

Implementing a Model of Human Plausible Reasoning

Michelle Baker
Columbia University, Department of Computer Science

Mark H. Burstein
Allan M. Collins
BBN Laboratories Inc.

Abstract

This paper describes the current state of implementation of a cognitive computer model of human plausible reasoning, based on the theory of plausible reasoning described by Collins and Michalski. Our goal is to use the simulation as a means of testing and refining the theory. This requires developing appropriate memory organization and search techniques to support of this style of inference, finding ways to estimate similarity in specific contexts and investigating ways of combining the sometimes contradictory conclusions reached when inferences of different types are used to answer questions.

1. INTRODUCTION

Over the last 15 years, Collins and his colleagues (Carbonell and Collins, 1973, Collins et al., 1975, Collins, 978a, Collins, 978b) have collected and categorized a wide variety of human plausible inferences made from incomplete and inconsistent information. This work led to the development of a partial theory of plausible inference (Collins and Michalski, in press) for situations where the most appropriate or specific information was not available. This paper describes some current work in progress, the development of a computer simulation of a portion of that theory. Our goal is to use the simulation as a means of testing and refining the theory.

The popularity of expert systems has generated great interest in developing techniques to reason with uncertain information. To date, research on reasoning under uncertainty has emphasized the role of statistical theory. (Pearl, 1986, Duda et al., 1976). Unfortunately, in most real-world problems neither the data nor the inference rules themselves are known to apply with precise certainties. Methods of combining uncertain evidence from multiple sources are also often required. With the exception of Cohen (Cohen, 1985), it has usually been assumed that the appropriate certainty parameters and the methods of combination were independent of the type of inference performed. Furthermore, these techniques usually require some form of closed world assumption for correct interpretation. Unfortunately, in most real-world situations, the available information is incomplete as well as uncertain. People deal with this problem continually, and quite effectively, using techniques for reasoning by similarity, reasoning from negative information, and reasoning from their own lack of knowledge about particulars (e.g., "I would know it if Ronald Reagan was 10 feet tall.") It is these kinds of inferences that we seek to model.

Collins' theory of plausible reasoning is based on a corpus of people's answers to everyday questions (Collins, 978b). In general, he found that these answers had the following characteristics:

1. There are usually several different inference types used to answer any question.
2. The same inference types recur in many different answers.

3. People weigh different evidence (and different kinds of evidence) they find that bears on a question.
4. People are more or less certain depending on the certainty of their information, the certainty of the inferences used, and on whether different inferences lead to the same or opposite conclusions.

Also apparent from the protocols is that subjects faced with answering a question for which they have no specific knowledge launch a search for relevant information that they do have. As relevant pieces of information are found (or are found to be missing), they trigger particular types of inferences. The type of inference applied is determined by the relation between the information found and the question asked. For example, when a tutor was asked whether they grow coffee in the Llanos region of Colombia, he responded:

I don't think that the savanna is used for growing coffee. The trouble is the savanna has a rainy season and you can't count on rain in general. But I don't know, this area around Sao Paulo (in Brazil) is coffee region, and it is sort of getting into the savanna region there.

Initially, the tutor said no because he knew that coffee growing depends on factors like rainfall, temperature, soil, etc. and the savannas do not have the correct value on the rainfall factor. (This is called a derivation from mutual implication in the theory.) Secondly, he did not know specifically that the Llanos was used for coffee growing, and believed that he would know if it was (lack of knowledge). Later, he backed off when he found positive evidence; i.e., that the region in Brazil was near an area where coffee was grown (a similarity transform). His final answer weighed all of these pieces of evidence together, albeit inexactly.

In the remainder of this paper, we will describe an initial implementation of one part of Collins' theory of plausible reasoning, based on examples like this one. Initially, we have concentrated on modeling the class of functional inferences, where the inference is based on a functional dependence such as that coffee growing depends on climate and vegetation.

The primary purpose of the system is to act as a testbed for the theory. As such, it is not designed to produce one "right" answer, but a number of plausible positive and negative inferences each of which may be a weak (or not so weak) reason for believing that the question asked could be answered in a particular way. Our goals are to demonstrate that the theory produces only plausible answers, to develop ways of searching memory for the kinds of relevant information that are needed in order to apply each inference type, and to investigate methods for combining the various kinds of evidence that are produced.

The Plausible Reasoning Simulation System (PRSS) we have developed is thus quite different from other systems that have been developed to reason with incomplete and/or uncertain information. Since it is intended to simulate human reasoning, it generates multiple proofs of both the truth and the falsity of a given proposition. The types of inferences made depend on the particular information found in memory, and the nature of their relevance to the question asked. Finally, the certainty of the overall conclusions

¹This work was done at BBN Laboratories, and sponsored by ARI under Contract number MDA9038SC0411.

reached depends on both the certainty of the evidence and the types of inferences used.

2. AN EXAMPLE

To give a sense of the behavior of the simulation system, consider how it behaves when asked a question like "Does coffee grow in Llanos?".

(? crop :of llanos := coffee)

HO DIRECT EVIDENCE FOUND.

TRYING NEGATIVE IMPLICATION FROM

CROP - COFFEE \implies RAINFALL - HIGH
(certainty .8)

Since HIGH is not a known value for RAINFALL(LLANOS), and set of values for RAINFALL(LLANOS) is CLOSED.

Conclude that COFFEE is not a value for CROP(LLANOS) with MEDIUM certainty.

TRYING ARGUMENT BASED DEPENDENCY TRANSFORMS

LLANOS and SAOPALLO match on CLIMATE
(aim -0.8)

LLANOS and SAOPALLO match on VEGETATION.
(aim -0.6)

Using a SIM transform:

Since CLIMATE and VEGETATION \iff CROP and SAOPALLO is similar to LLANOS with respect to CLIMATE and VEGETATION (aim -0.7)

and CROPSAOPALLO - COFFEE
Conclude that CROP(LLANOS) - COFFEE is TRUE with MEDIUM certainty.

Evidence is evenly mixed. I cannot make a judgement.

For this example, PRSS finds two kinds of evidence. First, it reasons from the implication that coffee growing requires heavy rainfall, and from the fact that it does not believe the Llanos to have heavy rainfall, that the Llanos is not a coffee growing region. This conclusion is given medium certainty primarily because of the certainty of the implication. Secondly, it finds that the SAO-PAULO region does have coffee as a crop and matches Llanos on CLIMATE and VEGETATION, two variables involved in a mutual dependency with CROP. Since the evidence is evenly divided, no final conclusion is reached.

3. SYSTEM OVERVIEW

Unlike an expert system, which must generate a solution, PRSS tries both to verify and disconfirm each proposition that it is given as an input question. Some examples of the kinds of queries the system may receive as input are:

(? CLIMATE :OF ENGLAND :- TEMPERATE)
(? FLOWERTYPE :OF HOLLAND :- ROSE)
(? WATERREQUIREMENT :OF ROSE :- HIGH).

The system responds to each query with a determination of whether the statement is TRUE or FALSE along with an estimate of the certainty of its answer and an explanation of its reasoning. When presented with a query the system first checks whether it has the answer stored directly. If so, the answer is returned along with the certainty that was recorded at the same time the fact was recorded. If it does not have the fact stored it attempts to use every plausible inference for which it has adequate information and explains what it is doing as it performs each inference. The evidence from each plausible inference is then weighed together to generate a final guess of TRUE or FALSE along with the estimated certainty of that guess.

In general, people use many different, possibly independent, arguments to convince themselves of the truth or falsity of a

proposition. It is a bit like using a theorem prover that returns every possible proof. Unlike Bayesian inference networks (Pearl, 1986), which can be viewed as combining probabilistic evidence from multiple proofs to verify the truth of a proposition, our system tries to prove both the truth and, separately, the falsity of a proposition in as many ways as are possible given the information available.

Each inference made by PRSS is like a proof in that it may require backchaining to generate information necessary for the top level inference. Each top level inference (i.e. proof based on uncertain information) becomes a separate bit of evidence. Proofs that the query proposition are true are gathered together as evidence for the proposition and proofs of falsity are pooled as evidence against the proposition. Each bit of evidence has a certainty parameter that has been derived by combining the certainty parameters of the stored propositions used and parameters that measure the goodness of matches required in the applications of inference rules. The final judgment and the system's certainty of that judgment depend on the certainties of the evidence and on how contradictory the evidence was.

4. THE KNOWLEDGE BASE

We have tried to model the system on the behavior of people when generating functional inferences. This has required a highly redundant, crossreferenced memory organization. The knowledge representation system we developed for this purpose provides mechanisms for automatic crossreference of every input proposition, allowing for redundancies in set/subset relations, and multiple indexing of declarative inference rules. Collins and Michalski's theory assumes that inferences are made when relevant information is found by a parallel search for information associated with the argument and the referent of the query. While our current simulation does not do this directly, we have implemented a set of specialized search routines that collect all information potentially useful for (possibly several of) the inference types so far implemented.

PRSS has a database consisting of prepositional knowledge and functional relations (implications and mutual dependencies), organized in a multiply-indexed semantic network. In the existing implementation each proposition is a binary relation. We are currently working on extending the representation to include structured objects and n-ary relations.

Collins and Michalski (in press) identified four different certainty parameters associated with the propositions or declarative knowledge in this network. Two parameters, *certainty* and *frequency* are associated with each proposition in the knowledge base. For example, we might have

CLIMATE(AFRICA)-TEMPERATE, freq=3, cert-.9
CLIMATE(AFRICA)-TROPICAL, freq=.5, cert=HIGH

Following the notation of Collins and Michalski (in press), we call the predicate a *descriptor*, which, together with its argument (here, AFRICA) forms a *term*. The predicate CLIMATE is the descriptor, mapping its *argument* (a place) to various *referents* (values for climates). The *certainty* parameter is a measure of degree of certainty that a statement is believed to be true. The *frequency* parameter² measures the estimated proportion of the referent out of all possible referents for that descriptor and argument. The example above represents the belief that 30% of AFRICA is temperate and 50% is tropical.³

In addition to certainty, a likelihood parameter is attached to each implication and dependency. For example we might have the dependency,

²Corresponding to the all some distinction in logic.

³At present we assume that potential ambiguities associated with the meaning of the frequency parameter - e.g. does it refer to space or time - are accounted for by consistent interpretation by the user.

For all Places p,
 TEMPERATURE(p) \iff LATITUDE(p)
 certainty - .9, likelihood - HIGH.

where the likelihood is intended to be a measure of the conditional probability of the right-hand side given the left hand side. For an implication like the one below, it is a measure of the likelihood that the right hand side of the implication is in the given range when the left hand side is in its specified range.

For all Places p,
 GRAIN(p) - rice \implies rainfall(p) - heavy
 certainty - .9, likelihood - HIGH.

The fourth type of certainty parameter stored with the declarative knowledge of the system is *dominance*. A dominance parameter is associated with every set/subset link in the system. It measures the proportion of elements in the subset out of all elements in the set. For example, PART-OF(ENGLAND) - SURREY would have low dominance, since Surrey is a small part of England.

5. MULTIPLE TYPES OF INFERENCE

The current version of PRSS implements three basic types of functional inferences on statements retrieved from its memory, depending on the kind of dependency found and the resulting kind of contextually-based similarity match required. The three types are *functional analogies*, which are based on mutual dependencies between descriptors, *implication inferences*, and *set/subset inferences*.

In the example below, we show how the system is able to construct three separate "proofs" that the climate of England is temperate. Given the data in memory provided for this example, the system is unable to construct a single proof that the climate of England is not temperate.

(? climate :of england := temperate)

Using an Inheritance transform:

Since ENGLAND = PART-OF(EUROPE) (dom = LOW)
 And EUROPE has CLIMATE = TEMPERATE (cert = HIGH)
 Conclude that CLIMATE(ENGLAND) = TEMPERATE
 is TRUE with MED certainty.

Using an Implication transform:

Since LATITUDE = SECONDQUAD or THIRD QUAD
 \implies CLIMATE = TEMPERATE
 and LATITUDE(ENGLAND) = THIRDQUAD
 Conclude that CLIMATE(ENGLAND) = TEMPERATE
 is TRUE with MEDIUM certainty.

TRYING ARGUMENT BASED DEPENDENCY TRANSFORMS...

Using a SIM transform I reason:

Since LATITUDE \iff CLIMATE
 and HOLLAND is similar to ENGLAND with respect
 to LATITUDE (sim = 1.0)
 and CLIMATE(HOLLAND) = TEMPERATE
 Conclude that CLIMATE(ENGLAND) = TEMPERATE
 is TRUE with MEDIUM certainty.

TRYING REFERENT BASED DEPENDENCY TRANSFORMS

Insufficient Information Available.

I conclude CLIMATE(ENGLAND) = TEMPERATE, (cert = HIGH).

One general class of functional inference is called *statement transforms* (Collins and Michalski, in press). This type of inference requires a declarative rule called a *dependency*. In the example above, an analogy is made between England and Holland. The system is aware of a general relationship that the climate of a place is dependent upon the latitude of a place. In order to determine whether a specific relation exists between a latitude in the third-

quad (45-67.5 deg.) and a temperate climate the system must find an instance analogous to England which is known to have a temperate climate. Holland is such an instance. Since Holland and England have the same latitude the system can conclude that England can have a temperate climate as well.

Argument-based Transforms

GEN: flower-type(Europe)={daffodils, roses...}
 SPEC: flower-type(Surrey)={daffodils, roses...}
 SIM: flower-type(Holland)={daffodils, roses...}
 DIS: flower-type(Brazil)⁴ {daffodils, roses...}

Reference-based Transforms

GEN: flower-type(England)=temperate flowers...
 SPEC: flower-type(England)={yellow-roses...}
 SIM: flower-type(England)={peonies...}
 DIS: flower-type(England)= {bougainvillea...}

Figure 5-1: Eight Transforms on "flower-type(England)=Daffodils, roses..."

Within the class of statement transforms, Collins and Michalski (in press) describe eight different kinds of *transforms*, four *argument-based* transforms, and four *reference-based* transforms. The eight inference transforms were derived by considering concepts related to the ones mentioned in the question asked, where the relationship could be any of *generalization*, *specialization*, *similarity*, and *dissimilarity*. Each of these *operators* could be applied to either the argument or the referent in the question statement, giving the total of eight specific transforms. Figure 5-1 gives an example of each of the eight transforms for the statement FLOWER-TYPE(ENGLAND)={daffodils, roses...}. The overall certainty of an inference based on one of these transforms depends on the degree of similarity or typicality of the concepts related, as compared along the dimensions specified in the dependency used, and the degree of certainty of the dependency itself. The dependency used in the example above can be paraphrased "If two places match on latitude then they will match on climate."

The simplest type of functional inference is based on a type of declarative inference rule called an *implication*. Implication inferences can be used to infer values for properties on the basis of other properties of the same concept. Since the precise relation is completely specified in an implication, an analogous instance is not required for its application. The implication used in the example above can be paraphrased as "If the latitude of a place is third-quad then the climate of that place is temperate."

In the next example, the system first generates an argument-based statement transform using a dependency whose consequent is the queried descriptor, FLOWER-TYPE. It finds a place where tulips are grown (Holland) and compares that place to Venezuela on the antecedent descriptor of the dependency, CLIMATE. Since they do not match, it concludes that tulips don't grow in Venezuela. The second inference is a reference-based transform. Here, a dependency is required whose consequent is the inverse of the query descriptor FLOWER-TYPE (i.e GROWS-IN), since one needs to find a flower that grows in Venezuela and which is similar to tulips with respect to the factors that affect flower growth in a place.⁴

⁴The system uses a knowledge representation in which the descriptor definitions may specify an inverse. The descriptor FLOWER-TYPE has been defined as having a domain that must be a PLACE, a range that must be a FLOWER, and an inverse named GROWSIN. Thus while FLOWERTYPE maps from PLACES into the FLOWERS that grow there, GROWSIN maps FLOWERS into the PLACES where they grow.

⁵In the future, we plan to extend the matcher to compare multiple features with multiple values.

(? flower-type :of Venezuela :s tulip)

TRYING ARGUMENT BASED DEPENDENCY TRANSFORMS . . .

Using a DIS transform I reason:
Since CLIMATE \longleftrightarrow FLOWERTYPE
and HOLLAND is dissimilar to VENEZUELA with respect to
CLIMATE (sim = -1.0)
and FLOWERTYPE(HOLLAND) = TULIP.
Conclude that FLOWERTYPE (VENEZUELA) - TULIP is FALSE
with LOW certainty.

TRYING REFERENT BASED DEPENDENCY TRANSFORMS

Using a DIS transform I reason:
Since CLIMATE-OF \longleftrightarrow GROWSIN
and BOUGAINVILLEA is dissimilar to TULIP with respect
to CLIMATE-OF (sim = -1.0)
and GROWSIN (BOUGAINVILLEA) » VENEZUELA
Conclude that GROWSIN(TULIP) - VENEZUELA is FALSE
with LOW certainty.

I conclude TULIP IS NOT FLOWERTYPE of VENEZUELA
(cert - MED).

6. COMPUTING THE CERTAINTY OF AN INFERENCE

Each of the examples shown so far involves several types of inference, and the certainty of each inference is based on a combination of several certainty parameters and a similarity or typicality measure.

The two similarity parameters computed by the matcher are *similarity* and *typicality*. At present, these two parameters measure the quality of a match and are computed in exactly the same way. The difference between them is that typicality applies when a property (properties) of a set is being matched with those of a subset and similarity is computed as the quality of a match between two subsets. In the theory, similarity (or typicality) measures the quality of the match either of a single feature or of a bundle of features.

In the current implementation we compute the similarity (or typicality) of a single feature with multiple known values by an urn model type algorithm.⁵ The similarity parameter is currently computed as the probability that two values for a given feature, chosen at random within their frequency distributions, match or mismatch.

The certainty of each individual inference is currently computed as the minimum of all the certainty parameters and match certainties used. This includes the certainties associated with every proposition used, the certainty and the likelihood of the *inference* rule and similarity measure returned by the matcher.

Once the system has constructed every possible proof for a given proposition it must determine whether the proposition is true or false and estimate the certainty of its guess. Currently this is done by weighing the evidence for the proposition with the evidence against that proposition. The certainties of all of the positive conclusions are combined, and all of the negative conclusions are combined. Multiple lines of evidence in a given direction increases the certainty of the conclusion for that direction. The final judgment is the direction with the greater certainty, and the certainty of that judgment is downweighted by the certainty of the conclusion in the opposite direction.

7. CONCLUSION

This work is still in its early stages, and yet already we see a number of interesting issues that will require further study. To date, we have not run the simulation with large numbers of facts in memory, and we foresee that this will cause the number of inferences the system makes to grow exponentially. Clearly, techniques will be needed to control this growth, such as the filtering of weak and redundant inferences, the use of prototypes when many similar

examples exist, and more sophisticated representations for complex dependencies and implications. We also need to develop better and more efficient techniques for similarity matching, if we are to do matches on many contextual features at once. As the model continues to develop, we will also begin a new round of protocol experiments, in order to test our model, and answer some of the questions discovered by computer modeling.

References

- Carbonell, J. R., and Collins, A. M. Natural semantics in artificial intelligence. In *Proceedings of the Third IJCAI*. Morgan Kaufman, 1973.
- Cohen, Paul R. *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach*, Pitman, 1985.
- Collins, Allan. Fragments of a Theory of Human Plausible Reasoning. In D. L. Waltz (Ed.), *Theoretical Issues in Natural Language Processing*. Urbana-Champaign, IL: University of Illinois, 1978a.
- Collins, Allan; *Human Plausible Reasoning* (Tech. Rep. 3810). BBN, 1978b.
- Collins, A. and Michalski, R. The Logic of Plausible Reasoning: A Core Theory. *Cognitive Science*, in press..
- Collins, A., Warnock, E. H., Aiello, N. and Miller, M. Reasoning from Incomplete Knowledge. In D. Bobrow and A. Collins (Ed), *Representation and Understanding: Studies in Cognitive Science*. Academic Press, 1975.
- Duda, R. O, Hart, P. E., and Nilson, N. *Subjective Bayesian methods for rule-based inference systems* (Tech. Rep Technical Note 124). SRI International, 1976.
- Pearl, Judea. On the Logic of Probabilistic Dependencies. In *Proceedings of AAAI-86*. Morgan Kaufman, 1986.