Computational & Mathematical Organization Theory 3:2 (1997): 75–90

Validation of Computational Models Based on Multiple Heterogeneous Knowledge Sources

DANIEL E. O'LEARY
University of Southern California, 3660 Trousdale Parkway, Los Angeles, CA 90089-1421
email: oleary@rcf.usc.edu

Abstract

Theories of organizations have brought together multiple heterogeneous theories in computational models. In addition, in artificial intelligence, there has been an emphasis on the generation of knowledge-based systems that include multiple heterogeneous knowledge bases. As a result, increasingly, theory and model developers have called for the need to validate these computational models. Unfortunately, there has been only limited attention given to validation of multiple knowledge source programs.

The primary focus of this paper is on the identification of conflict between multiple knowledge bases. The existence of conflict is particularly critical in those situations where database evaluations are "averaged". For example, what would it mean to average the assessments of supply and demand economists, or surgeons and chemotherapists?

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.

Keywords: computational models of organizations, multiple agent systems, Bayes' nets, probabilistic systems, validation

1. Introduction

Theories of organizations have brought together multiple heterogeneous theories (e.g., Cyert and March 1992). As a result, it is probably not surprising that computational models representing organization theories (e.g., Cyert and March 1992; Burton and Obel 1984; Baligh et al. 1996) also bring together multiple theories. Accordingly, heterogeneous knowledge is represented within these models. That heterogeneous knowledge can either be maintained as integrated wholes or as heterogeneous knowledge.

In addition, increasingly, in artificial intelligence, there has been an emphasis on the generation of knowledge-based systems that include multiple heterogeneous knowledge bases (e.g., Botten et al. 1989; Jennings 1994). For example, researchers (e.g., Gelernter 1992) are developing multiple knowledge base 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.

Theory and model developers also have called for the need to validate these computational models (e.g., Cyert and March 1992; Burton and Obel 1984). Since the resulting knowledge-based models are software, they ultimately need to be validated as software.

Thus, as noted by Adrion et al. (1982) we are concerned with "determination of the correctness of the final program or software...." Unfortunately, there has been only limited attention given to validation of multiple knowledge source programs. As a result, the purpose of this paper is to develop approaches that help us validate systems with multiple heterogeneous knowledge sources in order to determine their correctness.

This paper proceeds as follows. Section 2 provides a brief background on multiple knowledge-based systems, including rule-based systems and Belief Network/Influence Diagram. Section 3 briefly reviews the previous validation research. Section 4 examines the validation of multiple knowledge sources, including issues in ontologies, knowledge scope and detail and problems of conflicting knowledge. Section 5 summarizes the case study from which the data used in this paper is generated. Section 6 investigates two metrics for determining if the distribution estimates associated with two knowledge bases are in conflict. Section 7 uses the discussion here to extend the methodology in organizational social network research. Section 8 briefly summarizes the paper and its contributions, and analyzes some extensions.

2. Background

This section provides a brief background on influence diagrams and rule-based systems. In addition, it briefly discusses multiple heterogeneous knowledge sources.

2.1. *Rule-Based Systems and Belief Nets*

Rule-based systems and Belief Nets are similar in their structure. In addition, both use 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 and inference through the rule base using heuristic approaches. Rules are a very robust way to capture knowledge in organizational settings. For example as noted by Cyert and March (1992, p. 230)

> ...theories of rational, anticipatory, calculated consequential action underestimate both the persuasiveness and intelligence of an alternative decision logic—the logic of appropriateness, obligation, identity, duty and rules. Much of the decision making behavior we observe reflects the routine way in which people do what they believe they are suppose to do. Much of the behavior in an organization is specified by standard operating procedures, professional standards, cultural norms and institutional structures. Decisions in organizations, as in individuals, seem often to involve finding 'appropriate' rules to follow.

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 (most rule-based problems can be represented 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. *Heterogeneous Knowledge Sources*

Knowledge bases generated from heterogeneous knowledge sources come from more that a single source of knowledge and as a result, represent multiple views, multiple situations, multiple times of when the knowledge was valid, etc. Representation of these multiple heterogeneous knowledge bases can take a number of different approaches, including constraints (e.g., Burton and Obel 1984), multiple rule sets (Ngwenyama and Bryson 1992), multiple rule sets with multiple objectives (e.g., O'Leary 1986), multiple sets of weights on the same set of rules (e.g., Reboh 1982), and other combinations.

2.3. *Integration of Judgments in Heterogeneous Knowledge Bases*

Multiple knowledge based systems can combine or choose between the judgments of multiple knowledge sources at basically two different times: the time the system is built or 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 and each set of weights would be maintained separately. Then when the system was run, 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 by allowing developers to phase in and out the knowledge of particular individuals. 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.

If the heterogeneous knowledge is maintained separately, then it is possible to determine which bits of knowledge come from different sources for purposes of comparing the different knowledge sources. Unfortunately, once the multiple knowledge sources are aggregated, the knowledge may take a form other than the original intention (figure 1).

However, if knowledge is kept in its disaggregated, multiple knowledge base format the means by which the system aggregates or employs the multiple sources to make its decisions will also require validation. If that approach is simple averaging across expert estimates then that task is not substantial. However, if the multiple knowledge sources are integrated using negotiation or other approaches then the effort can be substantial.

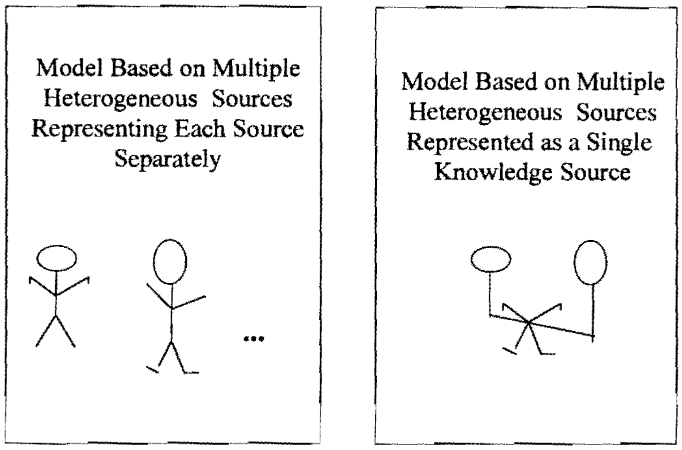

Figure 1.

3. Previous Validation Research

The previous validation research relating to validation of models with multiple heterogeneous knowledge bases can draw from at least two general lines of research. First, there is a substantial literature relating to validation of artificially intelligent and knowledge based systems (e.g., O'Keefe and O'Leary 1993). Second, validation of computational models of organizations is emerging as an important source of advances (e.g., Burton and Obel 1984, 1995; Carley 1996). In each line of research, the literature has grown around the knowledge representation requirements of the models being developed. As a result, in the knowledge based systems arena there has been a focus on rule-based systems. Further, the knowledge base systems validation literature has begun to address issues in the validation of multiple knowledge sources as developers begin to generate more such models. However, an analysis of both those literatures reveals that there has been limited, if any, attention given to multiple heterogeneous knowledge bases. Thus far the literature has focused on two primary topics: using the knowledge representation structure (e.g., rules) to find errors in the way the knowledge has been developed and the use of test data.

Using the Knowledge Representation Structure. If the knowledge is represented in a single manner, such as rules, then it is possible to use the structure of the rules to determine consistency, completeness and correctness of the rules from a "syntactic" perspective (O'Leary 1995). Similarly, with knowledge represented as constraints, inconsistency between knowledge sources can take other identifiable forms such as infeasibility.

Test Data. Brown et al. (1995) employed a test data approach in order to evaluate heterogeneous and competing knowledge bases. Their analysis found that for complex models it is difficult to generate test data cases. However, test data was effective in ascertaining "gaps" and incorrect knowledge in the knowledge bases.

4. Validating Multiple Knowledge Sources

One of the primary ways of validating a knowledge base is to compare it to some alternative. The set of alternative bases of comparison is quite extensive (e.g., O'Leary 1988a) and includes other models, other knowledge bases, other experts and a gold standard. If each knowledge source is represented separately then they can be compared against each other or test sources to facilitate determination as to the correctness of the knowledge (figure 2). If the knowledge is aggregated then the aggregate can be compared to each of the original sources to determine the extent to which it conforms with the original knowledge or it can be compared to test sources.

Comparisons of heterogeneous knowledge sources can yield a number of differences. First, there can be differences in ontologies being used by the different knowledge sources. Second, there may be different scopes to knowledge, so that one knowledge base has additional rules when compared to another knowledge base. Third, the knowledge sources may be in conflict.

4.1. Different Ontologies

Onologies define a sub language for a specific topic area, thus defining the terms and relations that constitute the vocabulary (e.g., Gruber 1993). In order to use multiple knowledge sources unambiguously each of the knowledge sources must employ the same ontology. For example, do all knowledge sources mean measure the same thing when they reference the concept "decentralization"? As a result, validation must determine if the heterogeneous

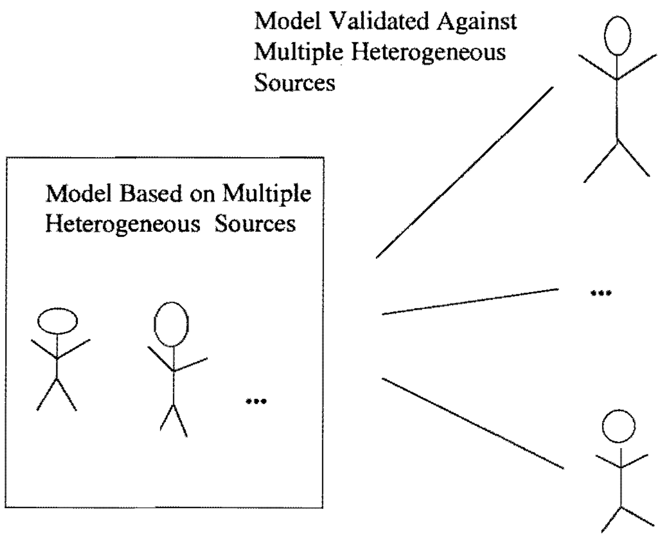

Figure 2.

sources use the same ontology or if they are integrated using some other approach. One such approach to integrating multiple heterogeneous knowledge sources with different ontologies uses brokers. The broker is an intermediary who is able to translate from one ontology to another and back again to facilitate communication. Unfortunately, there are a number of impediments to the use of explicit ontologies (e.g., O'Leary 1997).

Determining whether heterogeneous knowledge sources employ the same ontologies is not an easy task. However, there are some simple tests that can be employed. First, if there are any words used by one knowledge source but not the other then that can signal usage of different ontologies. Further, if there are different variables or different values for variables then that also can signal different ontologies. Other tests can also be generated.

In some situations there may be a referenced ontology list or an established ontology. In these cases, similarity of ontology can be directly established.

4.2. *Different Knowledge Scope and Detail*

If knowledge sources are heterogeneous then it is likely that they have a different scope. For example, one knowledge source may have rules about events that another knowledge source does not. Another knowledge source may have greater detail or knowledge fineness than other knowledge sources (e.g., O'Leary 1988b).

However, the scope or detail per se does not make the model valid or invalid. If the knowledge sources are heterogeneous then we would expect there to be differences between them. Such differences may be a signal of model validity or differences may signal the quality of one knowledge base compared to another. Again some simple tests can provide immediate insight into knowledge scope and detail: How many rules does each knowledge base have?; How many values are associated with each of the same variables?; How many rules contain each particular concept, e.g., "decentralization"?; etc. Ultimately, heterogeneous knowledge bases should be compared to their requirements to determine if they comform to those requirements.

4.3. *Conflicting Knowledge*

In heterogeneous multiple knowledge base systems it is necessary to determine if multiple knowledge sources are nonconflicting or conflicting (unless the purpose is to capture and/or model conflict). 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 determination of conflicting knowledge is an important issue in the development of multiple 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 judgments is likely to result in difficulties when the system is validated. Tests of the data at the extreme points (e.g., x and $\sim x$) 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 of the conflicting knowledge bases should be used. Alternatively, if the system is provided with knowledge about the experts, then it might be able choose the most expert, applicable or some other criteria for the knowledge base.

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. 1976, 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.

4.4. *Determining Conflicting Qualitative Knowledge*

Conflicts between knowledge sources can arise in qualitative or quantitative knowledge. In qualitative knowledge, assuming that each knowledge base (KB) employs the same ontology and a rule-based structure, a number of different kinds of conflicts can occur

- Direct Conflicts: KB 1 "If a then b" and KB 2 "If a then $\sim b$"
- Indirect Conflicts: KB 1 "If a then b" and KB 2 "If a then c"
- Subsumed Conflicts: KB 1 "If a then b then c" and KB 2 "If a then c".

In addition, conflicts can relate to variable values. For example, rules of the following type are in conflict: KB 1—If $a = \Lambda$ then $c = \Gamma$ and KB 2—If $a = \Lambda$ then $c = \Omega$. A wide range of variable conflicts in qualitative knowledge can be further identified.

4.5. *Determining Conflicting Probability Knowledge*

The primary focus of this paper is on those situations where probability distribution for multiple knowledge bases, on the same rules or arcs in Bayes' Nets, 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 $\sim x$. 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 knowledge base, 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 probability judgments are in conflict.

The remainder of this paper focuses on those situations where probability judgments are available from multiple knowledge bases. It is assumed that each knowledge base has an estimate of a discrete probability distribution, for each specific rule or Bayes' Net arc. As a result, there are no differences in scope of knowledge. Further, discrete probability distribution are estimated across a number of "categories". In the example earlier, there were probability estimates of x and $\sim x$ of 1 and 0, and 0 and 1, respectively. Since the concern is with predefined categories, there is no discernible ontology problem or scope of knowledge problem.

5. 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; 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 knowledge base system that incorporates the judgment of multiple pathologists. As such, it provides an organization-like model of diagnosis. Ng and Abramson (1994) list the multiple probability distributions associated with a single feature and kept in the system. Pathfinder contains thirteen node diseases and a number of different nodes features. Arcs were used to connect diseases to the features and a probability distribution was associated with each disease and feature pair. 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 knowledge bases. However, these are qualitative assessments, quantitative measures of the extent of similarity would be helpful in determining when the distributions of the knowledge bases are similar and "substantially different".

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 knowledge bases (or more), by the system. This generally meant averaging the probability distributions. Although in many cases the knowledge bases had

Table 1. Complete set of probability assessments.*

	Category					
Arc#	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

Source: Ng and Abramson (1994).
*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.

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).

6. Analysis of Knowledge Base Probability Distributions

The purpose of this section is to investigate methods for identifying whether or not two knowledge base 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. These approaches are summarized in figure 3.

6.1. *Correlational Analysis*

Assume that for each of two knowledge bases, 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 knowledge bases probability distributions are "in conflict" or are "similar".

In terms of the case, the Pearson correlation coefficients, between the two 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 those particular sets of 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 knowledge base probability distributions are in conflict for those arcs.

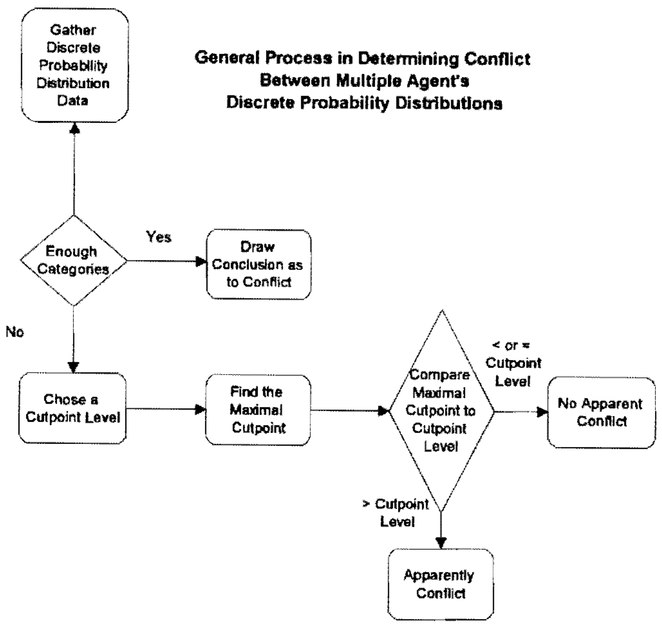

Figure 3.

Unfortunately, the analysis of the statistical significance of the correlation coefficient has some limitations in the context of multiple heterogeneous knowledge base systems. First, in the generation of most multiple knowledge base 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). Third, the probability distributions seem to have a number of zero probability categories, which can cause statistical debates. As a result, consider an alternative approach.

6.2. *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 knowledge base 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 $n = 1, \ldots, 6$. 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.

6.3. *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.

Table 2. Maximal cutpoints for the sample of probability assessments.

	Expert #1		Expert #2			
Arc #	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)).

Zero and double zero cutpoints define alternative ways to define the entire distribution, with two indices, say x and $\sim x$. That revised distribution, with a double zero cutpoint, has zero probability associated with x and $\sim x$ 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.

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 knowledge bases for those arcs are sufficiently different than the other distributions to warrant treatment as a conflict.

6.4. *More Than Two Distributions*

The approach presented here can be extended to the analysis of $m \geq 3$ distributions. Since the concern is with finding those distributions that are most disparate, the focus is on the maximum difference. Thus, the multiple distributions can be compared on a pairwise basis for all such pairs (i, j). Then, the maximal pairwise cutpoint would be generated by determining the maximal cutpoint across all such pairs (i, j). Similarly, the concern with zero and double zero cutpoints becomes, "does there exist a pair of arcs for which there is a zero or double zero cutpoint?"

6.5. *The Use Cutpoints With Other Multiple Knowledge Base Forms of Interaction*

Cutpoints can be used with other multiple knowledge base protocols in the development of multiple agent systems, such as consensus or negotiations. First, cutpoints can be used to indicate that there is a substantial difference between the knowledge bases that needs to be resolved. Zero and double zero cutpoints might signal the need for negotiation. Second, it can indicate that consensus may not be appropriate as a basis for generating a solution. The existence of a double zero cutpoint indicates substantial differences. It can be important to resolve those differences prior to generating a consensus judgment.

7. Extension: Organization Social Network Data[1]

Researchers in organizational social network research have addressed a problem that also generates multiple databases. For example, researchers like Krackhardt (1990, 1991) collect cognitive social structure data in which they ask each respondent, "who (j) would this person (i) go to for help or advice at work? (Krackhardt 1990, p. 348)". If the respondent indicates that i would go to j then there is an arc in the network from i to j. The result is a network of relationships for each respondent, across all the people in the organization. These respondent networks can then be compared to what is call the "actual" network that has an arc from i to j, if both agree.

Unfortunately, there is an arc in the network whether someone would ask for help or advice once or a thousand times. Rather than an unqualified "who would this person go to..." a probability based model could be generated associated with each relationship pair. The distribution could measure, e.g., probability of frequency of communications, or probability of importance of problems, and provide a more descriptive model. As a result, Bayes' Nets could be used to model this problem. In addition, the discussion here regarding conflicts could be used to investigate the resulting data to determine which distribution pairs were in conflict.

8. Summary, Extensions and Contributions

This section briefly summarizes the paper, reviews some potential extensions and discusses some of the contributions of the paper.

8.1. Summary

This paper has investigated the issue of when knowledge bases are similar or in conflict for purposes of validating a computational model. Particular interest was in identifying those situations where probability assessments of multiple knowledge bases are in conflict. For example, in the situation where knowledge base identifies the probability of x ($\sim x$) as 0 (1) and another identifies the probability of x ($\sim x$) 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.

8.2. Other Extensions

This paper can be extended in a number of ways. First, although the probability distributions were used in a Belief Net, the analysis is not limited to that context.

Second, in some cases we may be able to determine that the sets of probability distribution assessments come from specific distributions, e.g., binomial. In such a case, rather than using traditional statistical approaches to generate estimates of statistical significance, we could generate specific distribution-based, distributions of test statistics.

Third, nonparametric approaches could be used to evaluate the statistical significance. For example, computer intensive statistics (Efron 1979; Noreen 1989) could be used to generate distributions of test statistics in order to assess the statistical significance of a particular correlation coefficient. Alternatively, Kendall's tau could be used to measure correlation (Hollander and Wolfe 1973).

Fourth, the concept of cutpoint could be generalized beyond the constraint of assuming the indices are ordered. A "generalized" maximal cutpoint sets would be the generation of two sets of mutually exclusive indices such that the difference in the probability associated with those sets is a maximum. For most arcs it appears that the generalized and original definitions of cutpoints will be the same. However, that is not always the case. In the case of arc 11 the generalized maximum cutpoint would change from .08 $[(1), (2, \ldots, 6)]$ to .085 $[(1, 3), (2, 4, \ldots, 6)]$. The limitation of this approach is that it ignores the cumulative probability distribution.

8.3. Contributions

This paper has a number of contributions. First, this paper provides a basis for the validation of computational models that employ multiple knowledge bases.

Second, it was argued that probability distributions generated for multiple knowledge base systems should not arbitrarily be averaged. If there are substantial differences in distributions, then that is likely to indicate that additional knowledge acquisition is required, there is an error or there are different models being used. It is up to the developer to determine which is the case. Alternatively, the model can be structured so that the user is asked to choose between disparate probability distributions.

Third, different metrics were proposed to measure the extent of those differences and try to determine the existence of conflict. Correlation coefficients were used to establish the extent to which different discrete distributions provided in two different knowledge bases, were related. Cutpoints were used to investigate the similarity of distributions. In one situation a double zero cutpoint was developed, indicating that the two distributions could be partitioned into two mutually exclusive distributions. Cutpoints appear to be a useful tool in the investigation of the similarity of probability judgments for multiple agent systems.

Fourth, these concepts were illustrated in the context of a case study. Fifth, the use of Bayes' Nets and conflict analysis discussed here were extended to organizational social network problems.

Acknowledgments

The author would like to acknowledge the comments of the anonymous referees and Kathleen Carley on an earlier version of this paper. An earlier version of this paper was presented at the 1995 Spring Workshop on Mathematical and Computational Organizational Theory, Los Angeles, April 1995. The author would like to thank the participants for their comments on an earlier version of this paper.

Note

1. The author would like to thank an anonymous referee for pointing out this application and directing the author to this literature.

References

Adrion, W., M. Branstad and J. Cherniavsky (1982), "Validation, Verification and Testing of Computer Software," *ACM Computing Surveys*, 14(2), 159–192.

Baligh, H., R. Burton and B. Obel (1996), "Organizational Consultant: Creating a Usable Theory of Organizational Design," *Management Science*, 42(12), 1648–1662.

Botten, N., A. Kusiak and T. Raz (1989), "Knowledge Bases: Integration, Verification and Partitioning," *European Journal of Operational Research*, 42(1), 111–128.

Brown, C., N. Nielson, D. O'Leary and M.E. Phillips (1995), "Validating Heterogeneous and Competing Knowledge Bases Using a Blackbox Approach," *Expert Systems with Applications*, 9(4), 591–598.

Burton, R. and B. Obel (1984), *Designing Efficient Organizations*. North Holland, Amsterdam.

Burton, R. and B. Obel (1995), "The Validity of Computational Models in Organizational Science: From Model Realism to Purpose of the Model," *Computational and Mathematical Organization Theory*, 1(1), 57–72.

Carley, K. (1996), "Validating Computational Models," Working paper, Department of Social and Decision Sciences, Carnegie Mellon University.

Cyert, R. and J. March (1992), *A Behavioral Theory of the Firm*. Blackwell Business, Cambridge, MA. Reprinted 1994.

Duda, R., P. Hart and N. Nillsson (1976), "Subjective Bayesian Methods for Rule-Based Inference Systems," *National Computer Conference*, 1075–1082.

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

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

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

Edgington, E. (1980), *Randomization Tests*, Marcel Dekker.

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

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

Freund, J. (1971), *Mathematical Statistics*, Prentice-Hall, Englewood Cliffs, NJ.

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

Gasser, L. and R. Hill (1990), "Coordinated Problem Solvers," *Annual Review of Computer Science*, 4, 203–253.

Gelernter, D. (1992), *Mirror Worlds*, Oxford, New York.

Gruber, T. (1993), "A Translational Approach to Portable Ontologies," *Knowledge Acquisition*, 5(2), 199–220.

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

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

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

Krackhardt, D. (1990), "Assessing the Political Landscape: Structure, Cognition and Power in Organizations," *Administrative Science Quarterly*, 35, 342–369.

Krackhardt, D. (1991), "The Strength of Strong Ties," in N. Nohira and R. Eccles (Eds.) *Organizations and Networks: Theory and Practice*, Cambridge MA: Harvard Business School.

LeClair, S. (1985), "Justification of Advanced Manufacturing Technology Using Expert Systems," in N. Botten and T. Raz (Eds.) *Expert Systems*, Norcross, GA: Industrial and Management Press, pp. 191–216.

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

Ng, K. and B. Abramson (1994), "Probabilistic Multi-Knowledge Base Systems," *Journal of Applied Intelligence*, 4(2), 219–236.

Ngwenyama, O. and N. Bryson (1992), "A Formal Method for Analyzing and Integrating the Rule Sets of Multiple Experts," *Information Systems*, 17(1), 1–16.

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

O'Keefe, R. and D. O'Leary (1993), "Expert System Verification and Validation," *Artificial Intelligence Review*, 7, 3–42.

O'Leary, D. (1986), "Multiple Criterion Decision Making Evaluation Functions in Expert Systems," *The 6th International Symposium on Expert Systems and Their Applications*, Avignon, France, pp. 1017–1035.

O'Leary, D. (1988a), "Methods of Validating Expert Systems, " *Interfaces*, 18(6), 72–79.

O'Leary, D. (1988b), "Expert System Prototyping as a Research Tool," in E. Turban and P. Watkins (Eds.) *Applied Expert Systems*, North Holland, pp. 17–32.

O'Leary, D. (1995), "Verification of Multiple Agent Knowledge-Based Systems," *International Joint Conference on Multiple Agent Systems*, San Francisco.

O'Leary, D. (1997), "Impediments in the Use of Explicit Ontologies for KBS Development," *International Journal of Human-Computer Studies*, 46, 327–337.

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

Reboh, R. (1982), "Extracting Useful Advice from Conflicting Expertise," *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pp. 319–325.