Improving the Presentation of Argument Interpretations Based on User Trials

Ingrid Zukerman, Michael Niemann, and Sarah George

School of Computer Science and Software Engineering,
Monash University,
Clayton, Victoria 3800, Australia
{ingrid, niemann, sarahg}@csse.monash.edu.au

Abstract. The interpretation of complex discourse, such as arguments, is a difficult task that often requires validation, i.e., a system may need to present its interpretation of a user's discourse for confirmation. In this paper, we consider the presentation of discourse interpretations in the context of a probabilistic argumentation system. We first describe our initial approach to the presentation of an interpretation of a user's argument; this interpretation takes the form of a Bayesian subnet. We then report on the results of our preliminary evaluation with users, focusing on their criticisms of our system's output. These criticisms motivate a content enhancement procedure that adds information to explain unexpected outcomes and removes superfluous content from an interpretation. The discourse generated by this procedure was found to be more acceptable than the discourse generated by our original method.

1 Introduction

The interpretation of complex discourse, such as arguments, is a difficult task that often requires validation. That is, if a system is uncertain about its understanding of a user's discourse, it should present its interpretation for confirmation or correction. In this paper, we describe the content enhancement component of a probabilistic argumentation system currently under development. Our system receives arguments from users, interprets these arguments in the context of its domain knowledge, and generates responses. The arguments take the form of Natural Language (NL) sentences linked by means of argument connectives. A sample argument is shown in the top left of Figure 1, and its gloss appears in the top right of the Figure. Given an internal representation of an interpretation or an argument (right-hand side of Figure 1), our content enhancer determines information to be added to or removed from this representation to make its presentation more acceptable to people.

Our system uses Bayesian networks (BNs) [1] as its knowledge representation and reasoning formalism. Our domain of implementation is a murder mystery, for which we have designed several BNs. Each BN 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 a 32-node BN, which is illustrated in Figure 2 (the evidence

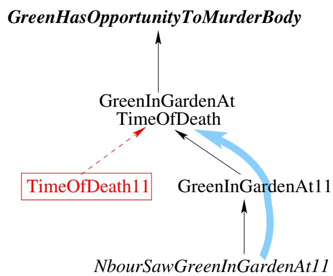
G.I. Webb and Xinghuo Yu (Eds.): AI 2004, LNAI 3339, pp. 587–598, 2004.

Argument
The neighbour saw Mr Green in the garden at 11
IMPLIES
Mr Green had the opportunity to murder MrBody
[ALittleUnlikely]

Gloss
Even though the neighbour saw Mr Green in the garden at 11, Mr Green possibly did not have the opportunity to murder Mr Body

Original interpretation
The neighbour seeing Mr Green in the garden at 11 implies that he very probably was in the garden at 11, which implies that he probably was not in the garden at the time of death. This implies that he possibly did not have the opportunity to murder Mr Body.

Interpretation (Bayesian subnet)



Enhanced interpretation
Even though the neighbour saw Mr Green in the garden at 11, the time of death not being 11 implies that Mr Green probably was not in the garden at the time of death, which implies that he possibly did not have the opportunity to murder Mr Body.

Fig. 1. Sample argument and interpretation

nodes are boxed, the observed evidence nodes are boldfaced and boxed, and the arrows show the causal relationships between the nodes).

The interaction between a user and the system proceeds as follows. The user first obtains information about a murder (in our examples Mr Body was murdered), and then builds an argument regarding the guilt or innocence of a particular suspect (Mr Green). The argument is typically composed of a sequence of implications leading from observable evidence to the argument goal, where each implication is composed of one or more antecedents and a consequent. Our system's discourse interpretation component matches the user's sentences with the nodes in the BN, and then derives an interpretation by finding a concise reasoning path or graph (a subnet of the domain BN) that connects the nodes in the argument, taking into consideration the information obtained by the user and the inference patterns within the BN. When generating an interpretation, the system attempts to make the structure and beliefs in the Bayesian subnet as similar as possible to what was stated by the user, within the limitations of its world model. This means that sometimes the beliefs in an interpretation do not match exactly the user's stated beliefs.

The black arcs and nodes in the Bayesian subnet at the bottom right of Figure 1 illustrate an interpretation generated for the argument at the top left of the Figure; the italicized nodes in the subnet are those mentioned in the argument (this interpretation may also be traced in the bottom right-hand side of Figure 2). The text corresponding to this interpretation appears in the middle left of Figure 1. Our content enhancer adds the grey, boxed node to the subnet in Figure 1, and skips the node overwritten by the thick grey arrow. The resultant interpretation is then presented to the user both in the same format as the argument and in the textual form shown in the bottom left of Figure 1.

In the next section, we describe our preliminary trial with users, focusing on their criticisms of our system's interpretations. In Section 3, we present the content enhancer we implemented to address these concerns, followed by the results of our evaluation. We then discuss related research, and present concluding remarks.

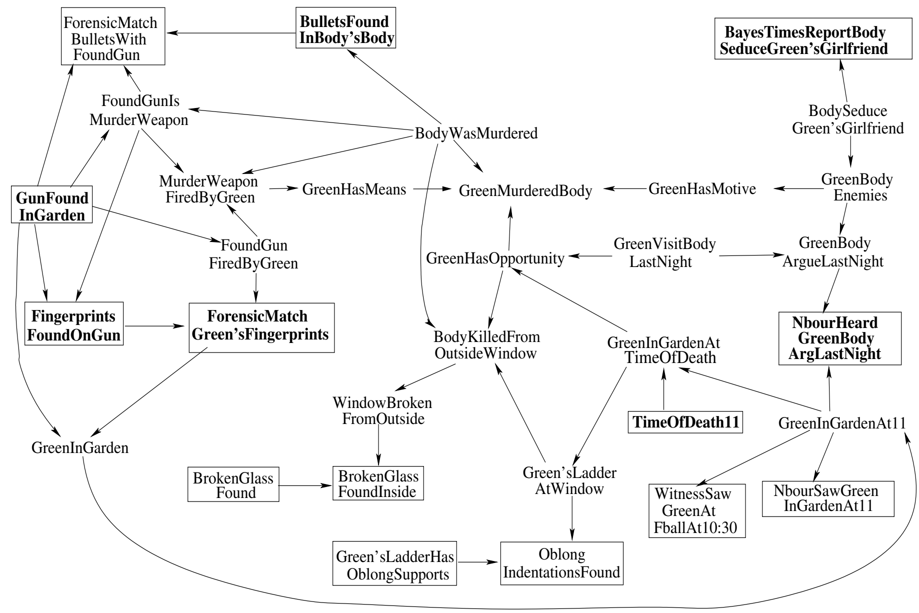


Fig. 2. Sample domain BN

2 Preliminary Trial – Results

Our initial generation technique consisted of simply following the links in the Bayesian subnet corresponding to an interpretation, and presenting the *canonical sentences* that are associated with each node in the subnet. For instance, the canonical sentence for GreenHasOpportunityToMurderBody is "Mr Green *have* the opportunity to murder Mr Body" (the verb "have" is inflected during the realization process). This process yields the text labelled "Original interpretation" in Figure 1.

The belief in a node being true is divided into the following seven belief categories: {VeryLikely, Likely, ALittleLikely, EvenChance, ALittleUnlikely, Unlikely, VeryUnlikely}, which are based on those offered in [2]. When rendered in English, they yield the following terms: {"very probable", "probable", "possible", "maybe"} – the probabilities under 0.5 are rendered in English by negating the verb, e.g., "Mr Green probably *did not* murder Mr Body".[1]

Our preliminary trial was designed to evaluate the performance of our argument interpretation component. We prepared four pencil-and-paper evaluation sets, each comprising of a short argument and between one and three candidate interpretations generated by our system. Each set was shown to between 15 and 23 subjects. Our subjects were asked to assess each interpretation, and to comment on aspects of the interpretations they liked or disliked. Although there was general acceptance of our system's interpre-

[1] We conducted trials to test people's understanding of these linguistic representations of the belief categories in comparison to alternative wordings. The above terms yielded the most consistent understanding (with the lowest standard deviation).

tations and its reasoning, our subjects' comments focused on the information included to describe these interpretations. Their concerns were that the interpretations were not presented in a concise yet complete manner that fully represents how the propositions in an interpretation influence each other. The original interpretation in Figure 1 illustrates two of the problems which are typical of those pointed out by our trial subjects: too little information and too much information. The inference from "Mr Green being in the garden at 11" to "Mr Green *not* being in the garden at the time of death" appears to be a non-sequitur, i.e., the information provided is insufficient to make sense of this inference. The inference from "the neighbour seeing Mr Green in the garden at 11" to "Mr Green being in the garden at 11" was viewed as being superfluous, i.e., too much information is being presented.

Below we summarize the main problems identified by our trial users.

- **Too Much Detail –** users became annoyed when obvious inferences (such as the second example above) were stated.
- **Discrepancies Between the Beliefs in an Argument and in Its Interpretation –** users were disconcerted when the system claimed to have interpreted an argument, but the beliefs that were presented differed from the beliefs stated within the argument (recall that our system cannot always match a user's stated beliefs with those obtained in the Bayesian subnet).
- **Increase in Certainty –** users objected to inferences where the consequent had a greater degree of certainty than its antecedents, e.g., "Mr Green *probably* had the means to murder Mr Body. Therefore, Mr Green *very probably* murdered Mr Body".
- **Large Change in Certainty –** users accepted small decrements in certainty, e.g., from "probable" to " possible". However, they objected to inferences where the belief in the consequent was substantially different from the belief in the antecedent (regardless of whether the difference represented an increase or a decrease in the level of certainty), e.g., the inference in Figure 1 from "Mr Green very probably being in the garden at 11" to "Mr Green probably *not* being in the garden at the time of death".

3 Enhancing the Content of a Presentation

Our content enhancement process identifies the above problems computationally, and addresses them by adding information to an interpretation or removing information.

3.1 Too Much Detail

In order to avoid unnecessary detail in the description of an interpretation, we used the following rule to identify potentially superfluous nodes, with the aim of removing them.

Rule 1

If $Node_B$ *is between* $Node_A$ *and* $Node_C$ ($Node_A \rightarrow Node_B \rightarrow Node_C$) *AND*
$Node_B$ *is similar to* $Node_A$ *AND*
BeliefCategory($Node_B$*) = BeliefCategory(*$Node_A$*) THEN*
$Node_B$ *is superfluous (its omission yields the implication '*$Node_A \rightarrow Node_C$*')*

Table 1. Sample nodes removed from our trial interpretations

(a) The neighbour saw Mr Green in the garden at 11 [VeryLikely] IMPLIES **[OMITTED] Mr Green was in the garden at 11 [VeryLikely]** **IMPLIES** Mr Green was in the garden at the time of death [VeryLikely]
(b) The neighbour heard Mr Green arguing with Mr Body last night [VeryUnlikely] IMPLIES **[OMITTED] Mr Green and Mr Body had a loud argument last night [VeryUnlikely]** **IMPLIES** Mr Green visited Mr Body last night [VeryUnlikely]

As stated in Section 2, each node in the BN is associated with a canonical sentence which represents the information in the node. These sentences are used in two ways: (1) to find the best match with the statements in a user's argument in order to map the argument onto the BN, and (2) to assess the similarity between two nodes for the application of Rule 1. For both usages, the similarity between two sentences is estimated by means of a modified version of the cosine similarity measure [3], which calculates the angle between the vectors comprising the words in the sentences. Our version ignores stop words (high frequency words that are generally ignored for retrieval purposes) and proper nouns that are common to most nodes, e.g., "Green" and "Body", and takes into account synonyms and near synonyms, e.g., "glass" and "window". For the node-similarity usage, if the (normalized) similarity score between an antecedent node and a consequent node exceeds a threshold, then the nodes are regarded as similar, thereby satisfying the second antecedent of Rule 1. If the nodes also have the same level of certainty, then the consequent node may be treated as superfluous and removed.

Clearly, Rule 1 may yield several candidates for omission along a reasoning path. However, we want to avoid removing too many nodes in order to safeguard against too much information being lost. This is done by applying a greedy algorithm that inspects each node along a reasoning path, and ranks the nodes according to their similarity score with respect to their antecedent. These nodes are then considered for removal in highest-to-lowest order of similarity score. However, no node can be removed if it was in the original argument or if it is the consequent of an already removed node. For instance, given $A \rightarrow B \rightarrow C \rightarrow D$, if B is similar to A and they have the same belief, then B is omitted, and C can not be removed even if it is similar to A. If only C and D were found to be similar, then C will be omitted, as the consequent D was in the original argument.

Table 1 displays fragments from two of the interpretations shown to our trial subjects, highlighting the information omitted as a result of the application of Rule 1. The application of this rule reduces the verboseness of the output by removing propositions that a reader can easily deduce from their antecedent.

3.2 Discrepancies Between Beliefs

As stated in Section 1, the interpretation generated by the system does not always match the beliefs in a user's argument. For instance, say the user states that fact A being true implies that fact B is Likely to be true. The system will endeavour to find a path from the node corresponding to A to the node corresponding to B that makes B Likely. However, the system's beliefs are restricted by what can be calculated from the *Conditional Probability Tables (CPTs)* of the nodes in the BN.[2] Hence, it is possible that the closest belief our system can obtain for B is ALittleLikely.

If such a discrepancy in belief occurs, then the following information is included in the presentation.

– A preamble such as the following: "I know this is not quite right, but it is the best I could do given what I believe about this situation".
– An explanation at each point of discrepancy, e.g., "I know that your belief is stronger, but this is the closest I can come up with".

At the knowledge representation level, the system could update its inference patterns to match the user's. In the context of BNs, this involves modifying the CPTs for the offending implications. However, this is a complex process, with far-reaching implications with respect to the system's inference patterns. This type of solution is left for future investigation, as the CPTs in our system are regarded as static.

3.3 Increase in Certainty and Large Change in Certainty

Given an implication such as $A \rightarrow B$, users generally accepted inferences that yielded the same or a slightly weaker belief in B than the belief in A (a belief is weaker if it leans towards greater uncertainty about whether a fact is true or false, i.e., even chance). However, as indicated above, they objected to all other belief changes between antecedents and consequents. This situation is identified by means of the following rule.

Rule 2

For Node$_A$ [BeliefCategory$_A$] $\rightarrow$ Node$_B$ [BeliefCategory$_B$]:
// increase in certainty
If BeliefCategory$_A$ is weaker than BeliefCategory$_B$ OR
// large change in certainty
BeliefCategory$_A$ differs from BeliefCategory$_B$ by more than 1 level THEN
a "leap" in belief has occurred

Some of the identified leaps in belief are due to the CPTs of the implications in the reasoning path, and can be justified only by explaining the CPTs (a task that is outside the scope of this paper). However, most of these leaps are due to influences from nodes that are not part of the initial interpretation, and should be included in the presentation of the interpretation to make it acceptable to users. These influencing nodes are siblings of the antecedents of the offending implications. That is, they are nodes

[2] The CPT of a node represents the influence of its parent nodes in the BN on the beliefs in this node. The CPTs in our BN were derived from human knowledge of the domain.

Algorithm *SelectInfluences(Implication, BN)*

1. Get Antecedents and Consequent for the Implication
2. For each Sibling of Antecedents that is an influence node
   (a) // Decreasing belief between antecedent and consequent
   If Belief(Antecedent) > Belief(Consequent) THEN
      i. If BeliefCategory(Sibling) = BeliefCategory(Consequent) THEN
         store Sibling as a `StandardInfluence`
      ii. Else if BeliefCategory(Sibling) < BeliefCategory(Consequent) THEN
         store Sibling as an `LargeInfluence`
      iii. Else //BeliefCategory(Sibling) > BeliefCategory(Consequent)
            if BeliefCategory(Antecedent) > BeliefCategory(Sibling) THEN
            store Sibling as a `WeakInfluence`
   (b) // Increasing belief between antecedent and consequent
   If Belief(Antecedent) < Belief(Consequent) THEN
      i. If BeliefCategory(Sibling) = BeliefCategory(Consequent) THEN
         store Sibling as a `StandardInfluence`
      ii. Else if BeliefCategory(Sibling) > BeliefCategory(Consequent) THEN
         store Sibling as an `LargeInfluence`
      iii. Else // BeliefCategory(Sibling) < BeliefCategory(Consequent)
            if BeliefCategory(Antecedent) < BeliefCategory(Sibling) THEN
            store Sibling as a `WeakInfluence`
3. Add the influences in the highest-ranked category to the implication

Fig. 3. Algorithm for selecting influence nodes

connected to the consequent of these implications, e.g., TimeOfDeath11 is a sibling of GreenInGardenAt11 in Figure 1. However, not every sibling is necessarily an influence, and not all siblings that are influences should be included in an interpretation. For instance, a sibling node that represents unobserved evidence is not an influence (the evidence has not been gathered, hence the value of this node is unknown). In contrast, an influencing node must either be an observed evidence node, or it must be influenced by a neighbouring node that is an observed evidence node. Also, we aim to include in an interpretation the minimum number of influences that explain the belief in a consequent. Hence, weak influences will be omitted if stronger influences are present.

After a leap in belief has been identified in an implication, algorithm *SelectInfluences* is activated in order to determine which influence nodes to present (Figure 3).[3] To this effect, our algorithm divides the influencing siblings of the implication's antecedents into three categories based on the strength of the belief in each sibling compared to the strength of the belief in the consequent and the antecedents: large > standard > weak. For example, a sibling with a large influence is one that has a more extreme belief than the consequent. Such a sibling is estimated to have a strong "pull" in its own direction, and hence provides a good explanation for the current (unintuitive) belief in the consequent. Since our system tries to minimize the number of inclusions in an interpretation, it adds to the implication only the nodes in the highest non-empty category.

[3] For clarity of exposition, we show only the "positive" version of the *SelectInfluences* algorithm. This version works for siblings that increase the belief in a consequent when they are true, and decrease its belief when they are false. The "mirror image" of this version is applied when a false sibling yields a true belief in the consequent, and a true sibling yields a false belief.

Upon completion of this enhancement step, the nodes added to each implication are incorporated in the presentation of the implication by means of appropriate connectives. For instance, additive expressions, such as “together with”, are used when presenting nodes that explain increases in certainty (provided the nodes are on the same side of EvenChance as the antecedent), while adversative expressions, such as “however” and “despite”, are used when presenting nodes that explain reductions in certainty or movements across the EvenChance divide.

To illustrate the workings of algorithm *SelectInfluences*, let us consider the argument fragments in Table 2, which appeared in interpretations shown to our trial subjects (the argument goal in these interpretations was either GreenMurderedBody or GreenHasOpportunity). These fragments exhibit changes in certainty which our subjects found confusing, and which were made more understandable by the addition of influence nodes.

The interpretation fragment in Table 2(a) goes from Mr Green *very probably* being in the garden to Mr Green *possibly **not*** being in the garden at 11. This is a case of a decrease in belief (Step 2a) coupled with a large drop in certainty. Our algorithm examines the siblings of GreenInGarden, which are NbourSawGreenInGardenAt11, WitnessSawGreenAtFootballAt10:30 and NbourHeardGreenBodyArgueLastNight (Figure 2), in order to find the strongest negative influences that explain this decrease in belief (even though GreenInGardenAtTimeOfDeath is a sibling of the antecedent, it is not considered because it is already part of the interpretation as the consequent of GreenInGardenAt11). First our algorithm determines whether a sibling has a negative influence, and if it does, the sibling is assigned an influence category. However, in this example the first two siblings are unobserved evidence nodes, which do not contribute to the information content of the interpretation. Hence, the only candidate for inclusion in the interpretation is NbourHeardGreenBodyArgueLastNight. This node is assigned the large influence category, as

Table 2. Sample nodes added to our trial interpretations

(a) Large change in certainty, decrease in belief – addition of `LargeInfluence` Mr Green was in the garden [VeryLikely] **BUT** **[ADDED] The neighbour heard Mr Green arguing with Mr Body last night [VeryUnlikely]** IMPLIES Mr Green was in the garden at 11 [ALittleUnlikely]
(b) Large change in certainty, decrease in belief – addition of `LargeInfluence` Mr Green had the means to murder Mr Body [Likely] **BUT** **[ADDED] Mr Green had the opportunity to murder Mr Body [ALittleUnlikely]** IMPLIES Mr Green murdered Mr Body [EvenChance]
(c) Increase in certainty, increase in belief – addition of `StandardInfluence` Mr Green visited Mr Body last night [ALittleUnlikely] **BUT** **[ADDED] Mr Green was in the garden at the time of death [VeryLikely]** IMPLIES Mr Green had the opportunity to murder Mr Body [VeryLikely]

its belief category is more extreme than that of the consequent, and it is then added to the interpretation.

The interpretation fragment in Table 2(b) goes from Mr Green *probably* having the means to murder Mr Body to him *maybe* murdering Mr Body. This is also a case of a decrease in belief (Step 2a), and decrease in certainty. Our algorithm examines the siblings of GreenHasMeans, which are BodyWasMurdered, GreenHasMotive and GreenHasOpportunity (Figure 2), in order to find the strongest negative influences. Since BodyWasMurdered has a positive influence, it is dropped from consideration, but the other two siblings have a negative influence: GreenHasOpportunity has a large influence, and GreenHasMotive has a standard influence. As indicated above, if stronger influences are present, weaker influences are not added to an interpretation, as we are trying to minimize the number of inclusions in an interpretation. Hence, only GreenHasOpportunity is added.

Finally, the interpretation fragment in Table 2(c) goes from Mr Green *possibly **not*** visiting Mr Body last night to Mr Green *very probably* having the opportunity to murder Mr Body. This is a case of an increase in belief (Step 2b) and an increase in certainty. Our algorithm considers the siblings of GreenVisitBodyLastNight, which are BodyKilledFromOutsideWindow, GreenInGardenAtTimeOfDeath and GreenMurderedBody (Figure 2), in order to find the strongest positive influences that explain this increase in belief. GreenInGardenAtTimeOfDeath has the strongest influence (standard), so it is the only node added to the interpretation.

4 Evaluation

Our evaluation of the content enhancer was conducted as follows. We constructed three evaluation sets, each consisting of two presentations of an interpretation. One of the presentations was generated using our original approach, and the other by post-processing this presentation with the content enhancer. Two of the evaluation sets were from the initial trial (Section 2) and one was new.[4] This set was added in order to evaluate all aspects of the content enhancer. For one of the evaluation sets, the content enhancer removed nodes from the interpretation that it felt contained superfluous information (Section 3.1). For the other two sets, the enhancer added influencing nodes to the interpretations (Section 3.3). One of these interpretations included a large decrease in belief and the other included a small increase in belief.

The three evaluation sets were shown to 20 subjects, including 6 of the subjects who participated in the initial trial (the other subjects of this trial were unavailable). The subjects came from several populations, which included staff and students in the School of Computer Science and Software Engineering at Monash University, and friends of the authors. The subjects belonged to several age groups and exhibited different levels of computer literacy. In our experiment, we first gave our subjects a definition and example of an interpretation, and told them that the aim of the experiment was to compare our original method for the presentation of BIAS' argument interpretations with our new method. The subjects were then shown the three evaluation sets. However, they were

[4] The interpretation for one of the original evaluation sets was not affected by the content enhancer. Also, due to modifications performed to the interpretation system since the initial trial, the interpretations generated for another evaluation set differed from the original ones.

not told which presentation was generated by the original method and which by the enhanced method, and had no knowledge of the BN used to derive the interpretations. This yielded a total of 60 judgments, where 48.3% favoured the new output, 15% were indifferent, and 36.7% favoured the old output.

For the evaluation set which had nodes removed from the presentation, the results were widely spread, suggesting that people's opinions regarding what is superfluous differ substantially, and may depend on contextual information. 30% of the trial subjects preferred the post-processed presentation, 35% were indifferent, and 35% preferred the original interpretation. 40% of the subjects felt that the original interpretation was verbose, but 25% of the subjects thought that the original interpretation was lacking in information to fully explain its reasoning path and beliefs. Also, our subjects' comments indicated that the information they found lacking from the post-processed interpretation was not necessarily related to the removed node. For the evaluation sets which had nodes added to the interpretations, the results clearly favour the enhanced presentations. 57.5% of the trial subjects preferred the new interpretations, compared to 37.5% who preferred the original ones, and 5% who were indifferent. 45% of the subjects felt that the expanded presentations were too verbose regardless of whether they preferred them or not, while 17.5% thought that the expanded presentations still lacked information. Only 7.5% thought that the original presentations were already too verbose. This indicates that the subjects preferred to know more about the system's reasoning, but still had problems with its presentation. These problems may be partially attributed to the presentation of the nodes as full canonical sentences, which makes the interpretations appear repetitive in style, and hence may have an adverse influence on acceptance.

In general, our subjects' comments point to the difficulty of conducting user trials in a commonsense domain. Our BN contains only limited domain knowledge (included by the authors), which may differ from the beliefs and expectations of our subjects. This explains why our subjects may have considered the presented information insufficient to account for the system's inferences, irrespective of the modifications made by the content enhancer. Some of the subjects also felt that for an interpretation to be acceptable, they had to make assumptions about what other information the system was basing its beliefs on, even though they had no knowledge of the structure of the BN. Future developments in the system will work on establishing these assumptions and including them in the presentations.

In addition, a limitation of our approach is that its similarity measure only approximates the similarity between the content of propositions. As a result, our system may omit propositions that appear similar to stated propositions according to our measure, but are in fact dissimilar in content. In contrast, our system retains propositions that are dissimilar in form to stated propositions, even if they convey a similar meaning. This indicates that a more sophisticated measure of propositional similarity is required to determine whether nodes may be omitted from a presentation.

5 Related Research

The mechanism presented in this paper enhances the content of discourse generated from BNs.

BNs have become pervasive in recent years. However, the explanation of Bayesian reasoning has been considered only by a few researchers [4, 5, 6]. Druzdzel [4] and McConachy *et al.* [5] studied different aspects of the presentation of BNs. Druzdzel focused on the reduction of the number of variables being considered, verbal expressions of uncertainty, and qualitative explanations, which were generated by tracing the influence of the nodes in a BN. McConachy *et al.* applied attentional models to the construction of probabilistic arguments, and studied probabilistic argumentation patterns and argumentation strategies. Jitnah *et al.* [6] extended this work by considering strategies for the presentation of rebuttals to users' arguments. Our work follows the last two contributions. However, it is worth noting that these systems generated arguments, while we present interpretations of arguments. In addition, the presentations generated by these systems hinged on discrepancies between the system's world model and the user's, while our presentations rely on the features of an interpretation itself.

Several NLG systems consider the addition or removal of information to improve planned discourse. The research reported in [7, 8, 9] considers the addition of information to planned discourse to prevent or weaken a user's erroneous inferences from this discourse. In contrast, the mechanism presented in this paper adds information to explain reasoning steps that a user may find difficult to understand. The work described in [5, 9, 10] considers the omission of information that may be inferred by a user from planned discourse. Our omission mechanism is most similar to that described in [5]. However, they used spreading activation and partial Bayesian propagation to determine whether a node may be omitted, while we use a simple word similarity measure and belief comparison.

6 Conclusion

We have offered a mechanism developed on the basis of user trials, which enhances the content of argument interpretations for presentation. This is done through the removal of superfluous information and the inclusion of information that explains unintuitive effects. Our mechanism was developed in the context of BNs. However, it is also applicable to non-Bayesian systems (provided belief is represented). Further, although our current results focus on interpretations, our procedures are also applicable to arguments.

Our evaluation of the content-enhancer shows that the post-processed presentations have a positive effect on users' acceptance of the system's interpretations, in particular in regard to the addition of information. In the near future, we propose to further refine the node-removal component of our algorithm, and to improve the node-addition component to include assumptions made by the system. We also intend to conduct additional user trials with more complex arguments.

Acknowledgments

This research is supported in part by the ARC Centre for Perceptive and Intelligent Machines in Complex Environments.

References

1. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Publishers, San Mateo, California (1988)
2. Elsaesser, C.: Explanation of probabilistic inference for decision support systems. In: Proceedings of the AAAI-87 Workshop on Uncertainty in Artificial Intelligence, Seattle, Washington (1987) 394–403
3. Salton, G., McGill, M.: An Introduction to Modern Information Retrieval. McGraw Hill (1983)
4. Druzdzel, M.: Qualitative verbal explanations in Bayesian Belief Networks. Artificial Intelligence and Simulation of Behaviour Quarterly (1996) 43–54
5. McConachy, R., Korb, K.B., Zukerman, I.: Deciding what not to say: An attentional-probabilistic approach to argument presentation. In: Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, Madison, Wisconsin (1998) 669–674
6. Jitnah, N., Zukerman, I., McConachy, R., George, S.: Towards the generation of rebuttals in a Bayesian argumentation system. In: Proceedings of the First International Natural Language Generation Conference, Mitzpe Ramon, Israel (2000) 39–46
7. Joshi, A., Webber, B.L., Weischedel, R.M.: Living up to expectations: Computing expert responses. In: AAAI84 – Proceedings of the Fourth National Conference on Artificial Intelligence, Austin, Texas (1984) 169–175
8. van Beek, P.: A model for generating better explanations. In: Proceedings of the Twenty-Fifth Annual Meeting of the Association for Computational Linguistics, Stanford, California (1987) 215–220
9. Zukerman, I., McConachy, R.: WISHFUL: A discourse planning system that considers a user's inferences. Computational Intelligence **17** (2001) 1–61
10. Horacek, H.: A model for adapting explanations to the user's likely inferences. User Modeling and User-Adapted Interaction **7** (1997) 1–55