Balancing Conflicting Factors in Argument Interpretation

Ingrid Zukerman, Michael Niemann and Sarah George
Faculty of Information Technology
Monash University
Clayton, VICTORIA 3800, AUSTRALIA
{ingrid,niemann,sarahg}@csse.monash.edu.au

Abstract

We present a probabilistic approach for the interpretation of arguments that casts the selection of an interpretation as a model selection task. In selecting the best model, our formalism balances conflicting factors: model complexity against data fit, and structure complexity against belief reasonableness. We first describe our basic formalism, which considers interpretations comprising inferential relations, and then show how our formalism is extended to suppositions that account for the beliefs in an argument, and justifications that account for the inferences in an interpretation. Our evaluations with users show that the interpretations produced by our system are acceptable, and that there is strong support for the postulated suppositions and justifications.

1 Introduction

The source-channel approach has been often used for word-based language tasks, such as speech recognition and machine translation (Epstein, 1996; Och and Ney, 2002). According to this approach, an addressee receives a noisy channel (language or speech wave), and decodes this channel to derive the source (idea). The selected source is that with the maximum posterior probability.

In this paper, we apply the source-channel approach to the interpretation of arguments. This approach enables us to cast argument interpretation as a trade-off between conflicting factors, viz model complexity against data fit, and structure complexity against belief reasonableness. This trade-off is inspired by the Minimum Message Length (MML) Criterion – a model selection method that is the basis for several machine learning techniques (Wallace, 2005). According to this trade-off, a more complex model might fit the data better, but the plausibility (priors) of the model must be taken into account to avoid over-fitting.[1]

Our argument interpretation mechanism has been implemented in a system called BIAS (Bayesian Interactive Argumentation System). BIAS presents to a user a set of facts about the world (evidence), and the user constructs an argument about a particular goal proposition in light of this evidence. BIAS then generates an interpretation of the user's argument, i.e., it tries to understand the argument. When people try to understand an interlocutor's discourse, their interpretation is in terms of their own beliefs and inference patterns. Likewise, our system's interpretations are in terms of its underlying knowledge representation – a Bayesian network (BN). The interpretations generated by BIAS include *inferences* that connect the propositions in a user's argument, *suppositions* that postulate a user's beliefs that are necessary to make sense of the argument, and *explanatory extensions* that justify the inferences in the interpretation (and in the argument). BIAS does *not* generate its own arguments, rather, it integrates these components to make sense of the user's argument.

In this paper, we first describe our basic formalism, which is used to calculate the probability of interpretations that include only inferences, and then show how progressive enhancements of this formalism are used for more informative interpretations.

In Section 2, we explain what is an argument interpretation, and describe briefly the interpretation process. Next, we discuss our probabilistic formalism for selecting an interpretation, which is the focus of this paper. In Section 4, we present

---

[1] Other model selection criteria such as Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC) (Box et al., 1994) also argue for model parsimony, but they do so by penalizing models with more free parameters.

the results of our evaluations, followed by a discussion of related work, and concluding remarks.

2 Argument interpretation

We define an interpretation of a user's argument as the tuple $\{SC, IG, EE\}$, where SC is a *supposition configuration*, IG is an *interpretation graph*, and EE are *explanatory extensions*.

- A **Supposition Configuration** is a set of suppositions attributed to the user (in addition to or instead of shared beliefs) to account for the beliefs in his or her argument.

- An **Interpretation Graph** is a domain structure, in our case a subnet of the domain BN, that connects the nodes mentioned in the argument. The nodes and arcs that are included in an interpretation graph but were not mentioned by the user fill in additional detail from the BN, bridging inferential leaps in the argument.

- **Explanatory Extensions** are domain structures (subnets of the domain BN) that are added to an interpretation graph to justify an inference. Contrary to suppositions, these explanations contain propositions believed by the user and the system. The presentation of these explanations is motivated by the results of our early trials, where people objected to *belief discontinuities* between the antecedents and the consequent of inferences, i.e., increases in certainty or large changes in certainty (Zukerman and George, 2005).

To illustrate these components, consider the example in Figure 1. The top segment contains a short argument, and the bottom segment contains its interpretation. The middle segment contains an excerpt of the domain BN which includes the interpretation; the probabilities of some nodes are indicated with linguistic terms.[2] The *interpretation graph*, which appears inside a light gray bubble in the BN excerpt, includes the extra node *GreenInGardenAtTimeOfDeath* (boxed). Note that the propagated beliefs in this interpretation graph do not match those in the argument. To address this problem, the system *supposes* that the user believes that *TimeOfDeath11*=TRUE, instead of the BN belief of *Probably* (boldfaced and

---

We use the terms *Very Probable, Probable, Possible* and their negations, and *Even Chance*. These terms, which are similar to those used in (Elsaesser, 1987), are most consistently understood by people according to our user surveys.

ARGUMENT

Mr Green *probably* being in the garden at 11 implies that he *possibly* had the opportunity to kill Mr Body, but he *possibly* did *not* murder Mr Body.

EXCERPT OF DOMAIN BN
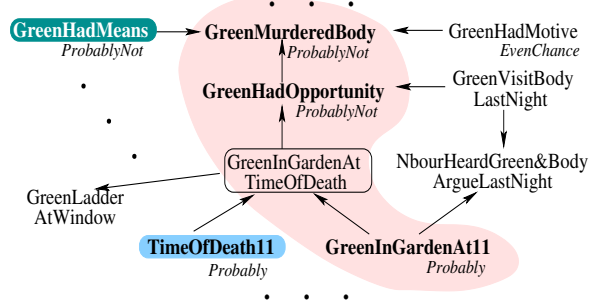
INTERPRETATION

Mr Green *probably* being in the garden at 11, and *supposing that the time of death is 11* implies that *Mr Green probably was in the garden at the time of death.* Hence, he *possibly* had the opportunity to kill Mr Body, but *Mr Green probably did not have the means.* Therefore, he *possibly* did *not* murder Mr Body.

Figure 1: Sample argument, BN excerpt and interpretation

gray-boxed). This fixes the mismatch between the probabilities in the argument and those in the interpretation, but one problem remains: in early trials we found that people objected to belief discontinuities, such as the "jump in belief" from *possibly* having opportunity to *possibly not* murdering Mr Body (this jump appears both in the original argument and in the interpretation, whose beliefs now match those in the argument as a result of the supposition). This prompts the generation of the explanatory extension *GreenHadMeans[ProbablyNot]* (white boldfaced and dark-gray boxed). The three elements added during the interpretation process – the extra node in the interpretation graph, the supposition and the explanatory extension – appear in boldface italics in the interpretation at the bottom of the figure.

2.1 Proposing Interpretations

The problem of finding the best interpretation is exponential. In previous work, we proposed an *anytime* algorithm to propose interpretation graphs and supposition configurations until time runs out (George et al., 2004). Here we apply our algorithm to generate interpretations comprising supposition configurations (SC), interpretation graphs (IG) and explanatory extensions (EE) (Figure 2).

Supposition configurations are proposed first, as instantiated beliefs affect the plausibility of inter-

135

Algorithm *GenerateInterpretations(Arg)*

while {there is time}

{

  1. Propose a supposition configuration SC that accounts for the beliefs stated in the argument.

  2. Propose an interpretation graph IG that connects the nodes in *Arg* under supposition configuration SC.

  3. Propose explanatory extensions EE for interpretation graph IG under supposition configuration SC if necessary.

  4. Calculate the probability of interpretation $\{SC, IG, EE\}$.

  5. Retain the top N (=6) most probable interpretations.

}

Figure 2: Anytime algorithm for generating interpretations

pretation graphs, which in turn affect the need for explanatory extensions. The proposal of supposition configurations, interpretation graphs and explanatory extensions is driven by the probability of these components. In each iteration, we generate candidates for a component, calculate the probability of these candidates *in the context of the selections made in the previous steps*, and probabilistically select one of these candidates. That is, higher probability candidates have a better chance of being selected than lower probability ones (our selection procedures are described in George et al., 2004). For example, say that in Step 1, we selected supposition configuration SC_a. Next, in Step 2, the probability of candidate IGs is calculated *in the context of the domain BN and SC_a*, and one of the IGs is probabilistically selected, say IG_b. Similarly, in Step 3, one of the candidate EEs is selected *in the context of SC_a and IG_b*. In the next iteration, we probabilistically select an SC (which could be a previously chosen one), and so on. To generate diverse interpretations, if SC_a is selected again, a different IG will be chosen.

3 Probabilistic formalism

Following (Wallace, 2005), our approach requires the specification of three elements: *background knowledge, model* and *data*. **Background knowledge** is everything known to the system prior to interpreting a user's argument, e.g., domain knowledge, shared beliefs with the user, and dialogue

history; the **data** is the argument; and the **model** is the interpretation.

We posit that the best interpretation is that with the highest posterior probability.

$IntBest = \text{argmax}_{i=1,\ldots,q}\Pr(SC_i, IG_i, EE_i | Arg)$

where q is the number of interpretations.

After applying Bayes rule, this probability is represented as follows.[3]

$$\Pr(SC_i, IG_i, EE_i | Arg) = \qquad (1)$$
$$\alpha \Pr(SC_i, IG_i, EE_i) \times \Pr(Arg | SC_i, IG_i, EE_i)$$

where α is a normalizing constant that ensures that the probabilities of the interpretations sum to 1 $\left(\alpha = \frac{1}{\sum_{j=1}^{n}\Pr(SC_j, IG_j, EE_j)\times\Pr(Arg|SC_j, IG_j, EE_j)}\right)$.

The first factor represents *model complexity*, and the second factor represents *data fit*.

- Model complexity measures how difficult it is to produce the model (interpretation) from the background knowledge. The higher/lower the complexity of a model, the lower/higher its probability.

- Data fit measures how well the data (argument) matches the model (interpretation). The better/worse the match between the argument and an interpretation, the higher/lower the probability that the speaker intended this interpretation when he or she uttered the argument.

Model Complexity

Model complexity is a function $\{\mathcal{B}, \mathcal{M}\} \rightarrow [0, 1]$ that represents the prior probability of the model $\mathcal{M}$ (i.e., the interpretation) in terms of the background knowledge $\mathcal{B}$. The calculation of model complexity depends on the type of the model: numerical or structural.

The probability of a ***numerical model*** depends on the similarity between the numerical values (or distributions) in the model and those in the background knowledge. The higher/lower this similarity, the higher/lower the probability of the model. For instance, a supposition configuration SC comprising beliefs that differ significantly from those in the background knowledge will lower the probability of an interpretation. One of the functions we have used to calculate belief probabilities is the Zipf distribution, where the parameter is the difference between beliefs, e.g., between the supposed

---

[3]In principle, $\Pr(SC_i, IG_i, EE_i | Arg)$ can be calculated directly. However, it is not clear how to incorporate the priors of an interpretation in the direct calculation.

beliefs and the corresponding beliefs in the background knowledge (Zukerman and George, 2005). That is, the probability of a supposed belief in proposition P according to model $\mathcal{M}$ (bel$_M(P)$), in light of the belief in P according to background knowledge $\mathcal{B}$ (bel$_B(P)$), is

$$\Pr(\text{bel}_M(P)|\text{bel}_B(P)) = \frac{\theta}{|\text{bel}_M(P) - \text{bel}_B(P)|^\gamma}$$

where θ is a normalizing constant, and γ determines the penalty assigned to the discrepancy between the beliefs in P. For example,
$\Pr(\text{bel}_M(P) = \text{TRUE}|\text{bel}_B(P) = Probable) >$
$\Pr(\text{bel}_M(P) = \text{TRUE}|\text{bel}_B(P) = EvenChance)$
as TRUE is closer to *Probable* than to *EvenChance*.

The probability of a ***structural model*** (e.g., an interpretation graph) is obtained from the probabilities of the elements in the structure (e.g., nodes and arcs) in light of the background knowledge. The simplest calculation assumes that the probability of including nodes and arcs in an interpretation graph is uniform. That is, the probability of an interpretation graph comprising n nodes and a arcs is a function of

- the probability of n,
- the probability of selecting n particular nodes from N nodes in the domain BN: $\binom{N}{n}^{-1}$,
- the probability of a, and
- the probability of selecting a particular arcs from the arcs that connect the n selected nodes.

This calculation generally prefers small models to larger models.[4]

Data fit

Data fit is a function $\{\mathcal{M}, \mathcal{D}\} \rightarrow [0, 1]$ that represents the probability of the data $\mathcal{D}$ (argument) given the model $\mathcal{M}$ (interpretation). This probability hinges on the similarity between the model and the data – the closer the data is to the model, the higher is the probability of the data.

The calculation of the similarity between ***numerical data*** and a numerical model is the same as the calculation of the similarity between a numerical model and background knowledge.

The similarity between ***structural data*** and a structural model is a function of the number and type of operations required to convert the model into the data, e.g., node and arc insertions and

---

[4]In the rare cases where $n > N/2$, smaller models do not yield lower probabilities.

deletions. For the example in Figure 1, to convert the interpretation graph into the argument, we must delete one node (*GreenInGardenAtTimeOfDeath*) and its incident arcs. The more operations need to be performed, the lower the similarity between the data and the model, and the lower the probability of the data given the model.

We now discuss our basic probabilistic formalism, which accounts for interpretation graphs, followed by two enhancements: (1) a more complex model that accounts for suppositions; and (2) increases in background knowledge that yield a preference for larger interpretation graphs under certain circumstances, and account for explanatory extensions.

3.1 Basic formalism: Interpretation graphs

In the basic formalism, the model contains only an interpretation graph. Thus, Equation 1 is simply

$$\Pr(IG_i|Arg) = \alpha \Pr(IG_i) \times \Pr(Arg|IG_i) \quad (2)$$

The difference in the calculations of model complexity and data fit for numerical and structural information warrants the separation of structure and belief, which yields

$$\Pr(IG_i|Arg) = \alpha \Pr(\text{bel } IG_i, \text{struc } IG_i) \times$$
$$\Pr(\text{bel } Arg, \text{struc } Arg|\text{bel } IG_i, \text{struc } IG_i)$$

After applying the chain rule of probability

$$\Pr(IG_i|Arg) =$$
$$\alpha \Pr(\text{bel } IG_i|\text{struc } IG_i) \times \Pr(\text{struc } IG_i) \times$$
$$\Pr(\text{bel } Arg|\text{struc } Arg, \text{bel } IG_i, \text{struc } IG_i) \times$$
$$\Pr(\text{struc } Arg|\text{bel } IG_i, \text{struc } IG_i)$$

Note that $\Pr(\text{bel } IG_i|\text{struc } IG_i)$ does *not* calculate the probability of (or belief in) the nodes in IG_i. Rather, it calculates how probable are these beliefs in light of the structure of IG_i and the expectations from the background knowledge. For instance, if the belief in a node is p, it calculates the probability of p. This probability depends on the closeness between the beliefs in IG_i and the expected ones. Since the beliefs in IG_i are obtained algorithmically by means of Bayesian propagation from the background knowledge, they match precisely the expectations. Hence, $\Pr(\text{bel } IG_i|\text{struc } IG_i) = 1$.

We also make the following simplifying assumptions for situations where the interpretation is known (given): (1) the probability of the beliefs in the argument depends only on the beliefs in the

Table 1: Probability – Basic formalism	
Model complexity (against background)	
$\downarrow \Pr(\text{struc } IG_i)$	$\uparrow$ structural complexity (model size)
Data fit with model	
$\uparrow \Pr(\text{struc } Arg \mid \text{struc } IG_i)$	$\downarrow$ structural discrepancy
$\Pr(\text{bel } Arg \mid \text{bel } IG_i)$	numerical discrepancy

Table 2: Probability – More informed model	
Model complexity (against background)	
$\Pr(\text{struc } IG_i \mid \mathbf{SC_i})$	structural complexity
$\downarrow \Pr(\mathbf{SC_i})$	$\uparrow$ numerical discrepancy
Data fit with model	
$\Pr(\text{struc } Arg \mid \text{struc } IG_i)$	structural discrepancy
$\uparrow \Pr(\text{bel } Arg \mid \mathbf{SC_i}, \text{bel } IG_i)$	$\downarrow$ numerical discrepancy

interpretation (and not on its structure or the argument's structure), and (2) the probability of the argument structure depends only on the interpretation structure (and not on its beliefs). This yields

$$\Pr(IG_i \mid Arg) = \alpha \Pr(\text{struc } IG_i) \times \quad (3)$$
$$\Pr(\text{bel } Arg \mid \text{bel } IG_i) \times \Pr(\text{struc } Arg \mid \text{struc } IG_i)$$

Table 1 summarizes the calculation of these probabilities separated according to model complexity and data fit. It also shows the trade-off between structural model complexity and structural data fit. As seen at the start of Section 3, smaller structures generally have a lower model complexity than larger ones. However, an increase in structural model complexity (indicated by the $\uparrow$ next to the structural complexity and the $\downarrow$ next to the resultant probability of the model) may reduce the structural discrepancy between the argument structure and the structure of the interpretation graph (indicated by the $\downarrow$ next to the structural discrepancy and the $\uparrow$ next to the probability of the structural data-fit). For instance, the smallest possible interpretation for the argument in Figure 1 consists of a single node, but this interpretation has a very poor data fit with the argument.

3.2 A more informed model

In order to postulate suppositions that account for the beliefs in an argument, we expand the basic model to include supposition configurations (beliefs attributed to the user in addition to or instead of the beliefs shared with the system). Now the model comprises the pair $\{SC_i, IG_i\}$, and Equation 2 becomes

$$\Pr(SC_i, IG_i \mid Arg) = \quad (4)$$
$$\alpha \Pr(SC_i, IG_i) \times \Pr(Arg \mid SC_i, IG_i)$$

Similar probabilistic manipulations to those performed in Section 3.1 yield

$$\Pr(SC_i, IG_i \mid Arg) = \quad (5)$$
$$\alpha \Pr(\text{struc } IG_i \mid SC_i) \times \Pr(SC_i) \times$$
$$\Pr(\text{bel } Arg \mid SC_i, \text{bel } IG_i) \times \Pr(\text{struc } Arg \mid \text{struc } IG_i)$$

(Recall that suppositions pertain to beliefs only, i.e., they don't have a structural component.)

Table 2 summarizes the calculation of these probabilities separated according to model complexity and data fit (the elements that differ from the basic model are **boldfaced**). It also shows the trade-off between belief model complexity and belief data fit. Making suppositions has a higher model complexity (lower probability) than not making suppositions (where SC_i matches the beliefs in the domain BN). However, as seen in the example in Figure 1, making a supposition that reduces or eliminates the discrepancy between the beliefs in the argument and those in the interpretation increases the belief data-fit considerably, at the expense of a more complex belief model.

3.3 Additional background knowledge

An increase in our background knowledge means that we take into account additional factors about the world. This extra knowledge in turn may cause us to prefer interpretations that were previously discarded. We have considered two additions to background knowledge: dialogue history, and users' preferences regarding inference patterns.

Dialogue history

Dialogue history influences the salience of a node, and hence the probability that it was included in a user's argument. We have modeled salience by means of an activation function that decays with time (Anderson, 1983), and used this function to moderate the probability of including a node in an interpretation (instead of using a uniform distribution). We have experimented with two activation functions: (1) a function where the level of activation of a node is based on the frequency and recency of the direct activation of this node; and (2) a function where the level of activation of a node depends on its similarity with all the (activated) nodes, together with the frequency and recency of their activation (Zukerman and George,

2005).

To illustrate the influence of salience, compare the preferred interpretation graph in Figure 1 (in the light gray bubble) with an alternative path through *NbourHeard-Green&BodyArgueLastNight* and *GreenVisit-BodyLastNight*. The preferred path has 4 nodes, while the alternative one has 5 nodes, and hence a lower probability. However, if the nodes in the longer path had been recently mentioned, their salience could overcome the size disadvantage. Thus, although the chosen interpretation graph may have a worse data fit than the smallest graph, it still may have the best overall probability in light of the additional background knowledge.

Inference patterns

In a formative evaluation of an earlier version of our system, we found that people objected to inferences that had increases in certainty or large changes in certainty (Zukerman and George, 2005). An example of an increase in certainty is
A [Probably] implies B [VeryProbably].
A large change in certainty is illustrated by
A [VeryProbably] implies B [EvenChance].

We then conducted another survey to determine the types of inferences considered acceptable by people (from the standpoint of the beliefs in the antecedents and the consequent). The results from our preliminary survey prompted us to distinguish between three types of inferences: *BothSides, SameSide* and *AlmostSame*.

- *BothSides* inferences have antecedents with beliefs on both "sides" of the consequent (in favour and against), e.g.,
  A[VeryProbably] & B[ProbablyNot] implies C[EvenChance].
- All the antecedents in *SameSide* inferences have beliefs on "one side" of the consequent, but at least one antecedent has the same belief level as the consequent, e.g.,
  A[VeryProbably] & B[Possibly] implies C[Possibly].
- All the antecedents in *AlmostSame* inferences have beliefs on one side of the consequent, but the closest antecedent is one level "up" from the consequent, e.g.,
  A[VeryProbably] & B[Possibly] implies C[EvenChance].

Our survey contained six evaluation sets, which were done by 50 people. Each set contained an ini-tial statement (we varied the polarity of the statement in the various sets), three alternative arguments that explain this statement, and the option to say that no argument is a good explanation. The respondents were asked to rank these options in order of preference.

All the evaluation sets contained one argument that was objectionable according to our preliminary survey (there was an increase in belief or a large change in belief from the antecedent to the consequent). The two other arguments, each of which comprises a single inference, were distributed among the six evaluation sets as follows.

- Three sets had one *BothSides* inference and one *SameSide* inference, each with two antecedents.
- Two sets had one *SameSide* inference, and one *AlmostSame* inference, each with two antecedents.
- One set had one *SameSide* inference with two antecedents, and one *BothSides* inference comprising three antecedents.

In order to reduce the effect of the respondents' domain bias, we generated two versions of the survey, where for each evaluation set we swapped the antecedent propositions in one of the inferences with the antecedent propositions in the other.

Our survey showed that people prefer *BothSides* inferences (which contain antecedents for and against the consequent). They also prefer *SameSide* to *AlmostSame* for antecedents with beliefs in the negative range (*VeryProbNot, ProbNot* and *PossNot*); and they did not distinguish between *SameSide* and *AlmostSame* for antecedents with beliefs in the positive range. Further, *BothSides* inferences with three antecedents were preferred to *SameSide* inferences with two antecedents. This indicates that persuasiveness carries more weight than parsimony.

These general preferences are incorporated into our background knowledge as expectations for a range of acceptable beliefs in the consequents of inferences in light of their antecedents. The farther the actual beliefs in the consequents are from the expectations, the lower the probability of these beliefs. Hence, it is no longer true that $\Pr(\text{bel } IG_i | SC_i, \text{struc } IG_i) = 1$ (Section 3.1), as we now have a belief expectation that goes beyond Bayesian propagation. As done at the start of Section 3, the probability of the beliefs in an interpretation is a function of the discrepancy between

these beliefs and expected beliefs. We calculate this probability using a variant of the Zipf distribution adjusted for ranges of beliefs.

Explanatory extensions are added to an interpretation in order to overcome these belief discrepancies, yielding an expanded model that comprises the tuple $\{SC_i, IG_i, EE_i\}$. Equation 2 now becomes

$$\Pr(SC_i, IG_i, EE_i|Arg) = \qquad (6)$$
$$\alpha \Pr(SC_i, IG_i, EE_i) \times \Pr(Arg|SC_i, IG_i, EE_i)$$

We make simplifying assumptions similar to those made in Section 3.1, i.e., given the interpretation graph and supposition configuration, the beliefs in the argument depend only on the beliefs in the interpretation, and the argument structure depends only on the interpretation structure. These assumptions, together with probabilistic manipulations similar to those performed in Section 3.1, yield

$$\Pr(SC_i, IG_i, EE_i|Arg) = \qquad (7)$$
$$\alpha \Pr(\text{struc } IG_i|SC_i) \times \Pr(SC_i) \times$$
$$\Pr(\text{bel } IG_i|SC_i, \text{struc } IG_i, \text{bel } EE_i, \text{struc } EE_i) \times$$
$$\Pr(\text{struc } EE_i|SC_i, \text{struc } IG_i, \text{bel } EE_i) \times$$
$$\Pr(\text{bel } EE_i|SC_i, \text{struc } IG_i, \text{struc } EE_i) \times$$
$$\Pr(\text{bel } Arg|SC_i, \text{bel } IG_i) \times \Pr(\text{struc } Arg|\text{struc } IG_i)$$

The calculation of the probability of an explanatory extension is the same as the calculation for structural model complexity at the start of Section 3. However, the nodes in an explanatory extension are selected from the nodes directly connected to the interpretation graph. In addition, as for the basic model (Section 3.1), the beliefs in the nodes in explanatory extensions are obtained algorithmically by means of Bayesian propagation. Hence, there is no discrepancy with expected beliefs, i.e., $\Pr(\text{bel } EE_i|SC_i, \text{struc } IG_i, \text{struc } EE_i) = 1$.

Table 3 summarizes the calculation of these probabilities (the elements that differ from the basic model and the enhanced model are **boldfaced**). It also shows the trade-off between structural and belief model complexity. Presenting explanatory extensions has a higher structural complexity (lower probability) than not presenting them. However, explanatory extensions can reduce the numerical discrepancy between the beliefs in an interpretation and the beliefs expected from the background knowledge, thereby increasing the belief probability of the interpretation. For instance,

Table 3: Probability – Additional background knowledge

Model complexity (against background)	
$\Pr(\text{struc } IG_i|SC_i)$	structural complexity
$\Pr(SC_i)$	numerical discrepancy
$\downarrow \Pr(\text{struc } \mathbf{EE_i}|\mathbf{SC_i}, \text{struc } \mathbf{IG_i}, \text{bel } \mathbf{EE_i})$	$\uparrow$ structural complexity
$\uparrow \Pr(\text{bel } IG_i|SC_i, \text{struc } IG_i, \text{bel } \mathbf{EE_i}, \text{struc } \mathbf{EE_i})$	$\downarrow$ numerical discrepancy
Data fit with model	
$\Pr(\text{struc } Arg|\text{struc } IG_i)$	structural discrepancy
$\Pr(\text{bel } Arg|SC_i, \text{bel } IG_i)$	numerical discrepancy

Table 4: Summary of Trade-offs

$\downarrow$ Pr model structure (IG)	$\Rightarrow \uparrow$ Pr struct. data fit
$\downarrow$ Pr model belief (SC)	$\Rightarrow \uparrow$ Pr belief data fit
$\downarrow$ Pr model structure (EE)	$\Rightarrow \uparrow$ Pr model belief

in the example in Figure 1, the added explanatory extension eliminates the unacceptable jump in belief.

Table 4 summarizes the trade-offs discussed in this section.

4 Evaluation

We evaluated separately each component of an interpretation – interpretation graph, supposition configuration and explanatory extensions.

4.1 Interpretation graph

We prepared four evaluation sets, each of which was done by about 20 people (Zukerman and George, 2005). In three of the sets, the participants were given a simple argument and a few candidate interpretations (ranked highly by our system). The fourth set featured a complex argument, and only one interpretation (other candidates had much lower probabilities). The participants were asked to give each interpretation a score between 1 (Very UNreasonable) and 5 (Very reasonable). Table 5 shows the results obtained for the interpretation selected by our formalism for each set, which was the top scoring interpretation. The first

Table 5: Evaluation results: Interpretation graph

Set #	1	2	3	4
Avg. score	3.38	3.68	3.35	4.00
Std. dev.	1.45	1.11	1.39	1.02
Stat. sig. (p)	0.08	0.15	0.07	NA

140

row shows the average score given by our subjects to this interpretation, the second row shows the standard deviation, and the third row the statistical significance, derived using a paired Z-test against alternative options (no alternatives were presented for the fourth set). Our results show that the interpretations generated by our system were generally acceptable, but that some people gave low scores. Our subjects' feedback indicated that these scores were mainly due to mismatches between beliefs in the argument and in its interpretation, and due to belief discontinuities. This led to the addition of suppositions and explanatory extensions.

4.2 Supposition configuration

We prepared four evaluation sets, each of which was done by 34 people (George et al., 2005). Each set consisted of a short argument, plus a list of supposition options as follows: (a) four suppositions that had a reasonably high probability according to our formalism, (b) the option to make a free-form supposition in line with the domain BN, and (c) the option to suppose nothing. We then asked our subjects to indicate which of these options was required for the argument to make sense. Specifically, they had to rank their preferred options in order of preference (but they did not have to rank options they disliked). Overall, there was strong support for the supposition preferred by our formalism. In three of the evaluation sets, it was ranked first by most of the trial subjects (30/34, 19/34, 20/34), with no other option a clear second. Only in the fourth set, the supposition preferred by our formalism was equal-first with another option, but still was ranked first 10 times (out of 34).

4.3 Explanatory extensions

We constructed two evaluation sets, each of which was done by 20 people. Each set consisted of a short argument and two alternative interpretations (with and without explanatory extensions). There was strong support for the explanatory extensions proposed by our formalism, with 57.5% of our trial subjects favouring the interpretations with explanatory extensions, compared to 37.5% of the subjects who preferred the interpretations without such extensions, and 5% who were indifferent.

5 Related Research

An important aspect of discourse understanding involves filling in information that was omitted by the interlocutor. In this paper, we have presented a probabilistic formalism that balances conflicting factors when filling in three types of information omitted from an argument. Interpretation graphs fill in details in the argument's inferences, supposition configurations make sense of the beliefs in the argument, and explanatory extensions overcome belief discontinuities.

Our approach resembles the work of Hobbs et al. (1993) in several respects. They employed an abductive approach where a model (interpretation) is inferred from evidence (sentence); they made assumptions as necessary; and used guiding criteria pertaining to the model and the data for choosing between candidate models. There are also significant differences between our work and theirs. Their interpretation focused on problems of reference and disambiguation in single sentences, while ours focuses on a longer discourse and the relations between the propositions therein. This distinction also determines the nature of the task, as they try to find a concise model that explains as much of the data as possible (e.g., one referent that fits many clues), while we try to find a representation for a user's argument. Additionally, their domain knowledge is logic-based, while ours is Bayesian; and they used weights to apply their hypothesis selection criteria, while our criteria are embodied in a probabilistic framework.

Plan recognition systems also generate one or more interpretations of a user's utterances, employing different resources to fill in information omitted by the user, e.g., (Allen and Perrault, 1980; Litman and Allen, 1987; Carberry and Lambert, 1999; Raskutti and Zukerman, 1991). These plan recognition systems used a plan-based approach to propose interpretations. The first three systems applied different types of heuristics to select an interpretation, while the fourth system used a probabilistic approach moderated by heuristics to select the interpretation with the highest probability. We use a probabilistic domain representation in the form of a BN (rather than plan libraries), and apply a probabilistic mechanism that represents explicitly the contribution of background knowledge, model complexity and data fit to the generation of an interpretation. Our mechanism, which can be applied to other domain representations, balances different types of complexities and discrepancies to select the interpretation with the highest posterior probability.

Several researchers used maximum posterior

probability as the criterion for selecting an interpretation (Charniak and Goldman, 1993; Gertner et al., 1998; Horvitz and Paek, 1999). They used BNs to represent a probability distribution over the set of possible explanations for the observed facts, and selected the explanation (a node in the BN or a value of a node) with the highest probability. We also use BNs as our domain representation, but our "explanation" of the facts (the user's argument) is a Bayesian subnet (rather than a single node) supplemented by suppositions. Additionally, we calculate the probability of an interpretation on the basis of the fit between the argument and the interpretation, and the complexity of the interpretation in light of the background knowledge.

Our work on positing suppositions is related to research on presuppositions (Kaplan, 1982; Gurney et al., 1997) – a type of supposition implied by the wording of a statement. Like our suppositions, presuppositions are necessary to make sense of what is being said, but they operate at a different knowledge level than our suppositions. This aspect of our work is also related to research on the recognition of flawed plans (Quilici, 1989; Pollack, 1990; Chu-Carroll and Carberry, 2000). These researchers used a plan-based approach to identify erroneous beliefs that account for a user's statements or plan, while we use a probabilistic approach. Our approach supports the consideration of many possible options, and integrates suppositions into a broader reasoning context.

Finally, the research reported in (Joshi et al., 1984; van Beek, 1987; Zukerman and McConachy, 2001) considers the addition of information to planned discourse to prevent a user's erroneous inferences from this discourse. Our mechanism adds explanatory extensions to an interpretation to prevent inferences that are objectionable due to discontinuities in belief. Since such non-sequiturs may also be present in system-generated arguments, the approach presented here may be incorporated into argument-generation systems.

6   Conclusion

We have offered a probabilistic approach to the interpretation of arguments that casts the selection of an interpretation as a model selection task. In so doing, our formalism balances conflicting factors: model complexity against data fit, and structure complexity against belief reasonableness. We have demonstrated the use of our basic formalism

for the selection of an interpretation graph, and shown how a more complex model and additional background knowledge account respectively for the inclusion of suppositions and explanatory extensions in an interpretation. Our user evaluations show that the interpretation graphs produced by our formalism are generally acceptable, and that there is strong support for the suppositions and explanatory extensions it proposes.

References

J.F. Allen and C.R. Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15(3):143–178.

J. R. Anderson. 1983. *The Architecture of Cognition*. Harvard University Press, Cambridge, Massachusetts.

G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. Prentice Hall.

S. Carberry and L. Lambert. 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, 25(1):1–53.

E. Charniak and R. Goldman. 1993. A Bayesian model of plan recognition. *Artificial Intelligence*, 64(1):53–79.

J. Chu-Carroll and S. Carberry. 2000. Conflict resolution in collaborative planning dialogues. *International Journal of Human Computer Studies*, 6(56):969–1015.

C. Elsaesser. 1987. Explanation of probabilistic inference for decision support systems. In *Proceedings of the AAAI-87 Workshop on Uncertainty in Artificial Intelligence*, pages 394–403, Seattle, Washington.

M.E. Epstein. 1996. *Statistical Source Channel Models for Natural Language Understanding*. Ph.D. thesis, Department of Computer Science, New York University, New York, New York.

S. George, I. Zukerman, and M. Niemann. 2004. An anytime algorithm for interpreting arguments. In *PRICAI2004 – Proceedings of the Eighth Pacific Rim International Conference on Artificial Intelligence*, pages 311–321, Auckland, New Zealand.

S. George, I. Zukerman, and M. Niemann. 2005. Modeling suppositions in users' arguments. In *UM05 – Proceedings of the 10th International Conference on User Modeling*, pages 19–29, Edinburgh, Scotland.

A. Gertner, C. Conati, and K. 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.

J. Gurney, D. Perlis, and K. Purang. 1997. Interpreting presuppositions using active logic: From contexts to utterances. *Computational Intelligence*, 13(3):391–413.

J. R. Hobbs, M. E. Stickel, D. E. Appelt, and P. Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.

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

A. Joshi, B. L. Webber, and R. M. Weischedel. 1984. Living up to expectations: Computing expert responses. In *AAAI84 – Proceedings of the Fourth National Conference on Artificial Intelligence*, pages 169–175, Austin, Texas.

S. J. Kaplan. 1982. Cooperative responses from a portable natural language query system. *Artificial Intelligence*, 19:165–187.

D. Litman and J.F. Allen. 1987. A plan recognition model for subdialogues in conversation. *Cognitive Science*, 11(2):163–200.

F.J. Och and H. Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In *ACL'02 – Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 295–302, Philadelphia, Pennsylvania.

M.E. Pollack. 1990. Plans as complex mental attitudes. In P. Cohen, J. Morgan, and M.E. Pollack, editors, *Intentions in Communication*, pages 77–103. MIT Press.

A. Quilici. 1989. Detecting and responding to plan-oriented misconceptions. In A. Kobsa and W. Wahlster, editors, *User Models in Dialog Systems*, pages 108–132. Springer-Verlag.

B. Raskutti and I. Zukerman. 1991. Generation and selection of likely interpretations during plan recognition. *User Modeling and User Adapted Interaction*, 1(4):323–353.

P. van Beek. 1987. A model for generating better explanations. In *Proceedings of the Twenty-Fifth Annual Meeting of the Association for Computational Linguistics*, pages 215–220, Stanford, California.

C.S. Wallace. 2005. *Statistical and Inductive Inference by Minimum Message Length*. Springer, Berlin, Germany.

I. Zukerman and S. George. 2005. A probabilistic approach for argument interpretation. *User Modeling and User-Adapted Interaction, Special Issue on Language-Based Interaction*, 15(1-2):5–53.

I. Zukerman and R. McConachy. 2001. WISHFUL: A discourse planning system that considers a user's inferences. *Computational Intelligence*, 1(17):1–61.

143