A DYNAMIC DIALOG MODEL FOR HUMAN MACHINE COMMUNICATION

D. LUZZATI

LIMSI/CNRS - BP 133 - 91403 ORSAY CEDEX - FRANCE

Abstract

With vocal or written input, the real problem of human machine dialog deals with the dialog function itself. Taking into account analysis and recognition performance, we consider that the task of a dialog module is to avoid breaking off of communications as much as possible through the use of a dialog strategy. To this purpose, we describe a model using two axes and a set of variables to describe a model using two axes and a set of variables to build a graph allowing any dialog state to be situated in a metric space at any time. Due to the model, the system may control all dialog phases to avoid some critical zones or, as a last resort, be aware of their proximities.

1. Purpose of the model

Human machine dialog is rarely analysed as a complete phenomenon for itself; on one hand it is taken in account to show comprehension phenomena (Winograd 83); on the other hand, there is research that deals with human-machine dialogs as well as with human-human ones (Allen & Ferrault 80). But understanding does not mean the same thing when the purpose is to parse an utterance or to guide a dialog; the presence of the machine means to include the linguistic behavior of the speakers. With dialog, comprehension is a dynamic phenomena, instead of the static one of parsing. The problem is more one of pursuing interaction so the dialog can succeed rather than on of perfectly understanding what was said.

When the human-machine dialog is considered in itself (Pierrel 87), the 4 following problems are generally pointed out:

(a) The dialog history : the system has to be able to know what has been said since the beginning of the discussion.
(b) The task model : according to the task, the system has to be able to identify the goals and the plans of the speakers.
(c) The user model : the system has to identify the kind of speaker it is talking to and take this in account.
(d) The dialog model : the problem is to determine how the dialog can continue, and particularly when the system can admit precision, reformulation, explanation or confirmation questions.

We consider that the main point is that understanding and interaction are very close. In a dialog situation, an utterance has to be considered in respect to the preceding and the following ones. To understand supposes carrying out the necessary inferences between what has been said and what is said (dialog history). It supposes that the goals of the speakers are determined in order to respond to his wants as well as to the wording (task model). It also supposes that it can adapt to the speaker, to his age, his knowledge, his origin ... (user model). It supposes that it knows when interaction can go on ; when it is suitable to continue answering and when it is preferable to stop (dialog model).

(a), (b) and (c) increase the understanding level, but the system needs a robust computational dialog model to be able to take these advantages in account. In Example 1 for instance, even if, to understand U2, the system should consider that the goal of the speaker is to find a city with a garage and a hotel, the main dialog problem is to be able to use this understanding and, eventually, to assume partial or complete misunderstanding.

U1 : I am looking for a garage
M1 : where ?
U2 : is there any hotel here ?

Example 1 : short dialog containing a misunderstanding risk.

In U2, the dynamic dialog model does not allow us to discover the speaker's goals (Allen & Ferrault 80). It allows, when comprehension is uncertain or defective (i.e. each time the system experiences comprehension or each time dialog is supposed to be spontaneous), either anticipation or the pursuit of interaction and recovery of possible errors. The dynamic dialog model thus offers a choice between various answers even if these answers suppose various comprehension levels, according to the state of dialog. A dialog system has to be able to apply its reaction to the context and therefore react differently to the same utterance when it appears in different places in a dialog.

Example 2 for example, gives various possible answers to follow U2 in Example 1. Every automatic dialog
system has to make a choice according both to the comprehension level and the state of the dialog:
- If U2 hasn’t been understood, the system can choose between (a), (b) and (h): (a) leave the user free to ask another question to replace U2; with (b), which is a reformulation of M1, the user is constrained to follow the system; with (h), which can appear at any point in a dialog, the user has to reinitiate the whole dialog phase.
- If U2 has been understood, the system can choose between (c), (d), (e) and (f) according to its goals. (c) and (d) are confirmation questions supposing one more exchange before giving the answer. (e) and (f) are true answers supposing either that understanding has been achieved or that some adjustment rejoinders will follow.

(a) Sorry, could you repeat?
(b) Where do you want a garage?
(c) Why do you want a hotel?
(d) Do you want to get your car repaired in a city with a hotel?
(e) Would you prefer a hotel close to a garage?
(f) You have the 2 following hotels...
(g) You have 3 garages and 2 hotels. The following one is close to a garage...
(h) I don’t understand anything: would you ask the whole question again?

Example 2 : list of possible answers to follow U2 of Example 1.

2. Model function

The model has been designed based on both from descriptive studies of human machine dialogs recorded with a simulated machine and from the Geneva dialog model (Roulet & al. 85). The main point is to allow the system to always have a representation of the state of interaction. Such consciousness is necessary to give the system the possibility of knowing when an interruption risk might occur, i.e. when dialog can fail, or when it is not too dangerous if the system continues to make mistakes.

As we can see in Figure 1, the model is based on 2 axes. The horizontal one is called the managing axis, and the vertical one the incident axis:
- The managing axis deals with the real questions the system is asked, and the corresponding answers (managing question = MQ; managing answer = MA).
- The incident axis deals with the precision, reformulation, explanation or confirmation questions, which are necessary to understand MQs and MAs, and the corresponding answers (incident question = IQ; incident answer = IA).

There are 2 kinds of incident axes:
- The system incident axis, which appears after an MQ and which depends on the system. The corresponding answers are thus given by the user.
- The user incident axis, which happens after an MA, and is asked by the user. The corresponding answers are thus given by the system (1).

For each dialog, the system builds a graph in which it moves vertically when MQ or MA occur, and horizontally when IQ or IA occur. There is also another kind of movement on the incident axis: for each IQ, the system moves to the left; it moves to the right for each IA (2). To come back from the incident axis to the managing one is only possible if the initial incident level has been joined again. The one exception to this is if the system is obliged to a break back: to avoid dialog failure, the system comes back to the beginning of the incident axis, despite the spontaneity of interaction.

The whole dialog can thus be managed using the 4 following variables:
- A depth variable (DP), pointing out to what depth the different incident axis has driven dialog.
- An advancement variable (AV), pointing out how much exchanges have taken place on the managing axis.
- An incidence variable (IV), pointing out the length of each incident axis.
- A removal variable (RV), pointing out, during each incident axis, the interval between the state of the dialog and the return level.

Figure 1 : graph of the dialog model
(QP=MQ; QI=IQ; RP=MA; RI=IA)
- DV and AV are general variables, remaining active during the whole dialog. IV and RV are local variables, limited to one incident axis. The use of these variables gives various indications to the system:
- DV + AV indicates the number of turn-takings, and therefore the number of exchanges.
- HV/DV measure the height of incidence to give an idea of the foreseeable tolerance of the user to new IQ.
- IV and RV help manage each incident axis depending on general criteria (i.e. criteria belonging to the system) as supposed to local criteria (i.e. criteria depending on the state of the graph).

In this paper, we do not consider how the system will interpret that an utterance of the user is an IQ instead of an MQ, because it depends both on the parsing technique and the task model. We have used this for a system dealing with train timetable information (Luzzati 89), a frame-driven approach (Bobrow & al. 77) in which these distinctions depend on frame instantiation.

3. Use of the model

The main utility of the model is to allow the system to manage incidence. If the comprehension level of the system is high and the dialog is easy, the system will liberate itself from incidence and spontaneity will improve. If the comprehension level is limited and the dialog difficult to manage, the system will repair any incidence, or even make break backs, and rigidity will improve.

To repair incidence is to repair RV, generally at the first level. In Example 1, for example, repairing incidence means repeating M1 (i.e. to answer (a) in Example 2) instead of confessing that U2 hasn’t been perfectly understood (i.e. (b), (c) or (h) in Example 2). In Example 3, M3 and M7 are concerned: in the case of fixed incidence, the system would ask the last question again, instead of asking for a repetition. In the system incident axis, as in this example, the machine can immediately repair incidence, while this is only possible one degree later in a user incident axis. These cases are thus more difficult to manage; a system with a limited comprehension level should avoid as much as possible giving incomplete MAS, which would lead to these axes.

To liberate incidence means to let the system admit to any internal representation problem that occurs, as we can see in Example 3. The spontaneity of dialog increases, as well as the risk of incommunicability. At this point, the dynamic aspect of the dialog model helps modulate incidence according to the state of interaction, or repair it when an incommunicability risk occurs. In Example 3, the first reformulation request (M3) can be considered tolerable, because it appears at the beginning of an incident axis. However, the second one (M7) is not tolerable, because IV has intolerably increased.

The dynamic dialog model also allows anticipation of such risks. Increasing RV supposes a precise number of exchanges before being able to continue along the managing axis. It also allows for use of various thresholds according to the capacities of the system and to the nature of the task. For example, the system can define different areas in the graph in which the system can either liberate, modulate or repair incidence.

Example 3: dialog including a complex incident axis.
Finally, this dynamic dialog model permits management of an automatic dialog. First, it freezes interaction, so as to make sure that dialog will continue, despite the lack of spontaneity. Second, it improves system understanding qualities: incidence can be selectively liberated, for instance when particular capacities of the system occur. Finally, it anticipates the incommunicability risks, making a break back, restarting a phase of dialog or a whole communication, or even activating a new parser or calling a human operator.

(1) This description supposes one of the most common human-machine dialog tasks, i.e. tasks where the users question the machine, and not the contrary.

(2) It could be useful to build a 3-dimensional graph, instead of drawing left and right movements.

References