An Information-theoretic Approach for Argument Interpretation

Sarah George and **Ingrid Zukerman**
School of Computer Science and Software Engineering
Monash University
Clayton, VICTORIA 3800, AUSTRALIA
email: {sarahg,ingrid}@csse.monash.edu.au

Abstract

We describe an information-theoretic argument-interpretation mechanism embedded in an interactive system. Our mechanism receives as input an argument entered through a web interface. It generates candidate interpretations in terms of its underlying knowledge representation – a Bayesian network, and applies the Minimum Message Length principle to select the best candidate. The results of our preliminary evaluations are encouraging, with the system generally producing plausible interpretations of users' arguments.

Keywords: Minimum message length, discourse interpretation, Bayesian networks.

1 Introduction

Discourse interpretation is an essential component of any dialogue system. However, most interactive systems developed to date afford users limited opportunities to express their views. The discourse interpretation mechanism described in this paper constitutes a step towards solving this problem.

This research builds on our previous work on BIAS – a *Bayesian Interactive Argumentation System* which uses Bayesian networks (BNs) (Pearl, 1988) as its knowledge representation and reasoning formalism. BIAS is designed to be a complete argumentation system which will eventually engage in unrestricted interactions with users. However, in this paper, we focus on its discourse interpretation mechanism and the impact of attentional focus on the interpretation process.

The contributions of this paper are as follows.

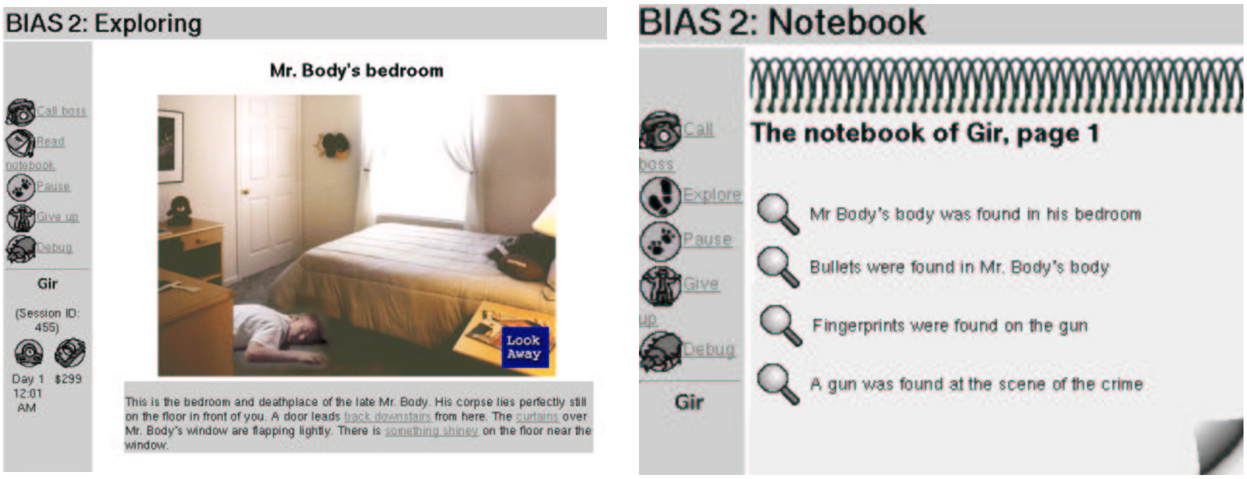
1. We incorporate attentional focus into the discourse-interpretation formalism described in (Zukerman and George, 2002), which uses the Minimum Message Length (MML) Principle (Wallace and Boulton, 1968) to evaluate candidate discourse interpretations.

2. We investigate a web-based argumentation facility for the detective game described in (Zukerman, 2001).

In the following section, we describe our detective game, and discuss our knowledge representation. Next, we outline the argument interpretation process. In Section 4, we provide an overview of our Minimum Message Length approach to discourse interpretation, and describe how attentional focus is incorporated into this formalism. The results of our evaluation are reported in Section 5. We then discuss related research, followed by concluding remarks.

2 Detective Game

As for the system described in (Zukerman, 2001), our experimental set up takes the form of a game where the user and the system are partners in solving a murder mystery, and neither the user nor the system is omniscient. That is, they have access only to information they can find out by investigating the murder. However, our current set up differs from that of our previous work in that the user is a junior detective, and the system is a desk-bound boss,



(a) Screen shot of Mr Body's bedroom

(b) Detective Gir's notebook

Figure 1: Sample screen of the WWW interface and Detective's Notebook

who knows only what the user tells him. Thus, the user does all the leg-work, navigating through a virtual crime scene, making observations and interviewing witnesses, and reports periodically to the boss. These reports consist of successively evolving arguments for the main suspect's guilt or innocence. Further, the user has limited resources, i.e., time and money, which are depleted as the investigation progresses. To win the game, the user must build a cogent argument regarding the guilt or innocence of the main suspect prior to exhausting his/her resources.

In order to evaluate the discourse interpretation capabilities of the system, in this paper we restrict the users' interaction with the system to a single round. That is, a user reads the initial police report, optionally explores the virtual scenario, and then presents an argument to his/her boss. The system interprets the argument, and presents its interpretation back to the user for validation. The results of this validation are discussed in our evaluation (Section 5). In the future, the boss will present counter-arguments, point out flaws in the user's argument or make suggestions regarding further investigations.

2.1 Playing the game – initial interaction

The game starts with the presentation of a police report that describes the preliminaries of the case for a particular scenario. The following police report is presented for the scenario used in this paper.

Yesterday, Mr Body was found dead in his bedroom. Fatal bullet wounds were found in Mr Body's chest.

Broken glass was found inside the bedroom window. A gun was found in the garden outside the house, and fingerprints were found on the gun.

Fresh footprints were found near the house, and some peculiar indentations were observed in the ground. Also, blue car paint was scraped on the letter box.

After reading the police report, the user may navigate through a virtual space to gather additional information (Figure 1(a) shows a screen shot of the victim's bedroom). The user may record information s/he considers interesting in his/her *Notebook* (Figure 1(b)), which is consulted by the user during the argument construction process. Upon completion of his/her investigation, the user builds an argument composed of a sequence of implications leading from evidence to the argument goal. Each implication is composed of one or more antecedents and consequents. In the current implementation, the antecedents and consequents are obtained by copying propositions from a drop-down menu into slots in the argument-construction interface.[1] Figure 2 shows a screen-shot of the argument-construction interface, and an argument built by a particular user

[1] An alternative version of our system accepts free-form Natural Language (NL) input for antecedents and consequents. However, in our current version this capability has been replaced with a web-based interface.

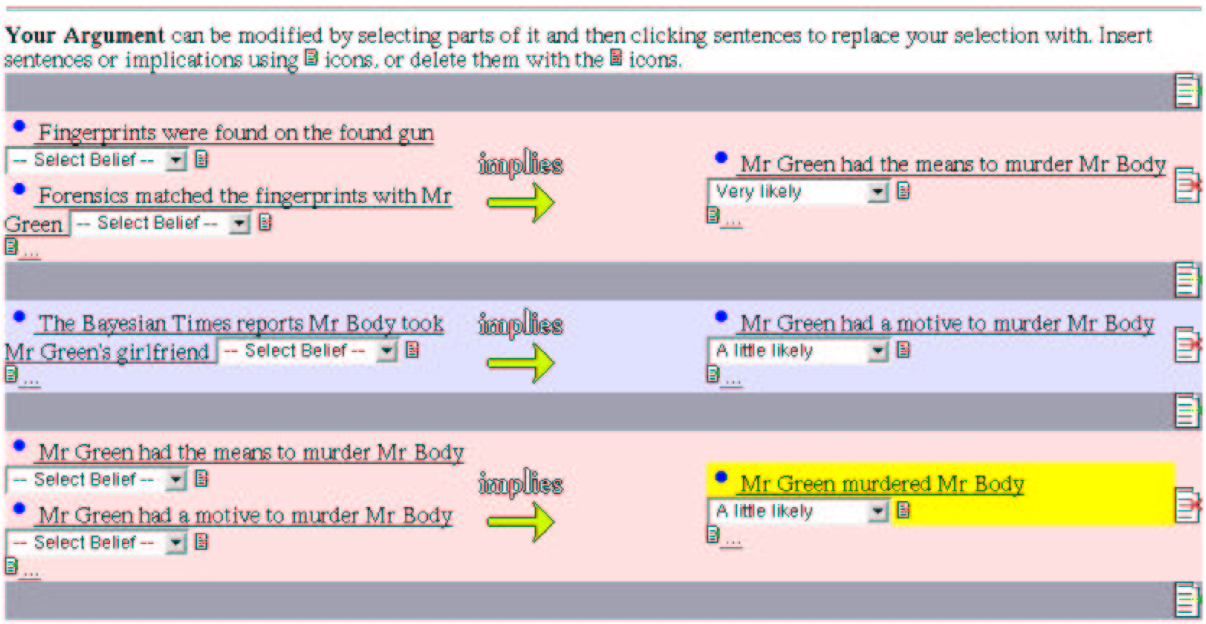


Figure 2: Argument-construction screen and user's argument

after she has read the police report, seen the newspaper and spoken to the forensic experts. Figure 3 shows the interpretation generated by BIAS for the argument in Figure 2. In it the system fills in propositions and relations where the user has made inferential leaps, and points out its beliefs and the user's.

2.2 Domain representation

The domain propositions and the relationships between them are represented by means of a Bayesian network (BN) (Pearl, 1988). Each BN in the system can support a variety of scenarios, depending on the instantiation of the evidence nodes. The murder mystery used for this paper is represented by means of an 85-node BN (similar to that used in (Zukerman, 2001)). Figure 4 shows a portion of this BN: the evidence nodes are boxed, and the goal node – GmurderedB – is circled. The five evidence nodes mentioned in the police report are boldfaced and shaded, the two evidence nodes obtained by the user in her investigation are boldfaced and dark shaded ([BayesianTimes Reports B Took G's Girlfriend] and [Forensics Match G's Fingerprints]), and the evidence nodes employed by the user in her argument have white text. The nodes corresponding to the consequents in the user's argument are boldfaced ([G Has Means], [G Has Motive] and [G Murdered B]).

3 Proposing Interpretations

BIAS generates interpretations of the user's argument in terms of its own beliefs and inferences, which may differ from those in the user's argument. This may require adding propositions and relations to the argument structure proposed by the user, deleting relations from the user's argument, or postulating degrees of belief in the propositions in the user's argument which differ from those stated by the user.

Our system generates candidate interpretations for an argument by finding different ways of connecting the propositions in the argument – each variant being a candidate interpretation. This is done by (1) connecting the nodes in the argument, (2) removing superfluous nodes, and (3) building connected sub-graphs of the resultant graph.

Connecting nodes. This is done by retrieving from the domain BN neighbouring nodes to the nodes mentioned in the user's argument (each proposition accessible through the argumentation interface corresponds to a node in the domain BN). BIAS then retrieves the neighbours' neighbours, and so on for a few iterations. These retrieved neighbours are *inferred* nodes. This process of retrieving neighbours enables us to model "inferential leaps", i.e., connections made by a user between two nodes that are not

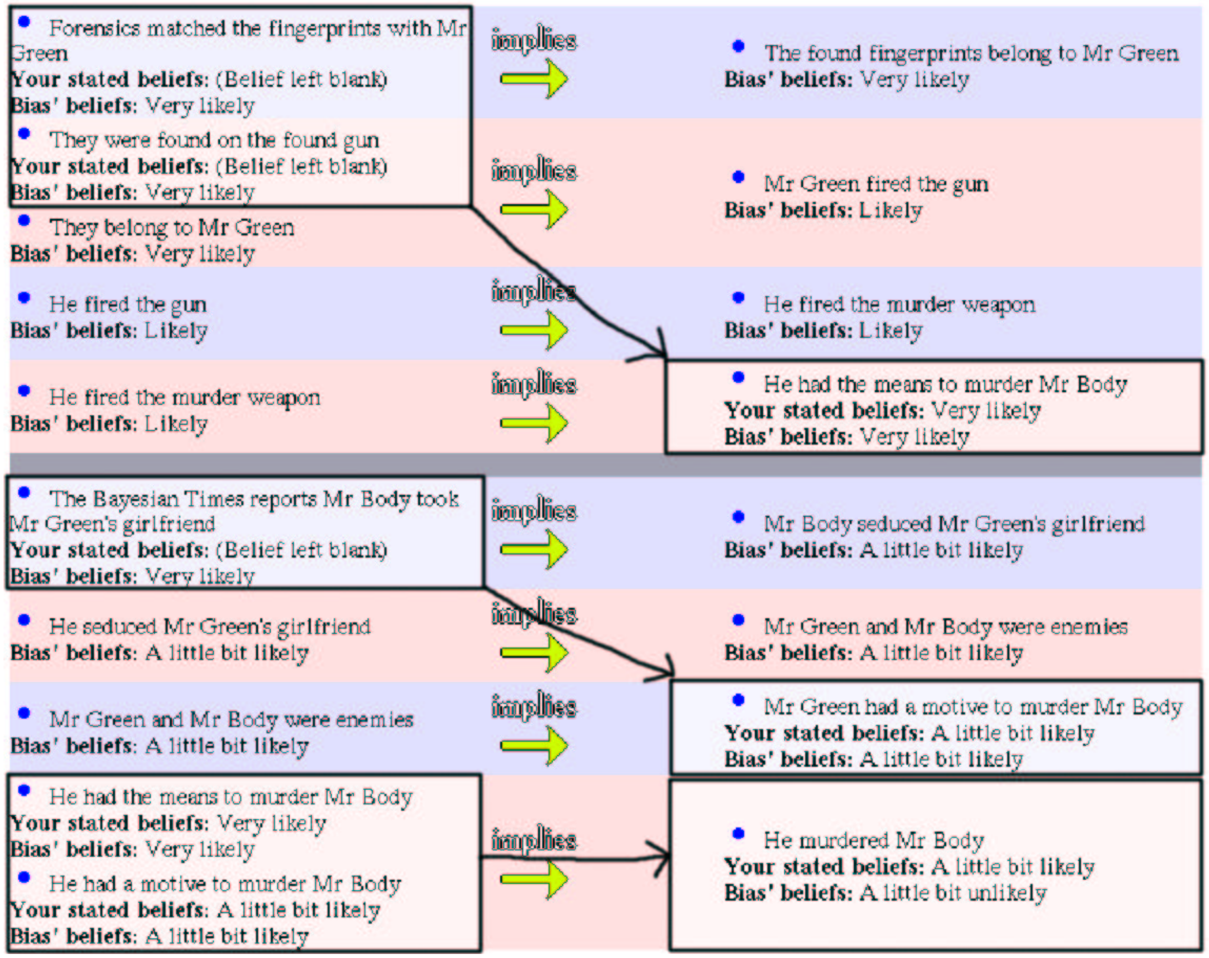


Figure 3: BIAS' interpretation of the user's argument

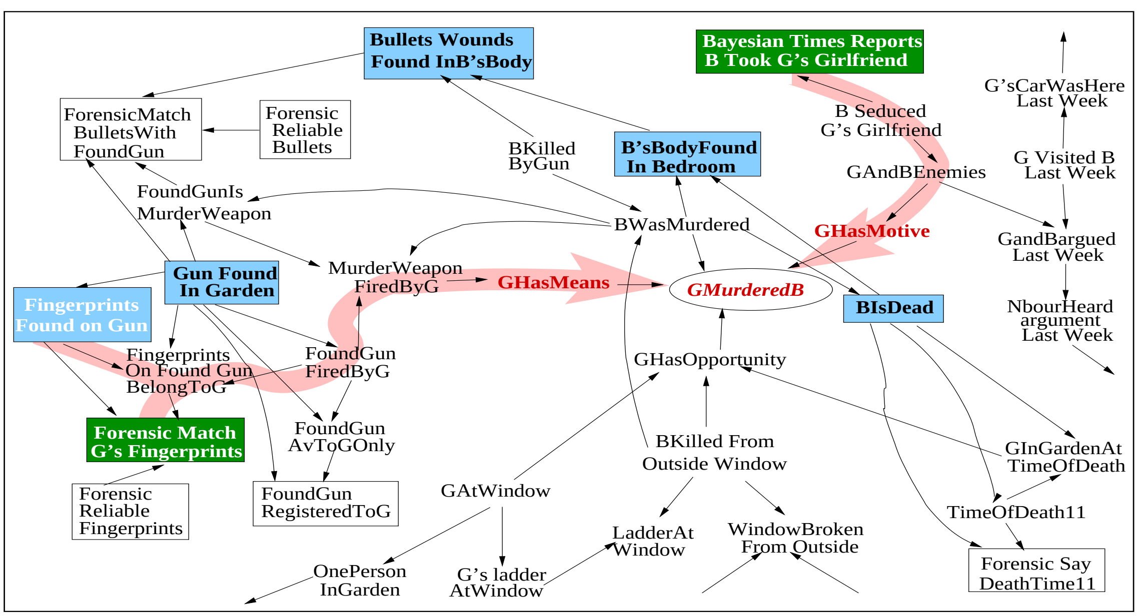


Figure 4: Interpretation of the user's argument and partial BN

adjacent in the domain BN. As a result of this process, mentioned nodes that are separated by a few inferred nodes in the domain BN will now be connected, but mentioned nodes that are far apart will remain unconnected. If upon completion of this process, a proposition in the user's argument is still unconnected, the system will have failed to find an interpretation (in the future, the user will be asked to fill this gap).

Removing superfluous nodes. This is done by marginalizing out nodes that are not on a path between an evidence node and the goal node.

Building sub-graphs. BIAS derives all the interpretations of an argument by computing combinations of paths which produce connected graphs that incorporate the nodes mentioned by the user.

The Bayesian subnets generated in this manner are candidate interpretations of a user's argument in terms of BIAS' domain knowledge. However, these subnets alone do not always yield the beliefs stated by the user, as the user may have taken into account implicit assumptions that influence his/her beliefs. For instance, the argument in Figure 2 posits a belief of A Little Likely in Mr Green's guilt, while Bayesian propagation from the available evidence yields a belief of A Little **Un**likely. This discrepancy may be attributed to the user's lack of consideration of Mr Green's opportunity to murder Mr Body (her argument includes only means and motive), an erroneous assessment of Mr Green's opportunity, or an assessment of the impact of opportunity on guilt which differs from BIAS'. In the future, our mechanism will consider the first two factors for neighbouring nodes of an interpretation (the third factor involves learning a user's Conditional Probability Tables – a task that is outside the scope of this project).

4 Selecting an Interpretation

In this section we present the Minimum Message Length criterion, describe its use for argument interpretation, and show how attentional focus is incorporated into our MML model.

4.1 MML Encoding

MML (Wallace and Boulton, 1968) is a *model selection criterion* (used to select between candidate models that explain observed data). The MML criterion implements Occam's Razor, which may be stated as follows: "If you have two theories which both explain the observed facts, then you should use the simplest until more evidence comes along" (the same idea is embodied in Einstein's aphorism "Make everything as simple as possible, but not simpler"). This criterion balances data fit with model complexity. That is, the best model should fit the data well *and* it should be simple. In probabilistic terms, given data D, the MML criterion selects the model M with the highest posterior probability.

$$\begin{aligned}\text{argmax}_M \Pr(M|D) &= \frac{\Pr(D\&M)}{\Pr(D)} \\ &= \frac{\Pr(M)\times\Pr(D|M)}{\Pr(D)}\end{aligned}$$

(the constant denominator can be ignored when selecting the highest-probability model.)

An optimal encoding for an event E with probability $\Pr(E)$ has message length $\text{ML}(E) = -\log_2 \Pr(E)$ (in bits). Hence, in information theoretic terms, the MML criterion selects the model M which yields the shortest message that transmits the model and the data.

$$\text{argmin}_M \text{ML}(D\&M) = \text{ML}(M) + \text{ML}(D|M)$$

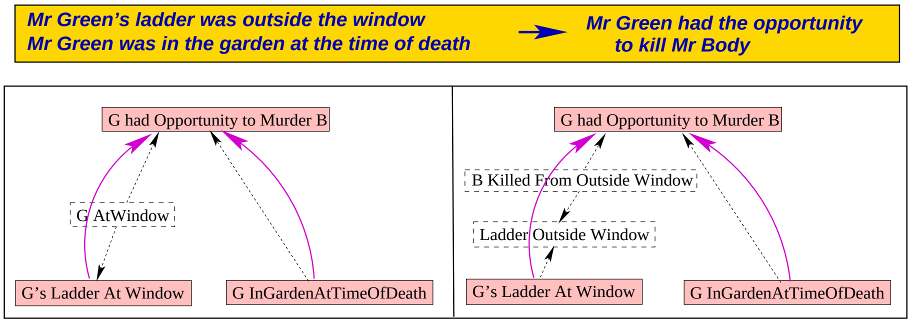
The message for the data and the model is composed of two parts: the first part transmits the model, and the second part transmits instructions for recovering the data from the model. The model for which $\text{ML}(D\&M)$ is minimal is the model with the highest posterior probability.

4.2 Evaluating Interpretations

The problem of selecting an interpretation for a user's argument among candidate interpretations may be viewed as a model selection problem, where the argument is the data, and the interpretation is the model. Let *Arg* be a graphical representation of an argument (with antecedents pointing to consequents), and *SysInt* an interpretation generated by our system. Thus, we are looking for the *SysInt* which yields the shortest message length for

$$\text{ML}(\textit{Arg}\&\textit{SysInt}) = \text{ML}(\textit{SysInt}) + \text{ML}(\textit{Arg}|\textit{SysInt})$$

The first part of the message describes the interpretation, and the second part describes how to reconstruct the argument from the interpretation. The



SysInt$_A$ *SysInt*$_B$

Figure 5: Interpretation of a simple argument

expectation from using the MML criterion is that in finding an interpretation that yields the shortest message for an NL argument (i.e., the interpretation with the highest posterior probability), we will have produced a plausible interpretation, which hopefully is the intended interpretation. This interpretation is determined by comparing the message length of the candidate interpretations, which are obtained as described in Section 3.

Since domain propositions (rather than NL sentences) are used to construct an argument, *Arg* can be directly obtained from the input.[2] *SysInt* is then derived from *Arg* by using the links and nodes in the domain BN to connect the propositions in the argument (Section 3). When the underlying representation has several ways of connecting between the nodes in *Arg*, then more than one candidate *SysInt* is generated (each candidate has at least one *inferred* node that does not appear in the other candidates). This is the case for the simple argument in Figure 5, which is composed of two antecedents and one consequent. This argument has two interpretations (the portion of the domain BN from which the interpretations are extracted appears at the bottom of Figure 4). The inferred nodes for the two interpretations in Figure 5 are [G At Window] for *SysInt*$_A$, and [Ladder Outside Window] and [B Killed From Outside Window] for *SysInt*$_B$; the inferred links are drawn with dashed lines (the nodes in *Arg* are shaded, and the links are curved).

After candidate interpretations have been postulated, the MML criterion is applied to select the best interpretation, i.e., the interpretation with the shortest message. The calculation of the message length takes into account the following factors: (1) the size of an interpretation, and (2) the structural and belief similarity between the interpretation and the argument. These factors influence the components of the message length as follows.

- ML(*SysInt*) represents the probability of *SysInt*. According to the MML principle, concise interpretations (in terms of number of nodes and links) are more probable than more verbose interpretations.

- ML(*Arg*|*SysInt*) represents the probability that a user uttered *Arg* when s/he intended *SysInt*. This probability depends on the structural similarity between *SysInt* and *Arg* (in terms of the nodes and links that appear in *Arg* and *SysInt*), and on the similarity between the beliefs in the nodes in *Arg* and the corresponding nodes in *SysInt* (the beliefs in the nodes in *SysInt* are obtained by performing Bayesian propagation through *SysInt*; thus, different *SysInt*s may yield different beliefs in the consequents of an argument). According to this component, interpretations that are more similar to *Arg* are more probable (i.e., yield a shorter message) than interpretations that are less similar to *Arg*.

Table 1 summarizes the effect of these factors on the message length for *SysInt*$_A$ and *SysInt*$_B$ (Figure 5). *SysInt*$_A$ is simpler than *SysInt*$_B$, thus yielding a shorter message length for the first component of the message. *SysInt*$_A$ is structurally more similar to *Arg* than *SysInt*$_B$, yielding a shorter message length for this aspect of *Arg*|*SysInt*$_A$ (*SysInt*$_A$ differs from *Arg* by 1 node and 3 links, while *SysInt*$_B$

[2] This is not the case in the version of the system which takes NL input, since there may be more than one proposition that is a reasonable interpretation for a sentence in an argument.

Table 1: Message length comparison of two interpretations

ML of	Factor	$SysInt_A$	$SysInt_B$	Shortest ML
SysInt	Size	4 nodes, 3 links	5 nodes, 4 links	$SysInt_A$
$Arg \vert SysInt$	Structural similarity	1 node, 3 links difference	2 nodes, 4 links difference	$SysInt_A$
	Belief similarity	less similar	more similar	$SysInt_B$

differs from *Arg* by 2 nodes and 4 links[3]). In this example, we assume that the belief in [G had Opportunity to Murder B] in $SysInt_B$ is stronger than that in $SysInt_A$, and hence closer to the asserted consequent of the argument in Figure 5. This yields a shorter message length for the belief component of $\mathrm{ML}(Arg|SysInt_B)$. However, this is not sufficient to overcome the shorter message length of $SysInt_A$ due to structural similarity and conciseness. Thus, although both interpretations of the argument are reasonable, $SysInt_A$ is the preferred interpretation.

As seen in this example, the MML criterion weighs possibly conflicting considerations during discourse interpretation, e.g., a verbose interpretation that is similar to a user's argument is preferred to a more concise interpretation that is dissimilar.

4.3 Modeling Attentional Focus

The above-presented interpretation process assumes that all inferred propositions are equally likely to be included in a user's argument. However, this is not necessarily the case. We posit that a user is more likely to imply (inside an inferential leap) propositions previously seen by him/her than propositions s/he has never encountered. In this case, the length of the part of the message which conveys *SysInt* should not only depend on the size of the interpretation (in terms of number of nodes and links), but also on the probability that the user will employ in his/her argument the nodes in that interpretation.

We have modeled the probability that a user will include a proposition in his/her argument as a function of the proposition's presence in the user's focus of attention. Attentional focus in turn was modeled by means of an activation level which depends on the following factors:[4]

- the type of access of a proposition, e.g., whether the proposition was copied from the menu or seen when exploring the scenario; and
- the passage of time, i.e., the longer the time elapsed since the last access to the proposition, the lower its level of activation.

These factors were combined as follows to express the probability that a proposition will be included in an argument:

$$\mathrm{Pr}(Prop) \propto \sum_{i=1}^{n} AccessType_i(Prop) \times [CurTime - TimeStmp_i + 1]^{-b}$$

where n is the number of times the proposition was accessed, *AccessType* is a score that reflects the manner in which the proposition was accessed, $b = 1$ is an empirically determined exponent, *CurTime* is the current time, and $TimeStmp_i$ is the time of the ith access. According to this formula, when a proposition is accessed, activation is added to the current accumulated (and decayed) activation. That is, there is a spike in the level of activation of the proposition, which starts decaying from that point again.

To illustrate the effect of attentional focus on the argument interpretation process, let us reconsider the sample argument in Figure 5, and let us assume that [G At Window] was never seen by the user, while [Ladder Outside Window] and [B Killed From Outside Window] were seen recently. In this case, a high probability for these two propositions may overcome the factors in favour of $SysInt_A$, thereby making $SysInt_B$ the preferred interpretation.

[3]There are 2 links in $SysInt_A$ that are not in *Arg*, and 1 link in *Arg* that is not in $SysInt_A$ (total 3 links difference). A similar calculation is performed for $SysInt_B$.

[4]Other factors, such as intrinsic salience of a proposition, have not been modeled at present.

5 Evaluation

The system was evaluated in two modes: (1) automatic and (2) user based.

Our automatic evaluation, described in (Zukerman and George, 2002), consisted of having the system interpret noisy versions of its own arguments. These arguments were generated from different subnets of its domain BN, and they were distorted by changing the beliefs in the nodes, and inserting and deleting nodes and arcs. All these distortions were performed on BNs of different sizes (3, 5, 7 and 9 arcs). Our measure of performance was the edit-distance between the original BN used to generate an argument, and the BN produced as the interpretation of this argument. BIAS produced an interpretation in 86% of the 5400 trials. In 75% of the 5400 cases, the generated interpretations had an edit-distance of 3 or less from the original BN (e.g., the interpretation differed from the original argument by one node and two links), and in 50% of the cases, the interpretations matched perfectly the original BN.

A preliminary user-based evaluation was performed with 10 computer-literate staff and students from our University. Our evaluation was conducted as follows. We introduced the users to our system, and explained its aims. We then encouraged them to explore the scenario, and when they were ready, they built an argument using the interface shown in Figure 2. BIAS then generated an interpretation of the argument, presenting it as shown in Figure 3. The users were asked to assess BIAS' interpretation under two conditions: before and after seeing a diagram of the domain BN. In the initial assessment, the users were asked to give BIAS' interpretation a score between 1 (Very UNreasonable) and 5 (Very Reasonable), and to optionally provide further comments. In the second assessment, the users were given the complete diagram for the partial BN shown in Figure 4, and were asked to re-assess BIAS' interpretation in light of this domain knowledge. They were also asked to trace their preferred interpretation on the diagram (on paper).

Our users found the system somewhat daunting, and indicated that the interface for entering an argument was inconvenient. We believe that this was partly due to their lack of familiarity with the available domain propositions. That is, the users were faced with 85 new propositions, which they had to read in order to determine which one(s) they could use to express what they had in mind. Nonetheless, the users managed to construct arguments, which ranged in size from 2 propositions to 26, and gave a generally favourable assessment of BIAS' interpretations. Overall the average score of BIAS' interpretations was 4 before seeing the BN diagram and 4.25 after seeing the diagram. This indicates that understanding the constraints imposed by a system's domain knowledge may influence users' ability to interact with the system.

The main lessons learned from this preliminary evaluation pertain to two aspects: (1) the interface, and (2) the use of BNs for discourse understanding. In order to improve the usability of the interface, we will integrate it with BIAS' NL module. It is envisaged that a solution combining menus and NL input will yield the best results. Our evaluation also corroborates the insights from Section 3 regarding the difficulties of taking into account users' assumptions during the argument interpretation process. However, the results of our evaluation are encouraging with respect to the use of the MML principle for the selection of interpretations. In the future, we propose to conduct a more rigorous evaluation with additional users to confirm these results.

6 Related Research

Our research builds on work described in (Zukerman, 2001; Zukerman and George, 2002), which integrates plan recognition for discourse understanding with BNs. Zukerman (2001) used a domain model and user model represented as a BN, together with linguistic and attentional information, to infer a user's goal from a single-proposition rejoinder to a system-generated argument. However, the combination of these knowledge sources was based on heuristics. Zukerman and George (2002) developed the initial probabilistic model for applying the MML principle to argument interpretation. Here we extend this model to include attentional focus, integrate it into an interactive interpretation system, and evaluate it with users.

The MML principle (Wallace and Boulton, 1968) is a model-selection technique which applies information-theoretic criteria to trade data fit against model complexity. MML has been used in a variety of applications, several of which are listed in `http://www.csse.monash.edu.au/~dld/Snob.application.papers`. In this paper, we demonstrate the applicability of MML to a high-level NL task.

BNs have been used in several systems that perform plan recognition, e.g., (Charniak and Goldman, 1993; Gertner et al., 1998; Horvitz and Paek, 1999). Charniak and Goldman's system (Charniak and Goldman, 1993) handled complex narratives, using a BN and marker passing for plan recognition. It automatically built and incrementally extended a BN from propositions read in a story, so that the BN represented hypotheses that became plausible as the story unfolded. Marker passing was used to restrict the nodes included in the BN. In contrast, we use domain knowledge to constrain our understanding of the propositions in a user's argument, and apply the MML principle to select a plausible interpretation. Gertner *et al.* (1998) used a BN to represent the solution of a physics problem. After observing an action performed by a student, their system (Andes) postulated candidate interpretations (like BIAS' *SysInt*), each hypothesizing subsequent actions. Unlike Andes, BIAS is presented with a complete argument. Hence, it must also consider the fit between all the argument propositions and the interpretation ($Arg|SysInt$). Finally, the system developed by Horvitz and Paek (1999) handled short dialogue contributions, and used BNs at different levels of an abstraction hierarchy to infer a user's goal in information-seeking interactions with a Bayesian Receptionist. In addition, they employed decision-theoretic strategies to guide the progress of the dialogue. We expect to use such strategies when our system engages in a full dialogue with users.

7 Conclusion

We have offered a mechanism based on the MML principle that generates interpretations of extended arguments in the context of a BN. The MML principle provides a theoretically sound framework for weighing possibly conflicting considerations during discourse interpretation. This framework enables us to represent structural discrepancies between the underlying, detailed domain representation and the more sparse arguments produced by people (which typically contain inferential leaps). The results of our formative evaluation are encouraging, supporting the application of the MML principle for argument interpretation.

Acknowledgments

This research was supported in part by Australian Research Council grant A49927212. The authors thanks Charles Twardy on insightful comments regarding MML.

References

Eugene Charniak and Robert P. Goldman. 1993. A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):50–56.

Abigail Gertner, Cristina Conati, and Kurt VanLehn. 1998. Procedural help in Andes: Generating hints using a Bayesian network student model. In *AAAI98 – Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 106–111, Madison, Wisconsin.

Eric Horvitz and Tim Paek. 1999. A computational architecture for conversation. In *UM99 – Proceedings of the Seventh International Conference on User Modeling*, pages 201–210, Banff, Canada.

Judea Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann Publishers, San Mateo, California.

C.S. Wallace and D.M. Boulton. 1968. An information measure for classification. *The Computer Journal*, 11:185–194.

Ingrid Zukerman and Sarah George. 2002. A Minimum Message Length approach for argument interpretation. In *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, pages 211–220, Philadelphia, Pennsylvania.

Ingrid Zukerman. 2001. An integrated approach for generating arguments and rebuttals and understanding rejoinders. In *UM01 – Proceedings of the Eighth International Conference on User Modeling*, pages 84–94, Sonthofen, Germany.