Title
Eliciting Additional Information during Cooperative Consultations

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

Authors
University, Monash
Raskutti, Bhavani
Zukerman, Ingrid

Publication Date
1993

Peer reviewed

Eliciting Additional Information during Cooperative Consultations*

Bhavani Raskutti† and Ingrid Zukerman
Department of Computer Science
Monash University
Clayton, Victoria 3168
AUSTRALIA
email: {banu,ingrid}@bruce.cs.monash.edu.au

Abstract

Analysis of naturally occurring information-seeking dialogues indicates that information providers often query a user when there is insufficient information to formulate a plan that satisfies the user's intentions. In this paper, we present a mechanism that determines when queries are required to elicit additional information from a user and the manner in which these queries should be posed. Query generation is done by taking into account the amount of relevant information in the user's intentions as recognized by a plan recognition mechanism. The mechanism for query generation described in this paper has been implemented as a component of a computerized information providing system in the travel domain.

Introduction

Consider the following excerpt from a real-life conversation at a Melbourne travel agency (`TA` stands for the travel agent, and `USER` for the customer/user).

```
1.  USER: I need to get some information
           about flights to Bombay.
2.  TA:    There are no direct flights to
           Bombay from Melbourne.  The best
           fare I can get is via Colombo.
           Is that OK or ...?
3.  USER: No.  That is not what I want.
           What is the best you can do
           via Singapore?
4.  TA:    When are you planning to go?
```

The above excerpt illustrates a salient feature of information-seeking interactions, viz that human information providers often query a user to obtain information. Queries are generated by an information provider to (1) *disambiguate between multiple possibilities*, e.g., the query to determine which route the customer has in mind (line 2 in the above excerpt); and (2) *obtain additional specifications*, e.g., the query to determine when the traveller is planning the trip in order to provide information about the fares (line 4 in the excerpt). A disambiguating query is generated when an information provider[1] infers many possible goals on the basis of the user's request. A query to obtain additional specifications is issued when the information provider infers a unique goal underlying the user's request, but is unable to generate a plan that achieves this goal.

The importance of recognizing plans/goals for responding appropriately during an information-seeking interaction has been widely accepted in the artificial intelligence community. The incorporation of plan recognition capabilities into computerized information providers has enabled a range of cooperative behaviours, such as supplying more information than what is explicitly requested [Allen & Perrault 1980], responding to ill-formed queries [Carberry 1988], and understanding (a) indirect speech acts [Perrault & Allen 1980], (b) inter-sentential ellipsis [Carberry 1985, Litman 1986], (c) queries based on invalid plans [Pollack 1990], and (d) sub-dialogues initiated in order to debug or correct plans [Sidner 1985, Litman & Allen 1987]. More recently, there has been research to develop plan recognition systems that are capable of considering multiple alternatives in order to recognize intentions that are developed and revised over multiple utterances [Kautz & Allen 1986, Carberry 1990, Raskutti & Zukerman 1991, Goldman & Charniak 1991, Appelt & Pollack 1992].

However, the plan recognition systems described above are not concerned with query generation to promote the cooperative process. When there is doubt regarding a speaker's intentions, these systems use heuristics to decide which of their inferences is the one intended by the speaker. If the heuristics do not result

*This research was supported in part by grant Y90/03/22 from the Australian Telecommunications and Electronics Research Board.

†Currently employed at Telecom Research Laboratories, Australia.

[1]The terms *information provider*, *system* and *listener* are used interchangeably in this paper. Similarly, the terms *information seeker*, *user* and *speaker* have the same meaning.

in a single alternative that can be planned for, then the speaker is expected to provide information without being requested to do so. This, however, is an unrealistic expectation since the user does not have access to the system's inferences.

Query generation for the purpose of disambiguating between multiple possible interpretations of a user's request has been considered in [van Beek & Cohen 1991] and [Wu 1991]. However, these systems do not consider a situation where a single goal of a user is recognized, but this goal does not have sufficient information to enable the formulation of a successful plan. Thus, these systems do not elicit information when a recognized goal lacks detail. Query generation for eliciting information is performed in GUS, a frame-driven natural language understander operating in the travel domain [Bobrow et al. 1977]. However, GUS's domain of discourse is restricted to deal only with goals involving return flights from Palo Alto, and hence, it does not need a goal inference process that underlies the query generation mechanism.

In this paper, we present two approaches for generating queries that elicit information. The first approach aims to keep the interaction succinct, while the second attempts to generate queries in a hierarchical and sequential manner. Both approaches consider the amount of relevant information in an inferred goal in order to determine whether a query should be posed about the goal.

Query Generation: Input

The query generation mechanism described in this paper is implemented as part of the response generator module of RADAR, a computerized information provider that **R**ecognizes **A**nd **D**iscriminates between **A**lternatives and **R**esponds. RADAR consists of four modules: (1) a parser that converts the natural language input of a user to predicates, (2) a plan recognizer that infers possible intentions of the user on the basis of the predicates passed by the parser, (3) a plan generator that proposes one or more means to satisfy an intention recognized by the plan recognizer, and (4) a response generator that generates responses and queries.

The response generation is done on the basis of the alternatives inferred during plan recognition and their probabilities. In RADAR, each of these alternatives is called an *interpretation*. Each interpretation consists of a sequence of action instances, where each action instance specifies a domain action with a number of instantiated parameters. For instance, the interpretation that John intends to fly to Sydney on the 1st of January 1992 consists of one action instance that specifies the action of FLYing. The parameters *destination*, *departure-date* and *departure-time* of this action are instantiated.

The input to the response generator may consist of **zero**, **one** or **many** interpretations. If there are zero interpretations, then the user is asked to re-phrase the request. If there are many interpretations, then the response generator issues a disambiguating query in order to determine the interpretation intended by the user [Raskutti & Zukerman 1992]. If there is one interpretation, then the course of action chosen by the information provider depends on the sufficiency of the information contained in this interpretation. If the information is sufficient to formulate a plan and a plan is formulated by the plan generator, then the user is informed of the system's inferences and of the proposed plan [Raskutti 1992]. If a plan is not formulated, then the system engages in a negotiation process whereby the user's request is altered until a valid plan is formulated [Raskutti 1992]. Finally, if the information is not sufficient to formulate a plan, RADAR seeks additional information from the user. This is done as described in the subsequent sections.

Determining the Sufficiency of Information

In order to obtain additional information for completing an interpretation with insufficient information, the response generator takes into account both the amount of relevant information in the interpretation and the amount of information that is sufficient to support the generation of a plan. The amount of relevant information in an interpretation is determined by the information content of the interpretation. The amount of information that is considered sufficient to support the generation of a plan is determined by a threshold at which an interpretation is accepted as complete. These two measures are detailed in the following sections.

Information Content of an Interpretation

The information content of an interpretation is defined as the sum of the information content of each action instance in the interpretation. The information content of an action instance in turn is determined by the sum of the information content of the parameters in a set of *necessary parameters*. These are parameters that are essential for the definition of the action instance.

The determination of a set of necessary parameters is complicated by the fact that an action instance may have several sets of necessary parameters. For instance, {origin, destination, departure-date, departure-time} is one such set for the action instance TAKE-PLANE, and {origin, destination, arrival-date, arrival-time} is another set. Inferences from the instantiations of the parameters in one set of necessary parameters automatically define the instantiations of the parameters in the other sets. Hence, in order to compute the information content of an action instance, the system must first determine which set of necessary parameters should be considered. This is done by choosing the set that yields the highest information content for the action instance.

The information content of a parameter in an action instance depends on three factors: N – the number of possible values assigned to this parameter; S – the strength of the belief in the parameter; and Sig – the significance of the parameter. The strength of the belief in a parameter is directly proportional to the reliability of the source of information from which this parameter was obtained. For instance, the strength of a parameter inferred directly from a user's statements is 1, while a parameter inferred on the basis of common sense notions has a strength of 0.6. The significance of a parameter reflects the importance of this parameter's definition for the overall definition of an action instance, e.g., in the travel domain, $Sig(origin) = Sig(destination)$, and $Sig(departure/arrival\ time) < Sig(departure/arrival\ date) < Sig(origin)$.

Borrowing from Information Theory [Shannon 1948], the information content of a parameter p, $IC(p)$, is defined as follows:

$$IC(p) = \begin{cases} \log_2 \frac{S(p)}{N(p) \times Sig(p)} & \text{if } N(p) \neq 1 \\ \log_2 S(p) & \text{otherwise} \end{cases}$$

For example, if the departure date of a trip is directly inferred from a user's statements to be between the 9th and the 15th of May, 1993, then the information content of this parameter is $\log_2 \frac{1\ \{=S(p)\}}{7\ \{=N(p)\}\ \times\ 3\ \{=Sig(p)\}} = -4.39232$. This measure ranges over the non-positive values.

According to this formula, undefined parameters have the least information content, since they can take on all possible values in the domain. Further, inexactly defined parameters of high significance, e.g., origin, have a lower information content than inexactly defined parameters of low significance. This is based on the observation that ignorance about a significant parameter contributes more to the lack of information of an interpretation than ignorance about an insignificant parameter. Since there is no lack of knowledge when a parameter is exactly defined, i.e., $N(p) = 1$, the parameter's significance is not taken into account when computing its information content. Hence, in this case, $IC(p) = \log_2 S(p)$. Thus, a parameter inferred exactly from a reliable source, i.e., $S(p) = 1$, has a maximum information content of 0. The maximum value of the information content of an interpretation is also 0, and it is achieved when all the parameters in a necessary set of parameters of its action instances have a maximum information content.

Completeness of an Interpretation

An interpretation is deemed complete if its information content is greater than a particular threshold. This threshold is referred to as $IC_{complete}$, and it ranges over the non-positive values. The value chosen for $IC_{complete}$ affects the clarification interaction between the user and the system prior to the invocation of the plan generator. Hence, by judicious choice of a value for $IC_{complete}$ the system is capable of modelling different attitudes of an information provider. A high value for $IC_{complete}$, achieved by assigning to it a small negative value, results in the generation of a large number of queries to complete an interpretation, thus modelling an information provider who does not perform plan formulation until the inferred goal is well defined. On the other hand, a low value for $IC_{complete}$, achieved by assigning to it a large negative value, represents an information provider who will attempt plan formulation even when the goals of the user are not clearly defined.

```
Get me a ticket to Sydney.
I am going to Hawaii on the 11 am flight
     the day after tomorrow.
By the way, I'll be flying from Adelaide.
```

Figure 1: **Initial Request of a Traveller**

In RADAR, the value of $IC_{complete}$ is chosen so that an interpretation is accepted as complete for plan formulation when all the necessary parameters of high significance are specified exactly, and some of the necessary parameters of low significance are partially specified. Thus, in general, RADAR will accept those interpretations where the origin, destination and departure/arrival dates of the action instances are known exactly. Since the information content of an interpretation is the sum of the information content of the necessary parameters of its domain actions, in principle, an interpretation may be accepted as complete when a single significant parameter is not known exactly and all the other parameters are known exactly. However, due to the use of $Sig(p)$ in the computation of the information content, this situation is not likely to occur.

To illustrate an inferred goal with insufficient information, consider the interpretation in Figure 2 generated by the plan recognizer on the basis of the request in Figure 1. This interpretation is inferred when the request is issued at a Melbourne travel agency with today's date set to the 14th of March 1991. In the interpretation in Figure 2, the departure dates and times of three of the trips, viz *Melbourne–Adelaide*, *Adelaide–Sydney* and *Hawaii–Melbourne*, are not specified exactly[2]. Due to the lack of these specifications, the information content of this interpretation is less than $IC_{complete}$. While it is possible to infer some of these specifications, such as the departure times of the trips, by looking up the flight time tables, other specifications, such as the departure dates of the three trips, cannot be determined by such means. Hence, additional information must be acquired from the user in order to complete the interpretation in Figure 2.

[2]The *Melbourne–Adelaide* and *Hawaii–Melbourne* trips are postulated by RADAR based on the premise that people generally depart from and return to their place of residence.

location of travel agency = **Melbourne** today's date = **14th of March 1991**
Fly from *Melbourne* to *Adelaide* departing on date $\leq$ the 16*th* of *March* 1991 at time $\leq$ 8:00 *am* arriving on date $\leq$ the 16*th* of *March* 1991 at time $\leq$ 9:30 *am*
Fly from *Adelaide* to *Sydney* departing on date $\leq$ the 16*th* of *March* 1991 at time $\leq$ 9:30 *am* arriving on date $\leq$ the 16*th* of *March* 1991 at time $\leq$ 11:00 *am*
Fly from *Sydney* to *Hawaii* departing on date = the 16*th* of *March* 1991 at time = 11:00 *am* arriving on date = the 16*th* of *March* 1991 at time = 11:00 *pm*
Fly from *Hawaii* to *Melbourne* departing on date $\geq$ the 16*th* of *March* 1991 at time $\geq$ 11:00 *pm* arriving on date $\geq$ the 17*th* of *March* 1991 at time $\geq$ 11:00 *am*

Figure 2: **RADAR's Interpretation**

Seeking Additional Information

The process of eliciting information by means of queries is directed at increasing the information content of an interpretation so that it is greater than $IC_{complete}$. The information content of an interpretation may be increased by generating (a) *confirmation queries* to increase the strength of some inferences that led to the interpretation, and/or (b) *information-seeking queries* to obtain specifications for parameters that are unspecified or inexactly specified. The number of query/answer turns needed to complete an interpretation as well as the style of the interaction are influenced by the approach used to select parameters for querying. In RADAR, the initial attempt at query generation was directed at keeping the interaction **succinct**. However, the interactions resulting from the succinct approach are not suitable for the travel domain. Therefore, an alternative approach called the **hierarchical/sequential** approach was attempted. This approach yielded interactions that mirrored real interactions in the travel domain. The implementation of the two approaches to query generation and their advantages and disadvantages are discussed in the following sections.

In both approaches, once a parameter is chosen for querying, the response generator poses a query that requests the user to specify this parameter. The information supplied in response to the query is then used to perform additional inferences. The process of parameter selection, query generation and inference on the basis of the supplied information is repeated until the interpretation preferred by the plan recognizer has an information content that is greater than $IC_{complete}$. This interpretation is then sent to the plan generator.

The Succinct Approach

In order to keep the interaction dialogue succinct, the response generator must pose a query whose response increases the information content of an interpretation to the greatest degree, i.e., the parameter with the highest potential contribution to the information content of the interpretation is selected for querying. The selection of this parameter is achieved as follows. First, the action instance with the least information content is chosen. Next, the set of necessary parameters that yields the highest information content for the chosen action instance is selected as the set that must be specified. From this set of parameters, the parameter with the least information content is then selected for querying. The first parameter that is selected by this approach for completing the interpretation in Figure 2 is the departure date of the last action instance, i.e., the *Hawaii–Melbourne* trip. Figure 3 illustrates the interaction resulting from this approach for the request in Figure 1.

The Hierarchical/Sequential Approach

Analysis of consultation transcripts in the medical domain [Shortliffe 1975] and the travel domain [Bobrow et al. 1977] indicates that consultants in these domains proceed both hierarchically and sequentially, i.e., unspecified/under-specified parameters with a higher significance are sought prior to those with a lower significance, and the action instances are queried in the order in which the information provider expects them to be performed.

In order to query hierarchically, RADAR chooses the action instance that lacks the definition of a necessary parameter with the highest significance. If several action instances have unspecified parameters with the same significance, then RADAR achieves sequential behaviour by choosing the action instance that is to be performed first.

We have also observed that in the travel domain, if a user omits both departure and arrival information from his/her initial request, an information provider will usually ask questions regarding departure information instead of arrival information. This is because if the arrival date and time of a trip had been of importance to the traveller, e.g., s/he needs to arrive in time for a conference, then s/he would have presented this information at the outset as constraints that must be satisfied. However, if the traveller omits information regarding the dates and times of a trip, s/he is more likely to know the departure information of the trip than its arrival information. Thus, in general, there is a preferred set of necessary parameters that is used for

```
 2 RADAR: What day are you departing
          from Hawaii to Melbourne?
 3 USER: The 30th.
 4 RADAR: What time are you departing
          from Adelaide to Sydney?
 5 USER: 8:00 am.
 6 RADAR: What time are you arriving
          in Adelaide from Melbourne?
 7 USER: 9:30 am.
 8 RADAR: What day are you arriving
          in Adelaide from Melbourne?
 9 USER: Tomorrow.
10 RADAR: What day are you departing
          from Adelaide to Sydney?
11 USER: The day after tomorrow.
12 RADAR: Your travel itinerary is ...
```

Figure 3: **Succinct Dialogue**

querying, and this set contains parameters that a user is likely to be able to specify. In addition, it has been observed that travel agents often confirm the essential aspects of a trip prior to proceeding with its details.

The first of these observations is incorporated into the hierarchical/sequential approach by assigning to each type of action instance a designated set of necessary parameters that the response generator looks up when querying. In the travel domain, for primitive operators that perform transportation, the designated set is the one that contains origin, destination, and departure date and time.

The second observation is incorporated as follows: if a parameter selected for querying has a lower significance than the most significant parameters in the designated set, and the strength of at least some of the most significant parameters is less than 1, then a confirmation query is issued to assert the overall intent of the user. If the user answers this query in the affirmative, then the strength of all the parameters affected by this query is updated to that of a direct inference.

This approach selects for querying the departure date of the first action instance in the interpretation in Figure 2, i.e., the *Melbourne–Adelaide* trip. However, prior to this query, it generates a confirmation query (Figure 4).

Choosing an Approach

In order to determine the query generation approach that must be used by an information provider during an interaction, the two approaches were tried on 8 different types of requests with the following different values of $IC_{complete}$: 0 – which accepts only those interpretations where all necessary parameters are exactly defined by means of direct inferences; -8 – which accepts interpretations where some of the least significant parameters, such as departure/arrival times, are not exactly specified; and -17 – which accepts interpretations where all of the least significant parameters are inexactly specified. The value of 0 for $IC_{complete}$ was unacceptable, since the information provider asked for information at the slightest doubt, and hence information-seeking and confirmation queries were generated until all the necessary parameters of all the action instances were exactly specified by means of direct inferences. The values of -8 and -17 were found to be useful at different stages of the information-seeking process. If the user is making a preliminary exploration, then the lower value is preferable, while during the actual plan formulation stage, the higher value is preferable.

```
 2 RADAR: Let me get this right,
          You are going from Melbourne
          to Adelaide, then Sydney then
          Hawaii and then Melbourne?
 3 USER: Yeah.
 4 RADAR: What day are you departing
          from Melbourne to Adelaide?
 5 USER: Tomorrow.
 6 RADAR: What day are you departing
          from Adelaide to Sydney?
 7 USER: The day after tomorrow.
 8 RADAR: What day are you departing
          from Hawaii to Melbourne?
 9 USER: The 30th.
10 RADAR: What time are you departing
          from Melbourne to Adelaide?
11 USER: 8:00 am.
12 RADAR: What time are you departing
          from Adelaide to Sydney?
13 USER: 8:00 am.
14 RADAR: Your travel itinerary is ...
```

Figure 4: **Hierarchical/Sequential Dialogue**

To illustrate the effect of the query generation approach and $IC_{complete}$ on the resulting interaction, consider the interactions obtained with the two approaches when the user makes the initial request presented in Figure 1 (Figures 3 and 4[3]). These interactions are generated with $IC_{complete}$ set at -8. When RADAR processes the same initial request using the lower value of -17, then the user is asked only the first two questions when the succinct approach is used, while the hierarchical/sequential approach requires the first three information-seeking questions in Figure 4, in addition to the confirmation query. Hence, the approach aimed at succinctness increases the information content of an interpretation with a few queries, particularly when interpretations with a low information content are accepted as complete. However, the queries generated using this approach do not conform to the expectation of hierarchy and sequence in the travel do-

[3]Line 1 in both dialogues is the user's initial request.

main. This lack of conformity becomes more apparent when the value of $IC_{complete}$ is high, i.e., close to 0, and a large number of queries is posed.

These observations indicate that the approach aimed at succinctness is useful in the following situations: (1) in domains where a single action instance is considered, e.g., retail consultation where the user intends to buy a single consumer item, and domains where the information is not organized hierarchically, e.g., car repair assistance; and (2) during preliminary consultations in any domain, where $IC_{complete}$ is set to a low value.

Conclusion

In this paper we have presented two approaches for generating queries that obtain additional information necessary for plan formulation. Both approaches take into account the amount of relevant information in an interpretation, and thus generate a query only when it is necessary. The query generation mechanism is integrated into RADAR, a computerized information providing system in the travel domain. RADAR has been used to process a number of requests of different types with different values of $IC_{complete}$. The queries generated by RADAR with the hierarchical/sequential approach were consistent with those generated by human information providers in the travel domain.

References

Allen, J.F. and Perrault, C.R. (1980), Analyzing Intention in Utterances. In *Artificial Intelligence* 15, pp. 143–178.

Appelt, D.E. and Pollack, M.E. (1992), Weighted Abduction for Plan Ascription. In *User Modeling and User Adapted Interaction* 2(1–2), pp. 1–25.

Bobrow, D. G., Kaplan, R. M., Kay, M., Norman, D. A., Thompson, H. and Winograd, T. (1977): GUS-1, a Frame-Driven Dialog System. In *Artificial Intelligence* 8(2), pp. 155–173.

Carberry, S. (1985), A Pragmatics-based Approach to Understanding Intersentential Ellipsis. In *Proceedings of the 23rd Annual Meeting of the Association of Computational Linguistics*, pp. 188–197, Chicago, Illinois.

Carberry, S. (1988), Modeling the User's Plans and Goals. In *Computational Linguistics* 14(3), pp. 23–37.

Carberry, S. (1990), Incorporating Default Inferences into Plan Recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pp. 471–478, Boston, Massachusetts.

Goldman, R. and Charniak, E. (1991), A Probabilistic Model for Plan Recognition. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pp. 160–165, Anaheim, California.

Kautz, H. and Allen, J. (1986), Generalized Plan Recognition. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 32–37, Philadelphia, Pennsylvania.

Litman, D. (1986), Understanding Plan Ellipsis. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 619–624, Philadelphia, Pennsylvania.

Litman, D. and Allen, J.F. (1987), A Plan Recognition Model for Subdialogues in Conversation. In *Cognitive Science* 11, pp. 163–200.

Perrault, C.R. and Allen, J.F. (1980), A Plan-Based Analysis of Indirect Speech Acts. In *American Journal of Computational Linguistics* 6, pp. 167–182.

Pollack, M.E. (1990), Plans as Complex Mental Attitudes. In Cohen, P., Morgan, J. and Pollack, M.E. (Eds.), *Intentions in Communication*, MIT Press.

Raskutti, B. (1992), A Probabilistic Approach to Plan Recognition and Response Generation during Cooperative Consultations. Ph.D. Thesis, Department of Computer Science, Monash University (submitted).

Raskutti, B. and Zukerman, I. (1991), Generation and Selection of Likely Interpretations during Plan Recognition in Task-Oriented Consultations Systems. In *User Modeling and User Adapted Interaction* 1(2), pp. 323–353. A modified version is available as Technical Report 91/157, Department of Computer Science, Monash University.

Raskutti, B. and Zukerman, I. (1992), Generating Queries during Cooperative Consultations. Technical Report 92/172, Department of Computer Science, Monash University.

Shannon, C.E. (1948), A Mathematical Theory of Communications. In *Bell System Technical Journals*, October 1948.

Shortliffe, E.H. (1975), *Computer-Based Medical Consultations: MYCIN*, American Elsevier Publishing Company, New York.

Sidner, C.L. (1985), Plan Parsing for Intended Response Recognition in Discourse. In *Computational Intelligence* 1(1), pp. 1–10.

van Beek, P. and Cohen, R. (1991), Resolving Plan Ambiguity for Cooperative Response Generation. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pp. 938–944, Sydney, Australia.

Wu, D. (1991), Implications of Active User Model Acquisition for Decision-Theoretic Dialog Planning and Plan Recognition. In *User Modeling and User Adapted Interaction* 1(2), pp. 149–172.