# Identifying Conflicting Probabilistic Knowledge in Multiple Knowledge Base Systems

Daniel E. O'Leary
University of Southern California
3660 Trousdale Parkway
Los Angeles, CA 90089-1421

oleary@rcf.usc.edu



**Abstract**

Increasingly, models are being built that include the expertise of multiple experts. An important issue with such models is "when are the representations of those multiple experts in conflict with each other?" If the expertise conflicts then there are a number of concerns: Is there an error?; Do the experts belong to different schools?; Or is this conflict just a "signal" that there is a need for additional knowledge acquisition?

The existence of conflict is particularly critical in those situations where expert evaluations are "averaged." For example, what would it mean to average the assessments of supply and demand economists, or surgeons and chemotherapists?

Accordingly, the focus of this paper is on the identification of conflict situations, with particular emphasis on probability evaluations in multiple agent systems. Correlational statistics are used to identify conflict situations. In addition, a new approach, referred to as cutpoints, is developed to determine if probability distributions of multiple agents are in conflict. A case study is used to illustrate the problems of combining expertise in multiple agent systems and to demonstrate the approach.

## 1. Introduction

Increasingly, there has been an emphasis on the generation of knowledge-based systems that include multiple knowledge bases (e.g., Botten et al., 1989 and Jennings, 1994). For example, researchers (e.g., Gelernter, 1992) are developing multiple agent models of organizations that "mirror" their real-world counterparts. These mirror worlds can be used to assist in decision making, to make sense out of the large amounts of data that flows into an organization, to anticipate the outcome of sets of events, and a variety of other activities.

Storage of these multiple knowledge bases can take a number of different approaches, including multiple different rule sets (Ngwenyama and Bryson, 1992), or multiple sets of weights on the same set of rules (e.g., Reboh, 1983), or some combination. In such multiple knowledge base systems it often is necessary to determine if knowledge is nonconflicting or conflicting (e.g., Reboh, 1983). If the knowledge is nonconflicting, then the knowledge can be combined or one of the models can be used without concern with conflict. However, if there is conflicting knowledge, then steps can be taken to either choose one judgment, take alternative steps to combine the conflicting judgments (e.g., using negotiation) or search out new information (perform additional knowledge acquisition). As a result, a critical step in such systems is the determination of when the different knowledge bases are in "conflict." The focus of this paper is on those situations where probability distribution knowledge for multiple agents, on the same rules or arcs in influence diagrams, might be in conflict.

Consider a system where two experts have probability judgments of 1 and 0 for the same event x and 0 and 1 for the other event ~x. Such disparate judgments generally would signal that the experts have different models of the world. Alternatively, it may signal that there is an error in one of the assessments. In either case, combining these judgments, using approaches such as averaging, is likely to simply camouflage the disparate nature of the judgments. The resulting combination is unlikely to be representative of either agent, or the state of the world that the system is trying to capture. Thus, the purpose of this paper is to investigate methods for identifying those situations where multiple agent probability judgments are in conflict.

## 1.1 Importance of Determination of Conflicting Knowledge

The determination of conflicting knowledge is an important issue in the development of multiple agent knowledge-based systems for a number of reasons. First, unless such conflicts are investigated, system behavior may be affected. The combination of conflicting judgments is likely to result in system behavior that is not sensible. For example, if there are two schools of thought as represented by mutually exclusive probability distributions, what does it mean if the system combines them and uses the average. Second, the existence of conflicting judgments by multiple experts suggests that the system has been misspecified. If the system contains conflicting knowledge, one explanation is misuse or misinterpretation of information. If the system has been misspecified in one aspect, then it may be misspecified in others. As a result, it is critical to determine the correctness of those specifications. Third, the existence of multiple disparate probability judgments is likely to result in difficulties when the system is verified and validated. Tests of the data at the extreme points (e.g., x and ~x) will result in different responses from the system and the comparative human experts. Fourth, if a set of distributions is found to be conflicting, then the system needs to have abilities to account for such differences. For example, the system can have users or developers choose which distribution should be used. Alternatively, if the system is provided with knowledge about the experts, then it might be able choose, say, the more expert agent.

## 1.2 Outline of This Paper

This paper proceeds as follows. Section 2 provides a brief background on multiple agent knowledge-based systems, including rule-based systems and Belief Network/Influence Diagram. Section 3 summarizes the case study from which the data used in this paper is generated. Section 4 investigates two metrics for determining if the distribution estimates of two agents are in conflict. Section 5 briefly summarizes the paper and its contributions, and analyzes some extensions.

# 2. Background

## 2.1 Rule-based Systems and Belief Nets

Rule-based systems and Belief Nets are similar in their structure. In addition, both use of probabilities, either as weights on rules or directly. Rule-based expert systems represent knowledge using "if a then b" rules. Often those systems employ probability measures of uncertainty on the rules. Inference through the rule base is done using heuristic approaches.

Belief Nets (also called Bayes' Nets and Influence diagrams) are graphical structures that facilitate Bayesian reasoning (Pearl, 1988). They are acyclic graphs that are used to represent any decision problem that can be captured as a decision tree. Roughly the arcs mean that if you know the state of the node at one end then you can infer about the node at the other end. Typically, Belief Nets have probabilities associated with each arc, in a manner similar to the probabilities or uncertainty factors associated with rules. Probability inference through the network is done using Bayes' Theorem or some heuristic approximation.

## 2.2 Timing Integration of Judgments in Multiple Knowledge Bases

Multiple agent systems can combine or choose between the judgments of multiple experts at two different times: the time the system is built or at the time the system is run. The first approach uses assessments from multiple agents to establish a single system. Dungan (1983) and Dungan and Chandler (1985) built a rule-based system that integrated the judgments of multiple experts at the time it was built. Weights for those rules were gathered from four different sources and then combined into a single estimate on each of the rules through a process of averaging.

The second approach provides more flexibility, allowing for evolving sets of agents. In this approach, e.g., the weights or probabilities on the rules would be captured as the system was being built. Then on compilation, the system would combine (e.g., average) the weights at the time or the user would chose which weights should be combined. This would permit the ability to change one subset of weights or probabilities, without making major changes in the system. Such an approach would facilitate an evolutionary system design. For example, Garvey et al. (1981) suggested that the knowledge of specialists, with different information, should be integrated in the system. Reboh (1982) used a similar approach, integrating different sets of rules. LeClair (1985) developed a system that permitted the user to choose from or average different experts represented in the system.

## 2.3 Agent Probability Distribution Judgments

This paper focuses on those situations where probability judgments are generated from multiple agents, for multiple knowledge bases, for expert systems and influence diagrams. In particular, it is assumed that multiple agents provide an estimate of a discrete probability distribution, for each expert system rule or influence diagram arc. As is often the situation with influence diagrams, experts would provide a discrete probability distribution across a number of "categories." In the example in the introduction, experts provided probability estimates of x and ~x of 1 and 0, and 0 and 1, respectively.

The previous research in artificial intelligence on the existence of conflicts is very limited. In one of the few discussions on the topics, Reboh (1983) described how the well-known expert system Prospector, determines and processes the effect of conflicts. Reboh (1983, p. 149)) defines rules to be in conflict when there are "conflicting rule strengths." In the case of the Prospector system, this meant that the point-estimate Bayesian-based AL/X weights (e.g., Duda et al., 1979) are of different strengths. As noted by Reboh (1983, p. 149), "... when Prospector discovers conflicting rules with identical left- and right-hand sides, .. it declares an inconsistency; the knowledge engineer must then resolve the situation by talking to the experts ...." However, Prospector was not a multiple agent system, and thus did not have conflicts between multiple agents.

## 3. Case Study: Pathfinder

Pathfinder is a Bayes Net that uses a large number of disease - feature pairs (related by probabilities) to facilitate diagnosis of pathologies of the lymph system. The system reasons about approximately 60 malignant and benign diseases of lymph nodes. Pathfinder has been discussed in detail, in a number of sources (e.g., Heckerman et al. 1992 and Ng 1991).

According to Heckerman et al. (1992), pathologists apply knowledge about features on a slide to determine the likelihood of alternative diseases. That diagnosis is then given to an oncologist who, based on this recommendation, directs a patient's therapy. Accordingly, the therapy is greatly dependent on the accuracy of the diagnosis.

Pathfinder is a multiple agent system that incorporates the judgment of multiple pathologists. As such, it provides a model of the diagnosis process. Ng and Abramson (1994) list the multiple agent probability distributions associated with a small portion of the system. In particular, the distributions for thirteen different disease - feature "arcs" are summarized in Table 1.

An examination of the probability distributions in table 1 finds that in some cases the distributions are very similar, while in other cases they appear to be quite different. For example, the distributions for arc 1 appear to be about the same for both experts, while, the distributions for arc 9 appear substantially different (in conflict) for each of the agents. However, these are qualitative assessments, quantitative measures of the extent of similarity would be helpful in determining when the distributions of the agents are similar and "substantially different."

Table 1: Complete Set of Probability Assessments@

| Arc # | Category 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. | .990 | .010 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 2. | .990 | .010 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 3. | .985 | .015 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 4. | .985 | .015 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 5. | .990 | .010 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 6. | .990 | .010 | .000 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 7. | .000 | .010 | .400 | .500 | .090 | .000 |
| | .000 | .200 | .600 | .200 | .000 | .000 |
| 8. | .000 | .000 | .000 | .000 | .000 | 1.000 |
| | .000 | .000 | .600 | .200 | .200 | .000 |
| 9. | .980 | .015 | .005 | .000 | .000 | .000 |
| | .000 | .200 | .600 | .200 | .000 | .000 |
| 10. | .900 | .090 | .010 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 11. | .980 | .015 | .005 | .000 | .000 | .000 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | .900 | .100 | .000 | .000 | .000 | .000 |
| 12. | .900 | .090 | .010 | .000 | .000 | .000 |
| | 1.000 | .000 | .000 | .000 | .000 | .000 |
| 13. | .000 | .010 | .400 | .500 | .090 | .000 |
| | .000 | .800 | .200 | .000 | .000 | .000 |

@ For each "arc" the first (second) line corresponds to expert #1 (#2)

Categories - Lacunar SR: 1 = Absent; 2 = Rare; 3 = Few; 4 = Many; 5 = Striking; 6 = Sheets.

Source: Ng and Abramson (1994)

The developers choose the approach of generating a single probability for each arc, at run time. Thus, the development of the system required the integration of probability information from two expert agents (or more), by the system. This generally meant averaging the probability distributions. Although in many cases the expert agents generated virtually identical distributions, in some cases there was question as to the similarity of the judgments. In these cases of conflict, it likely is inappropriate to average distributions. As a result, it is necessary to determine if the distributions are disparate, in order to determine if one of the distributions should be chosen (e.g., because of greater expertise) or in order to determine the need for additional information (e.g., through knowledge acquisition or from the user as to their preferences).

## 4. Analysis of Agent Probability Distributions

The purpose of this section is to investigate methods for identifying whether or not two agents' probability distributions are in conflict. Two approaches are employed. First, a traditional statistical correlation analysis is employed. Second, an approach based on Kolmogorov-Smirnov, referred to as cutpoints, is developed and discussed.

### 4.1 Correlational Analysis

Assume that for each of two agents, for each rule or arc, there is a probability distribution across a set of n points. We can use the correlation to measure the extent of similarity. The statistical significance of the correlation can be used to determine if the agents' distributions are "in conflict" or are "similar."

In terms of the case, the Pearson correlation coefficients, between the two experts' distribution estimates are as follows: arcs 1-6 .999; arc 7, .686; arc 8, -.349; arc 9, -.345; arcs 10-12, .995; and arc 13, -.219. In the case of arcs 1-6, and 10-12, the arcs' correlations are highly statistically significant, at .03 and .01, respectively. Thus, we reject the hypothesis that the distributions are not correlated.

The correlation coefficient for arc 7 was not statistically significant. The correlation coefficients for arcs 8, 9 and 13 were negative and found not statistically significant. Thus, we reject the hypothesis that there is a correlation between the distributions, for arcs 7, 8, 9 and 13. As a result, this metric signals that the distributions on arcs 1-6 and 10-12 are not in conflict. However, the correlation coefficients for arcs 7, 8, 9 and 13 suggest that those agents' probability distributions are in conflict.

Unfortunately, the analysis of the statistical significance of the correlation coefficient has some limitations in the context of multiple agent systems. First, in the generation of most multiple agent systems, the number of categories n, will be small. However, as n approaches 3 the measure of statistical significance approaches 0, since the factor (n-3) is used in the determination of the statistical significance (e.g., Freund 1971). Second, this test of statistical significance of the correlation coefficient assumes a bivariate normal distribution. Unfortunately, that assumption is not always valid (e.g., Freund 1971). As a result, consider an alternative approach.

### 4.2 Cutpoints

This section presents an approach based on the Kolmogorov-Smirnov (K-S) test (Ewart et al. 1982). This approach, referred to as cutpoints, requires no distribution assumptions. Basically, the cutpoint approach compares the cumulative frequencies associated with different categories for two discrete distributions being compared. The difference between the probabilities of those two distributions is calculated at each category. If the maximum value exceeds a specified level then the hypothesis that the two distributions are the same is rejected, and the agents distributions will be said to be in conflict.

For the discrete probability distributions on the individual arcs, such as those listed in table 1, each category will be referred to as an index number. Some of those indices have interesting properties that will help us determine if the distributions of the two experts are in conflict.

Define a maximal cutpoint as an index (in the example ranging from 1 to 6) such that the difference in the cumulative probability ("distribution difference"), between the two distributions, at that index, that is maximal. For example, in the case of arc 7, at category 3 the distribution for expert 1 has probability of .410, while that of expert 2 has probability of .800. The difference of .390 is larger than that of any other cutpoint, for n = 1, ..., 6. The complete set of maximal cutpoints, for the case, is given in table 2.

Define a zero cutpoint as an index where the cumulative probability for one distribution is zero and the cumulative probability for the other distribution is nonzero. There may be more than one zero cutpoint for a distribution. For example, in the case of arc 8, zero cutpoints occur at indices 3, 4, and 5.

Define a double zero cutpoint as a maximal cutpoint, where the nonzero probability equals one. In that case, there is an index where all the probability for one expert is on one side of the index and all the probability for the other expert is on the other side of the index. For example, as shown for arc 8 there is a double zero cutpoint at the index 5. There may be multiple double zero cutpoints.

## 4.3 Use of Cutpoints

Cutpoints can be useful in the analysis of the similarity of two probability distributions on an arc. First, the occurrence of a double zero cutpoint is probably the most critical. Zero and double zero cutpoints define alternative ways to define the entire distribution, with two indices, say x and ~x . That revised distribution, with a double zero cutpoint, has zero probability associated with x and ~x for each of the two experts. This implies the two experts see certainty of mutually exclusive sets of events. Thus, rather than just defining level, there can be implications for structure: A zero probability between two events indicates no relationship between events.

Table 2: Maximal Cutpoints for the Sample of Probability Assessments

| Arc # | Expert #1 | | Expert #2 | | | |
|---|---|---|---|---|---|---|
| | x | x' | x | x' | Location | Amount |
| 1. | .990 | .010 | 1.000 | .000 | 1 | .010 |
| 2. | .990 | .010 | 1.000 | .000 | 1 | .010 |
| 3. | .985 | .015 | 1.000 | .000 | 1 | .015 |
| 4. | .985 | .015 | 1.000 | .000 | 1 | .015 |
| 5. | .990 | .010 | 1.000 | .000 | 1 | .010 |
| 6. | .990 | .010 | 1.000 | .000 | 1 | .010 |
| 7. | .410 | .590 | .800 | .200 | 3 | .390 |
| 8. | .000 | 1.000 | 1.000 | .000 | 5 | 1.000 |
| 9. | .980 | .020 | .000 | 1.000 | 1 | .980 |
| 10. | .900 | .100 | 1.000 | .000 | 1 | .100 |
| 11. | .980 | .020 | .900 | .100 | 1 | .080 |
| 12 | .900 | .100 | 1.000 | .000 | 1 | .100 |
| 13. | .010 | .990 | .800 | .200 | 2 | .790 |

---------------------------

Source: Ng and Abramson (1994)

"Location" refers to category at which maximal cutpoint occurs.

"Amount" is the absolute value of (Pr(x for expert 1) - Pr(x for expert 2))

Second, the maximal cutpoint provides insight into the similarity of the distributions of the two experts. The maximal cutpoint value provides a measure that allows us to assess the point of maximal difference between the experts. One approach would be to suggest that a maximal cutpoint of .10 or lower (or .05 or .01, as in classic probability theory) would be viewed as similar, while cutpoints with distribution differences larger than .10 would be viewed as in conflict. This approach indicates that arcs 7, 8, 9 and 13 would be viewed as in conflict at the .10 level. In this case the results are the same as the use of the correlation coefficient analysis.

Third, maximal cutpoints are useful in describing the index number behavior. In particular, the maximal cutpoints for a set of arcs provides a distribution of cutpoints. In the example, "1" is a maximal cutpoint ten times, "2," "3," and "5," (arcs 7, 8 and 13) are each cutpoints one time. As a result, we might assert that the comparison of the probabilities distributions for arcs 7, 8 and 13 behave differently than the comparison of the other arcs. This could suggest that the distributions of the two agents for those arcs are sufficiently different than the other distributions.

Practically, the correlational and cutpoint approach are substantially different. Whereas the correlation approach provides a point to point comparison, cutpoints provide more of a distribution to distribution analysis.

## 4.4 Additional Criteria

Although this paper has focused on two primary approaches (correlation analysis and cut points), potentially other approaches could also be used to investigate potential conflict. Decision theory and information theory potentially could be used to compare the distributions.

Decision theoretic approaches that employ utility functions could also be used to determine the existence of conflicts. This approach would require additional information to generate the utility function. In the case of the example system there is no data available regarding the utility of particular estimates or conflicts to illustrate the use of a utility function.

One information theory measure, I, uses the following formula to compute expected information, given two distributions of probabilities $q_i$ and $p_i$.

$$I = \Sigma\, q_i \log (q_i/p_i).$$

Unfortunately, I is not defined for virtually each of the distributions in table 1 because of the zero values.

## 5. Summary

This paper has investigated the issue of when probability assessments of multiple experts in the generation of rule-based systems and Belief Networks are similar or in conflict. Particular interest was in identifying those situation were the probability assessments of multiple experts were in conflict. For example, in the situation where one expert identifies the probability of x (~x) as 0 (1) and the other expert identifies the probability of x (~x) as 1 (0), there is a conflict. Two approaches were used to identify conflict: correlation coefficient and cutpoints. Cutpoints are a new approach that do not have some of the same limitations of correlation analysis.

## References


Botten, N., Kusiak, A. and Raz, T., "Knowledge Bases: Integration, Verification and Partitioning," European Journal of Operational Research, Volume 42, 1989, pp. 111-128.

Duda, R., Gaschnig, J., and Hart, P., "Model Design in the Prospector Consultant System for Mineral Exploration," in D. Mitchie, (Ed.) Expert Systems for the Microelectronic Age, Edinburgh, Edinburgh University Press, 1979.

Duda, R., Hart, P. and Nillsson, N., "Subjective Bayesian Methods for Rule-based Inference Systems," National Computer Conference, pp. 1075-1082, 1976.

Dungan, C., "A Model of Auditor Judgment in the Form of an Expert System," Unpublished Ph. D. Dissertation, University of Illinois, 1983.

Dungan, C. and Chandler, J., "Auditor: A Microcomputer-Based Expert System to Support Auditors in the Field," Expert Systems, Volume 2, Number 4, 1985, pp. 210-221.

Edgington, E., Randomization Tests, Marcel Dekker, 1980.

Efron, B., "Bootstrap Methods: Another Look at the Jackknife, "Annals of Statistics, Volume 7, 1979.

Ewart, P., Ford, J. and Lin, C., Applied Managerial Statistics, Prentice-Hall, Englewood Cliffs, NJ, 1982.

Freund, J., Mathematical Statistics, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.

Garvey, T., Lawrence, J. and Fishler, M., "An Inference Technique for Integrating Knowledge From Disparate Sources," Proceedings of the Seventh International Joint Conference on Artificial Intelligence, pp. 319-325, 1981.

Gasser, L. and Hill, R., "Coordinated Problem Solvers," Annual Review of Computer Science, volume 4, pp. 203-253, 1990.

Gelernter, D., Mirror Worlds, Oxford, New York, 1992.

Heckerman, D., Horvitz, E., and Nathwani, B., "Toward Normative Expert Systems: Parts I and II. The Pathfinder Project," Methods of Information in Medicine, Volume 31, pp. 90-105 and 106-116, 1992.

Hollander, M. and Wolfe, D., Nonparametric Statistical Methods, John Wiley & Sons, New York, 1973.

Jennings, N., Cooperation in Industrial Multiple Agent Systems, World Scientific, Singapore, 1994.

LeClair, S. "Justification of Advanced Manufacturing Technology Using Expert Systems," in Expert Systems, edited by N. Botten and T. Raz, pp. 191-216, Industrial and Management Press, Norcross, GA, 1985.

Ng, K. "Probabilistic Multi-Knowledge-Base Systems," Ph. D. Dissertation, University of Southern California, 1991.

Ng, K. and Abramson, B., "Probabilistic Multi-Knowledge Base Systems," Journal of Applied Intelligence, Volume 4, Number 2, 1994, pp. 219-236

Ngwenyama, O. and Bryson, N., "A Formal Method for Analyzing and Integrating the Rule Sets of Multiple Experts," Information Systems, Volume 17, Number 1, 1992, pp. 1-16.

Noreen, E. An Introduction to Testing Hypotheses Using Computer Intensive Statistics, John Wiley, 1989.

Pearl, J., Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan-Kaufman, 1988.

Reboh, R., "Extracting Useful Advice from Conflicting Expertise," Proceedings of the Eight International Joint Conference on Artificial Intelligence, pp. 319-325, 1982.