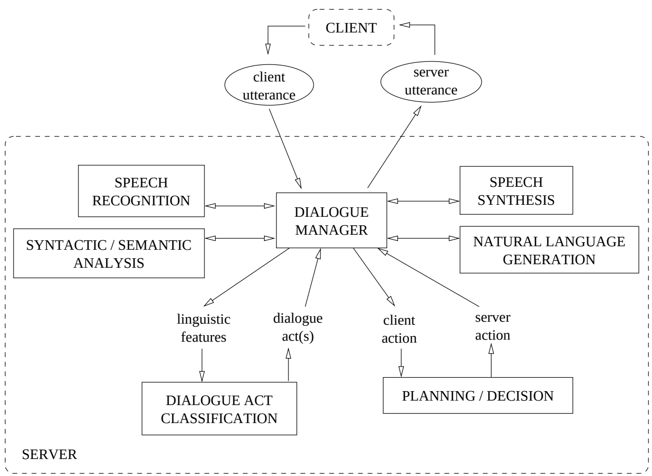

Figure 1.1: An architecture for a conversational agent

The dialogue manager coordinates the various steps involved in the interpretation of an incoming user utterance and the planning of what action to take next. First of all, an incoming utterance is submitted to the components of speech recognition (in case of spoken dialogue), and syntactic/semantic analysis. Next, the pragmatic aspect of identifying what *communicative action* was performed by the user in uttering the sentence is dealt with. This part is what we will be concentrating on in this paper. Finally, the resulting interpretation will have to lead to a decision on what actions to take, in particular, what communicative action to be addressed to the user.

In identifying the communicative action performed in an utterance, the server has to deal with uncertainty. This uncertainty arises because of the incompleteness of the information that is provided by the speech recognition and syntactic/semantic analysis components. In general, such components cannot provide all linguistic information that is contained in an utterance, especially because we are dealing with human speakers that may produce utterances that are partly unrecognised through bad pronunciation, sentences that are ungrammatical, or that contain unknown words. This is especially the case in mixed-initiative dialogues, where a client may tend to use more complex utterances.

In order to deal with the uncertainty, the server will have to make *educated guesses* in the interpretation of the client's utterances. Therefore, we will take a probabilistic approach to dialogue modelling, in the form of Bayesian networks.

In Section 2, we will describe our approach to identifying communicative actions in terms of *dialogue acts*. In Section 3, we will introduce the notion of Bayesian networks and how they can be used in *dialogue act classification*. In Section 4, we discuss some related work and finally, in Section 5 some conclusions are drawn and an indication of further research is given.

2 Dialogue Act Modelling

The notion of dialogue acts is originated in the work of Austin and Searle [Aus62, Sea69]. They observed that utterances are not merely sentences which can be either true or false, but should be seen as (communicative) actions. Searle introduced the theory of *speech acts*, which he defined as the whole act of uttering a sentence. He gave a categorisation of types of speech acts, including e.g. REQUEST, ASSERTION, and ADVICE.

When we speak of *dialogue acts* however, we emphasise the importance of looking beyond the boundaries of an utterance itself when analysing that utterance (see also [Tra99]). Besides linguistic information of the utterance in isolation (prosodic features, syntactic/semantic features, surface patterns, keywords, etc.), the meaning of dialogue utterances may also be determined by information concerning two other aspects:

1. the state of the dialogue: e.g. the current topic or the communicative act(s) performed in the previous utterance(s);

2. the state of the hearer's model of the speaker (e.g. the state of the model that our server has of the the client, the so-called *user model*).

Based on these three aspects, a categorisation of different dialogue act types can be made. The dialogue act hierarchy we use, is based on an existing dialogue act hierarchy underlying an annotation scheme, called DAMSL [AC97]. This scheme has been developed as a standard for annotating task-oriented dialogues. We will not go into the details of this hierarchy here, but just mention some of the dialogue act types that are relevant in the remainder of this paper:

- `q_ref`: the speaker requests the hearer for information in the form of references satisfying some specification also given by the speaker (e.g. a list of theatre performances scheduled on a specified date: "what operas are on next week?").
- `q_if`: the speaker asks the hearer if something is the case or not ("do you want to make reservations?").
- `req`: the speaker requests the hearer for a non-communicative action ("two tickets please").
- `neg_ans`: often in response to a `q_ref` previously performed by the hearer, the speaker indicates that no references satisfy the given specification ("there are no operas scheduled for next week").

Because of the need to be able to reason under uncertainty as explained in Section 1, we propose the use of probability theory in modelling the various relationships involved in interpreting dialogue utterances. Our model will consist of three interrelated components:

1. the Belief State of the server: this state is determined by beliefs concerning:
   (a) the course of the dialogue, and
   (b) the beliefs, desires and intentions of the client.
2. the Dialogue Act(s) performed in the client's utterance;
3. the relevant Linguistic Features that the client's utterance may contain.

Using this model, the server can calculate what most probably must have been the dialogue act performed in an utterance of the client, given new information w.r.t. linguistic features of that utterance, obtained from the speech recognition and syntactic/semantic components. We will now describe how this can be done, using probabilistic inference in a Bayesian network.

3 Bayesian Dialogue Act Classification

The model introduced in Section 2 is described by a set of discrete *Random Variables* (RVs), i.e. variables that describe events with different possible outcomes. In Table 1.1 some two-valued, Boolean RVs in this set are given with their meaning and which of the components, as indicated in Section 2, they are associated with. It should be noted that it is not a requirement that the all RVs are Boolean; we could also have a RV that represents the previous dialogue act of the server S, its values ranging over all possible dialogue act types.

To illustrate the influences that exist between the RVs, consider the following dialogue passage, where we are interested in the dialogue act performed in utterance (3):

(1) C: Wat gebeurt er komend weekend 19 maart in de schouwburg? (What is happening in the theatre next weekend March 19?)

(2) S: Op deze datum is er geen uitvoering. (On this date no performances have been scheduled.)

(3) C: En op 18 maart? (What about March 18?)

RV	meaning	component
PSN	the previous dialogue act of S was a `neg_ans`.	Belief State
PCQ	the previous dialogue act of C was a `q_ref`.	
CQ	C performed a `q_ref` in the current utterance.	Dialogue Acts
CR	C performed a `req` in the current utterance.	
CONT	the current utterance shows a continuation pattern, e.g. in Dutch, if it starts with the word "en".	Linguistic Features
QM	the current utterance contains a question mark.	

Table 1.1: The RVs of the example network of Figure 1.2.

Our judgement that a `q_ref` was performed in (3) may be determined by the observation that C previously performed a `q_ref` in (1), and that S's previous dialogue act in (2) was a `neg_ans`. In this context, C has decided to continue his previous dialogue act, but with a different specification accompanying that `q_ref`. Note that all this cannot be concluded by just analysing the linguistic features of utterance (3) only.

A probabilistic model is completely specified by a *joint probability distribution* (jpd), which assigns a number between 0 and 1 to every instantiation of the RVs. However, the number of probabilities to be assessed in order to specify the jpd increases exponentially with the number of variables. This problem can be overcome by identifying *conditional independencies* between RVs, reducing the number of probabilities needed for specifying the jpd. For example, we may indicate that if we know the value of CQ and CR, then learning the value of PCQ gives us no information on QM: CQ and QM are conditionally independent, given CQ and CR.

These conditional independencies can be specified by means of a *Bayesian Network*. A Bayesian network [Pea88] is a DAG (Directed Acyclic Graph) in which the nodes represent RVs and the arcs reflect the informational dependencies between these variables. In Figure 1.2, a Bayesian Network is depicted containing the RVs given above. It reflects a number of conditional independencies, including the one indicated above.

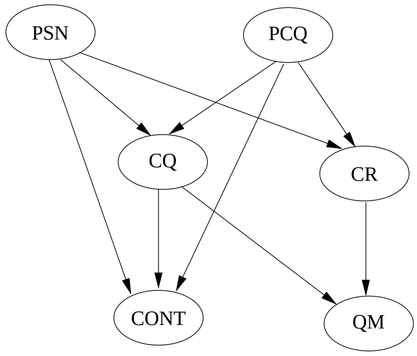

Figure 1.2: Simple Bayesian Network for utterance interpretation.

Associated with each RV is a *conditional probability distribution* (cpd) given its parents in the network. This means that for CQ a cpd is specified, given its parents PSN and PCQ. In the cpd's we should have numbers which reflect the qualitative relationships between the RVs involved. In Table 1.2, some of the chosen distributions are given. From the distribution of CQ, one can see that if PSN=*true* and PCQ=*true*, then CQ is more probably true (0.7) than false (0.3). For this example, the numbers have been assessed by using 'expert' knowledge; for more accurate and realistic models, statistical analysis on data using an annotated dialogue corpus is needed for the assessment (see for example [Hec95]).

PSN	
true	0.75
false	0.25

PCQ	
true	0.8
false	0.2

CQ				
PSN	true		false	
PCQ	true	false	true	false
true	0.7	0.2	0.2	0.5
false	0.3	0.8	0.8	0.5

Table 1.2: Some of the probability distributions specifying the Bayesian network of Figure 1.2.

The RVs concerning the Belief State, *PSN* and *PCQ*, have no parents, so their cpd's are not conditional, but 'prior' distributions (also given in Table 1.2). Although we have specified all prior distributions in the network as fixed, the distributions of RVs like these actually depend on the Belief State at the time-step in which the previous utterance was processed, and is therefore subject to updating. However, in order to keep our story within limits, we will not go into this dynamic extension of our model, but have based the prior distributions intuitively on the course of the dialogue passage given before. Here, the previous act by S, performed in utterance (2), was probably a `neg_ans` (probability 0.75) and the previous act of C, performed in utterance (1) a `q_ref` (probability 0.8).

This network can now be used for *probabilistic inferences*: we can determine the probability that e.g. a `q_ref` was performed, given the *partial* information that e.g. C's utterance contained a question mark, like in (3) of the dialogue passage given before. This process of determining the *posterior probability distribution* is called *belief updating*. In this process, the formula for the joint probability distribution (jpd) over all RVs in the network plays a central role. By making use of the conditional independencies implicitly given by the network structure, the jpd is given by the product of the specified cpd's:

$$\begin{aligned} P(PSN, PCQ, CQ, CR, CONT, QM) = P(PSN) \cdot P(PCQ) \cdot P(CQ|PSN, PCQ) \cdot \\ P(CR|PSN, PCQ) \cdot P(CONT|PSN, PCQ, CQ) \cdot P(QM|CQ, CR) \end{aligned} \tag{1.1}$$

Suppose the server gets to know that there was a question mark in the utterance. He will be interested in the updated probability that the dialogue act performed is a `q_ref`, given this new information. This probability can be calculated as follows:

$$P(CQ = true|QM = true) = \frac{P(CQ = true, QM = true)}{P(QM = true)} \tag{1.2}$$

Both numerator and denominator can be obtained from the jpd (1.1) by summing over all possible configurations of the other RVs in the network. Let $S = s$ and $T = t$ denote

prior distributions	$\mathcal{E}$	$P(CQ = true \vert \mathcal{E})$	$P(CR = true \vert \mathcal{E})$
$P(PSN) = 0.75$ and $P(PCQ) = 0.8$	(none)	0.515	0.298
	(QM=true)	0.562	0.380
	(QM=true, CONT=true)	0.804	0.313
$P(PSN) = 0.5$ and $P(PCQ) = 0.5$	(none)	0.400	0.413
	(QM=true)	0.521	0.537
	(QM=true, CONT=true)	0.624	0.485

Table 1.3: Results of Probabilistic Inferences.

instantiations of the RVs in $\{PSN, PCQ, CR, CONT\}$ and $\{PSN, PCQ, CR, CONT, CQ\}$ respectively. Then we get our posterior probability from 1.3 and 1.4.

$$P(CQ = true, QM = true) = \sum_s P(CQ = true, QM = true, S = s) \tag{1.3}$$

$$P(QM = true) = \sum_t P(QM = true, T = t) \tag{1.4}$$

We will now show some results of probabilistic inferences in the network. For three different cases of particular information (which will be called the 'evidence' $\mathcal{E}$), we have calculated the posterior probability distribution of *CQ* and *CR*, given that information. This has been done with two different choices for the prior distributions of *PSN* and *PCQ*. From the results in Table 1.3, one can observe that with the original priors (upper row of the table), the probability that `q_ref` was performed changes as the available information varies. Especially the case where S gets to know that the utterance shows a continuation pattern (starting with the word "en"), clearly reflects the correctness of classifying the utterance as a `q_ref` (with probability 0.804) and not `req` (with probability 0.313).

In the case of 'uniform' distributions for the priors (lower row of the table), i.e. the probabilities of both values $true$ and $false$ are 0.5 for both *PSN* and *PCQ*, one can observe that there is much more indifference between *CQ* and *CR* than before. This illustrates how the role of beliefs concerning the course of the dialogue (as part of the Belief State) in dialogue act classification can be taken into account.

4 Related Work

Other work on dialogue modelling which is based on the notion of dialogue acts includes e.g. Dynamic Interpretation Theory (DIT) and the Information State model [PT98]. In DIT [Bun95], dialogue acts are defined as functional units used by the speaker to change the context. They consist of a *semantic content* and a *communicative function*, so a dialogue act changes the context in a way that is given by the communicative function, using the semantic content as a parameter. According to the Information State model, both of the dialogue participants keep track of the *Conversational Information State* (CIS), in which *grounded conversational acts* are recorded and also *ungrounded contributions*. A CIS is characterised by a feature structure, containing embedded feature structures for both dialogue participants.

Concerning the classification of communicative actions, various research has been done. In plan-based approaches [PA80], communicative acts (speech acts) are predicted on the basis of recognition of plans that the speaker has. Therefore, the interpretation of utterances is extended from identifying direct speech acts from linguistic features, to indirect speech acts, taking into account the course of the dialogue in terms of a speaker's plans. This rule-based approach may lead to difficulties when dealing with uncertainty.

In other approaches, statistical methods are used to model dialogue, see for example [NM94, MNN+96, Sto00]. In [Sto00], a probabilistic model obtained from statistical analyses of a dialogue corpus is presented. Dialogues are modelled in a Hidden Markov Model, with states corresponding to dialogue acts and observations corresponding to utterances (in terms of word sequences, acoustic evidence and prosodic features). The transition probabilities are obtained from n-gram analysis of dialogue acts and the observation probabilities are given by local utterance-based likelihoods.

Also the use of Bayesian networks for interpreting utterances in a dialogue has been proposed before. In [Pul96], Stephen Pulman proposes a framework for classifying communicative actions, very similar to our approach, in which a framework of conversational games and moves is used, in stead of dialogue acts. A more interesting difference with our approach however, is in the structure of the Bayesian network used. While we have chosen for arcs from nodes representing dialogue acts to nodes representing linguistic features (like the arc from CQ to QM), Pulman has chosen arc in the opposite direction. One could say that in our approach a model of the speaker's behaviour is given, which is used to derive the most probable dialogue act he performed. Pulman's network however, models the hearer, using various sources of information as 'inputs' to derive the most probable conversational move made by the speaker. This difference is an interesting topic for further research.

Other work on the use of Bayesian networks in dialogue systems includes research, where the emphasis is on the *user modelling* part within a specific (task-)domain [AT94], in stead of the aspect of dealing with partial linguistic information in understanding communicative behaviour.

5 Conclusion

In this paper we have shown a probabilistic approach to interpreting natural language utterances in a dialogue. We have described how Bayesian networks can be used to interpret partial information from natural language analysis in terms of dialogue acts. Not only the linguistic information from utterances is taken into account, but also knowledge about the course of the dialogue and about the mental state of the speaker, the dialogue participant that performed the utterance. Using Bayesian networks, these various sources of information can be integrated into one probabilistic model.

Current research questions are related to the construction of a Bayesian network for dialogue act classification. This means that we have to find a set of random variables representing all relevant aspects of identifying a dialogue act type, identify a sufficient number of conditional independencies among these variables (i.e. finding the network structure), and assess the conditional probability distributions associated with the network. One way of doing that is using an annotated dialogue corpus to train a Bayesian network from data. Ongoing work includes the annotation of a corpus of dialogues, which was obtained from Wizard of Oz experiments. The corpus contains mixed-initiative dialogues between a human user typing utterances and a system also producing textual utterances. The dialogues concern the theatre domain, in which the user can get information about performances and ticket reservations can be made if required. The annotation currently concerns dialogue acts (from a dialogue act hierarchy based on the DAMSL scheme, as mentioned in Section 2) and superficial linguistic features like sentence type and punctuation.

Bibliography

[AC97] J. Allen and M. Core. Draft of DAMSL: Dialog Act Markup in Several Layers. Dagstuhl Workshop, October 1997.

[AT94] T. Akiba and H. Tanaka. A Bayesian approach for user modeling in dialogue systems. Technical Report TR94-0018, Tokyo Institute of Technology, 1994.

[Aus62] J. L. Austin. *How to Do Things with Words*. Harvard University Press, 1962.

[Bun95] H. C. Bunt. Dynamic interpretation and dialogue theory. In M.M. Taylor, D. G. Bouwhuis, and F. Neel, editors, *The Structure of Multimodal Dialogue*, volume 2. John Benjamins, A'dam, 1995.

[Hec95] David Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, 1995.

[MNN+96] M. Mast, H. Niemann, E. Noth, E. Guenter, and E. Schukat-Talamazzini. Automatic classification of dialog acts with semantic classification trees and polygrams. In S. Wermter, E. Riloff, and G. Scheler, editors, *Connectionist, Statistical, and Symbolic Approaches to Learning for Natural Language Processing*, volume 1040 of *Lecture Notes in Artificial Intelligence*, pages 217–229. Springer-Verlag, 1996.

[NM94] M. Nagata and T. Morimoto. First steps towards statistical modeling of dialogue to predict the speech act type of the next utterance. *Speech Communication*, 15:193–203, 1994.

[PA80] C. R. Perrault and J. Allen. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3-4):167–182, 1980.

[Pea88] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[PT98] M. Poesio and D. Traum. Towards an Axiomatization of Dialogue Acts. In J. Hulstijn and A. Nijholt, editors, *TwenDial'98: Formal Semantics and Pragmatics of Dialogue*, number 13 in TWLT, 1998.

[Pul96] Stephen G. Pulman. Conversational games, belief revision and Bayesian networks. In Kees van Deemter Jan Landsbergen, Jan Odijk and Gert Veldhuijzen van Zanten, editors, *Computational Linguistics in the Netherlands*, 1996. SRI Technical Report CRC-071.

[Sea69] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

[Sto00] A. Stolcke et al. Dialogue act modelling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–374, 2000.

[Tra99] David R. Traum. Speech acts for dialogue agents. In M. Wooldridge and A. Rao, editors, *Foundations of Rational Agency*, pages 169–201. Kluwer, 1999.