Reasoning about Updates

Marie Nilsenová

ILLC-UvA,
Nw. Doelenstraat 15,
1012 CP Amsterdam

Abstract

The aims of this paper are (i) to show advantages of using belief networks for natural language interpretation, and (ii) to argue that two layers of reasoning - *in* **and** *with* **the network - are needed in order to evaluate different ways of updating the network. The argument is empirically based on the phenomenon of model conversation failures which occur when a speaker is perceived to be inconsistent by the interpreter.**

1 Introduction

In the process of interpretation, human agents make use of their world knowledge together with the beliefs they hold of other agents. In particular, they evaluate the reliability of the information they receive (a) with respect to their previous beliefs about the topic, and (b) with respect to the reliability of **the source**: *Is (s)he well-informed? - Can I trust her/him?* Information update is influenced by the nature of the answer to the latter two questions. Therefore, from the point of view of interpretation modelling, it appears desirable to represent the impact of the degree of credibility and well-informedness ascribed by the interpreter to other agents. As argued by [7], it is in fact necessary to do so when representing the effect of *model conversation failures* on the interpretation process. Model failures (as opposed to *input failures*, viz [5]) occur when the speaker is perceived to be inconsistent by the interpreter (which can be either a human agent or a natural language dialogue system), either by contradicting herself, or by correcting or retracting previously asserted proposition. In previous approaches to information update (e.g., [1] - S. Larsson, p.c., [10]), corrections and retractions either result in bringing on an absurd state, or a "blind" revision (replacing the old information with the new update, with no further consequences). This is problematic not only for theoretical reasons, but also for practical purposes, as the interpretive module of a dialogue system should be able to distinguish between negotiable and non-negotiable information.[1] Instead, [7] propose to represent the belief state of an interpreter as a Bayesian (belief) network ([4]), in which interpreter's previous beliefs (about the world, about sources of information) influence the interpretation process.

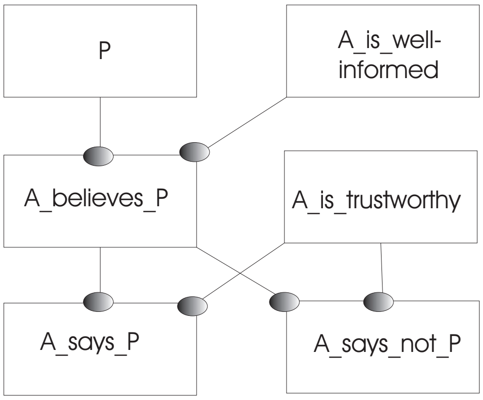

Figure 1.: Example of a belief network for interpretation

A Bayesian network is a directed acyclic graph with probability assignments associated with its nodes. The nodes (propositional variables) stand for beliefs of the agent and arrows in the graph are causal connections between beliefs. Information update is modelled by probability conditionalization locally propagating through the network.[2] Unlike in the previous approaches cited earlier, the process of information update is thus represented *quantitatively* and relations among beliefs directly influence the result of the update.

Although the approach of [7] leads to a simple and quite realistic representation of several cases of conversation failures, it cannot deal with the fact that agents appear to **reason** about ways of updating their beliefs. Especially if repeatedly communicating with the same agent, the interpreter can decide to disregard inconsistencies and still consider the agent trustworthy (the conversation failure can be ascribed to misperception instead). On the other hand, if the interpreter believes with a high probability in a proposition p and the speaker asserts $\neg p$, rather than revising her beliefs with respect to the proposition in question, the interpreter can instead decide that the speaker is not well-informed. (We know from experience that we often refuse to believe surprising news at first. The chance that we accept the news to be genuine increases if we acquire it iteratively from additional sources - a process which can, in fact, be captured in the belief network in [7].) Finally, reasoning about different kinds of update comes handy in cases of semantic ambiguities, when the interpreter decides for

[1] E.g., the information about user's travel destination is negotiable, while the price of the airplane ticket to the destination is not. See [2] for discussion.

[2] For an example of a network for one proposition and one agent, see Figure 1.

an update that conflicts the least with her previous beliefs (yet still has some information value). For example, if in a restaurant we see the sign "Please wait for the waiter to be seated", we would probably not look forward expectantly to the waiter taking his seat (and fastening his seat belt) but rather leading us to a free table.

In the present article, we will build on the previous work of [7]. In order to improve their model, we will explore three possible ways of representing reasoning about updates in Bayesian networks. Using the decision-theoretic terminology of, a.o., [9], we make the following general assumptions:

- agents are *utility maximizers* - they attempt to obtain the most relevant information with minimum effort;
- *utility* in the interpretation process is correlated with information value;
- *effort* increases if the interpreter has to revise strongly held beliefs.

2 The "Useful" Node

First, we attempt at a naive solution, consisting simply in adding a "useful" node into the original network. The table below shows (probabilistic) values associated with the node "Useful" given all possible combinations of the values of its parents. E.g., if the interpreter believes that q is true, p is true and the speaker A is credible, the positive value of the node will be 0.9. In general, the belief that a certain update is useful is high (0.9) when an agent is considered well-informed (the value TRUE), but goes down (0.6) if the agent is not considered well-informed, (the value FALSE). The network simulates situations where information received from the agent A is in contradiction with several of the interpreter's previous beliefs.

q	p	**A cred.**	**TRUE**	**FALSE**
TRUE	TRUE	TRUE	0.9	0.1
TRUE	TRUE	FALSE	0.4	0.6
TRUE	FALSE	TRUE	0.9	0.1
TRUE	FALSE	FALSE	0.4	0.6
FALSE	TRUE	TRUE	0.9	0.1
FALSE	TRUE	FALSE	0.4	0.6
FALSE	FALSE	TRUE	0.9	0.1
FALSE	FALSE	FALSE	0.4	0.6

The major objection to this approach is that no reasoning about an update is actually taking place. In cases where results of two kinds of updates have to be compared to each other (as for ambiguities), there is still no available mechanism to do so (e.g., one that would compare two different values of the node *Useful*). In other words, the network does not reflect the strife between *utility* and *effort* which we want to represent.

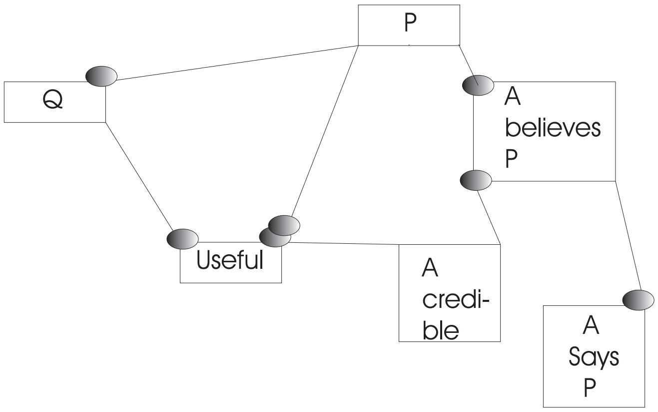

Figure 2.: The naïve approach.

As we will see in the next section, the difficulty with the second possible representation, via **decision networks**, will turn out to be on a par with the one just raised. We will explore the option nevertheless, because decision networks have been proposed as a general mechanism for rational decisions in the belief network literature ([6], [3]).

3 Decision Networks

"In its most general form, a decision network represents information about the agent's current state, its possible actions, the state that will result from the agent's action, and the utility of that state" ([6], p.484). Decision networks (sometimes also called *influence diagrams*) are belief networks enriched with nodes for decisions (rectangles) and utilities (diamonds). The usual chance nodes are represented as ovals.

Decisions about future actions are based on a utility evaluation of each possible setting of the decision node. In other words, for each possible action, the decision node is first set to that action and subsequent posterior probabilities of the oval nodes and the resulting value of the utility node are calculated. The selected action is the action with the highest expected utility. Obviously, a level of reasoning above the decision network is thus needed in order to compare the utility values of different states of the network.

In Figure 3., $< \alpha >$ and $< \beta >$ are possible actions the interpreter can take. By taking the action $< \alpha >$, she would decrease the probability that the speaker is credible and retain her previous beliefs concerning the actual world. By action $< \beta >$, she would maintain the belief that the speaker is credible and rather revise her beliefs about the world, based on the information obtained from the speaker.However, the network cannot update in such a manner that would maximize information value of the update, while minimizing its effort.

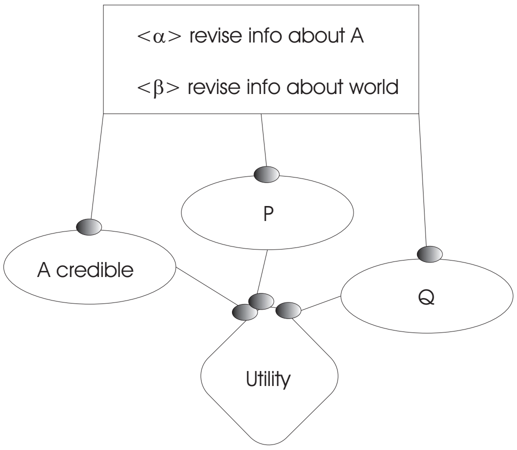

Figure 3.: Decision Network

For example, if the interpreter strongly believes that p and q are true but the information she receives from the speaker A is to the contrary, in the very limited model considered here it would be more advantageous for her to perform action $< \alpha >$ (by the reasoning that it requires less effort to revise beliefs in one proposition than in two), especially if her belief in speaker A's credibility is quite low to begin with. On the other hand, the interpreter is after information maximization, so she also does not disregard the information value resulting from the update. It is quite obvious that the *Utility* node cannot reflect the expected utility of the interpretation act, because it is itself a part of the network from which the utility values should be derived (by comparison of the effects of different types of updates). In consideration of this fact, we will propose that the expected utility be calculated in a process superimposed over the original belief network.

4 A Refined Architecture

The idea of representing the interpretation process on two levels appeared in the work of [8] on belief networks for purposes of medical diagnosis. There, a *control layer* containing non-probabilistic knowledge is used to control different types of problem solving, performed by the *probabilistic layer*. In our proposal, the control layer queries the probabilistic layer about (i) the probability change associated with the proposition under discussion, and (ii) the probability distribution, both prior and after an update. With the assumption that both positive and negative information is useful,[3] we propose to calculate the *utility* of the update as the difference in probability assigned to the proposition under discussion before an information update and after. Thus,

$$U = |P(X_n) - P'(X_n)|$$

where U is the utility of the update, $P(X_n)$ is the original probability of the node X_n and $P'(X_n)$ is the probability associated with the node after the update. On the other hand, there is effort associated with changes of probabilities at nodes influenced by the change of probability at X_n (non-d-separated nodes). Thus, for the part of the network of i nodes, locally affected by the update, we can calculate the effort of the update as

$$E = |P(X_1) - P'(X_1)| + ... + |P(X_{n-1}) - P'(X_{n-1})| + |P(X_{n+1}) - P'(X_{n+1})| + ... + |P(X_i) - P'(X_i)|$$

In order to compare the utility of two different updates (in the previous section referred to as actions $< \alpha >$ and $< \beta >$, respectively), the control layer calculates the utility and effort for (i) X_n = well-informedness (or credibility) of the speaker (A), i.e., for the case when the network would be updated with the proposition "The speaker is not credible", and (ii) X_n = an update with the proposition under discussion p. Of the nine possible outcomes, the first six (indicated in the table below) can be decided straightforwardly:

	Utility	**Effort**	**Decision**
1.	$U_i > U_{ii}$	$E_i = E_{ii}$	$< \alpha >$
2.	$U_i < U_{ii}$	$E_i = E_{ii}$	$< \beta >$
3.	$U_i = U_{ii}$	$E_i < E_{ii}$	$< \alpha >$
4.	$U_i = U_{ii}$	$E_i > E_{ii}$	$< \beta >$
5.	$U_i > U_{ii}$	$E_i < E_{ii}$	$< \alpha >$
6.	$U_i < U_{ii}$	$E_i > E_{ii}$	$< \beta >$
7.	$U_i = U_{ii}$	$E_i = E_{ii}$	arbitrary
8.	$U_i > U_{ii}$	$E_i > E_{ii}$	?
9.	$U_i < U_{ii}$	$E_i < E_{ii}$	?

In case of the seventh outcome, either action can be taken (though preference for $< \beta >$ could be assumed). To choose an action in case of the last two outcomes, we can foresee two possible strategies:

- give superiority to either *utility* or *effort*; [4]
- give preference to either action $< \alpha >$ or action $< \beta >$.

[3] For example, if the interpreter thinks with equal probability (0.5) that it is going to rain tomorrow, or that it is not going to rain, change in the probability in either direction is useful information, e.g., with respect to her decision problem *Should I take an umbrella?*

[4] This strategy could be relativized with respect to a constant α (e.g., for Outcome 8, if $E_i > \alpha$, then perform action $< \beta >$.)

It is not quite obvious which (if any) of the two options is in fact selected for by the interpreter in actual conversations.

5 Conclusion and Discussion

We have argued for a potential use of belief networks in one area of natural language interpretation. We indicated several difficulties regarding reasoning about updates in the approach of [7] and discussed three possible ways of modelling the process with belief networks. Of the three suggested options, a *Utility* node, decision networks and an additional *control layer*, the third appears to be most suitable for evaluating the utility versus effort of the update. We proposed to calculate the utility of an update as the absolute difference in probability value of the proposition under discussion before and after an update. As for effort, we have construed it as the sum of changes in probabilities at nodes affected by the update. By comparing the two values, the interpreter maximizes the general utility of the information update by either updating her beliefs about the source of information, or by updating the proposition under discussion. From a theoretical point of view, it is important that in our architecture, we have merged together the Gricean maxim of quality (here spelled-out as credibility and well-informedness) with the semantic module. This was crucial both for representing model conversation failures, but also for deciding about the general utility of different types of updates.

♢
The author would like to thank Balder ten Cate, Robert van Rooy and Henk Zeevat for helpful comments regarding the content of this article.
♢

References

[1] *TRINDI: Task Oriented Instructional Dialogue.* http://www.ling.gu.se/projekt/trindi//.

[2] Andreas Herzig and Dominique Longin. Belief dynamics in cooperative dialogues. *J. of Semantics*, 17(2), 2000. vol. published in 2001.

[3] R.E. Neapolitan. *Probabilistic Reasoning in Expert Systems*. John Wiley and Sons, New York, 1990.

[4] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, California, 1988.

[5] M.H. Ringle and B.C. Bruce. Conversation failure. In W.G. Lehnert and M.H. Ringle, editors, *Strategies for Natural Language Processing*, pages 203–221. 1982.

[6] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.

[7] B. ten Cate and M. Nilsenová. Representing model conversation failures with belief network. In B. Kroese, M. de Rijke, G. Schreiber, and M. van Someren, editors, *Proceedings of BNAIC 2001*, pages 365–372, Amsterdam, October 2001.

[8] L.C. van der Gaag. Bayesian belief networks: odds and ends. *The Computer Journal*, 39:97–113, 1996.

[9] R. van Rooy. Questioning to resolve decision problems. In P. Dekker, editor, *Proceedings of the Twelfth Amsterdam Colloquium*, pages 211–216, 1999.

[10] H. Zeevat and R. Scha. Integrating pragmatics into update semantics. In A. Ortony, J. Slack, and O. Stock, editors, *Communication from an AI perspective*, pages 17–33. Springer-Verlag, 1992.