Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language 24 (2010) 190–218
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

# Dynamic Bayesian networks and variable length genetic algorithm for designing cue-based model for dialogue act recognition

Anwar Ali Yahya [a,*], Ramlan Mahmod [b], Abd Rahman Ramli [a]

[a] *Intelligent System and Robotics Laboratory, Institute of Advanced Technology, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*
[b] *Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

Received 2 December 2007; received in revised form 24 March 2009; accepted 16 April 2009
Available online 6 May 2009

**Abstract**

The automatic recognition of dialogue act is a task of crucial importance for the processing of natural language dialogue at discourse level. It is also one of the most challenging problems as most often the dialogue act is not expressed directly in speaker's utterance. In this paper, a new cue-based model for dialogue act recognition is presented. The model is, essentially, a dynamic Bayesian network induced from manually annotated dialogue corpus via dynamic Bayesian machine learning algorithms. Furthermore, the dynamic Bayesian network's random variables are constituted from sets of lexical cues selected automatically by means of a variable length genetic algorithm, developed specifically for this purpose. To evaluate the proposed approaches of design, three stages of experiments have been conducted. In the initial stage, the dynamic Bayesian network model is constructed using sets of lexical cues selected manually from the dialogue corpus. The model is evaluated against two previously proposed models and the results confirm the potentiality of dynamic Bayesian networks for dialogue act recognition. In the second stage, the developed variable length genetic algorithm is used to select different sets of lexical cues to constitute the dynamic Bayesian networks' random variables. The developed approach is evaluated against some of the previously used ranking approaches and the results provide experimental evidences on its ability to avoid the drawbacks of the ranking approaches. In the third stage, the dynamic Bayesian networks model is constructed using random variables constituted from the sets of lexical cues generated in the second stage and the results confirm the effectiveness of the proposed approaches for designing dialogue act recognition model.


*Keywords:* Dialogue act recognition; Dynamic Bayesian networks; Variable length genetic algorithm; Lexical cues selection

## 1. Introduction

Dialogue act (hereafter DA) is a concise abstraction of speaker's intention-what a speaker is trying to achieve by his utterance. It has roots in several language theories of meaning, particularly speech act theory

* Corresponding author. Tel.: +967 734158210; fax: +967 06 509594.
*E-mail address:* anwar_althari@hotmail.com (A.A. Yahya).

0885-2308/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.csl.2009.04.002

(Austin, 1962) which interprets any utterance as a kind of action, called speech acts, performed by a speaker and categorised them into speech acts categories (Searle, 1975). DA, however, extends speech act by taking into account the context of the utterance (Bunt, 1994). Fig. 1 is a hypothetical dialogue annotated with DAs.

Dialogue act recognition (DAR) is a task of crucial importance for the processing of natural language dialogue at discourse level in various applications such as dialogue systems, machine translation, speech recognition, and meeting summarisation. For example, it conditions a successful interpretation of user's utterance which is the main function of natural language understanding unit in dialogue systems. Formally, it is defined as follows: given an utterance with its preceding context, how to determine the DA it realises.

On the other hand, the task is challenging because most often DA is not expressed directly in speaker's utterance, and consequently the meaning of the utterance is not the intended meaning. For instance, a dialogue system without DAR ability interprets the utterance "*Can you reserve three tickets for me*?" as if a user questioning its ability to reserve tickets, whereas his actual intention is a request to reserve three tickets. Obviously, such dialogue system is inadequate.

The literature of DAR indicates that the endeavours for modeling DAR started early seventies and resulted in two types of models (Jurafsky, 2004; Jurafsky and Martin, 2000). The models of the first type, known as plan-based model (Cohen and Perrault, 1979; Perrault and Allen, 1980; Allen and Perrault, 1980), are based, essentially, on belief logic to infer the meaning of the utterance, and then use it to infer the DA in a subsequent stage as depicted in Fig. 2. These models tend to be very time consuming of both human labour in development of plan inference heuristics and system time in running these heuristics (Jurafsky and Martin, 2000).

The models of the second type, known as cue-based model (Stolcke et al., 2000), are characterised by extensive use of Machine Learning (ML) approaches to automatically discover association rules between surface linguistic cues of utterances and DAs as shown in Fig. 3. This particular aspect of cue-based model displaces the burden of manual design of the association rules from human expert and makes these models more attractive from computational point of view (Jurafsky, 2004).

Among the wide spectrum of ML approaches investigated for constructing cue-based models (Fishel, 2007), the statistical approaches are the most prominent, due to their distinctive properties of modularity and ability to handle well both rules and exceptions to those rules. *N*-Gram (Reithinger and klesen, 1997), classification and regression tree (Shriberg et al., 2000), hidden Markov models (Wright, 1998; Chu-Carroll, 1998; Stolcke et al., 2000), naïve bayes (NB) (Grau et al. 2004; Ivanovic, 2005), static Bayesian networks (SBNs) (Keizer et al., 2002; Keizer and op den Akker, 2007), and maximum entropy (Lesch, 2005), as statistical ML approaches, have been explored. Besides that, non-statistical ML approaches have also been investigated such

| Speaker | Utterance | DA |
|---|---|---|
| A | Hello. | `greet` |
| A | I want to see you today at 2:00. | `suggest` |
| B | I' m busy at this time. | `reject` |
| B | but I'm free at 3:00. | `suggest` |
| A | Okay, that sounds fine to me. | `accept` |
| A | I'll see you then. | `bye` |

Fig. 1. Hypothetical dialogue annotated with DAs.

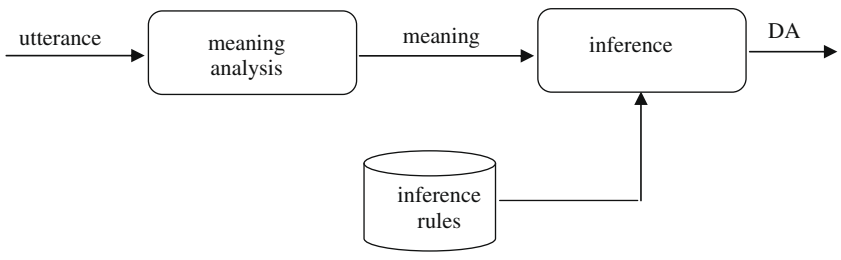


Fig. 2. Plan-based DAR model.