확률적 디폴트 규칙들을 이용한 비단조 상속추론 시스템

이 창 환†

요 약

상속추론(inheritance reasoning)은 인공지능의 추론분야에서 널리 사용되고있는 방법이다. 지금까지 인공지능 분야에서 다수의 상속추론 시스템이 제안되어 있지만, 대부분은 명확한 의미(semantics)가 정의되어있지 않으며, 이에 따라 부정확한 추론결과를 생성하는 원인이 되어왔다. 본 논문에서는 집합이론을 이용한 새로운 상속추론 시스템의 모델을 제안한다. 상속추론에 사용되는 규칙들의 의미는 통계적 분석에 의하여 정의되며, 규칙들의 확률은 과거 데이터의 값들에 의하여 정의된다. 특수성(specificity) 및 일반성(generality)의 두 가지 기본 추론 규칙이 제안되었으며 이들을 사용하여 기존의 상속추론 방법들의 오류들을 보정할 수 있음을 보였다. 제안된 상속추론 시스템의 알고리즘을 제시하고 전형적인 예제를 통하여 어떻게 비정상적인 결론 도출을 예방할 수 있는지를 보였다.

A Nonmonotonic Inheritance Reasoner With Probabilistic Default Rules

Chang-Hwan Lee†

ABSTRACT

Inheritance reasoning has been widely used in the area of common sense reasoning in artificial intelligence. Although many inheritance reasoners have been proposed in artificial intelligence literature, most previous reasoning systems are lack of clear semantics, thus sometimes provide anomalous conclusions. In this paper, we describe a set-oriented inheritance reasoner and propose a method of resolving conflicts with clear semantics of defeasible rules. The semantics of default rule is provided by statistical analysis of χ^2 method, and likelihood of rule is computed based on the evidence in the past.

Two basic rules, specificity and generality, are defined to resolve conflicts effectively in the process of reasoning. We show that the mutual tradeoff between specificity and generality can prevent many anomalous results from occurring in traditional inheritance reasoners. An algorithm is provided, and some typical examples are given to show how the specificity/generality rules resolve conflicts effectively in inheritance reasoning.

1. Introduction

Inheritance reasoning appears in various hierarchical forms in the literature of artificial intelligence and is becoming more important because inheritance hierarchies are simple, natural, and useful. Most inheritance reasoners have been characterized by algorithms which operate on an inheritance network, and some of them present translations of the inheritance network into some nonmonotonic logics, such as Reiter's

† 정 회 원 : 동국대학교 정보통신학과 교수
논문접수 : 1998년 10월 26일, 심사완료 : 1998년 11월 23일

default logic [1], Moor's autoepistemic logic [2], and circumscriptive theories [3] [4]. Although many inheritance reasoners have been proposed in artificial intelligence literature(e.g., [5] [6]), most previous works are missing clear semantics which is the central reason why these reasoners sometimes provide anomalous conclusions.

In this paper, instead of converting network structure to certain theories, we describe an inheritance reasoner and propose a method of resolving conflicts with clear semantics of defeasible rules. The semantics of default rule is provided by statistical analysis, and likelihood of rule is computed based on the evidence in the past. The χ^2 method is used to decide the existence of defeasible rules between predicate classes. In order to get presumably appropriate probabilities for each rule, probabilities are defined based on the observations of the occurrences in the past, and the probabilities are represented as intervals instead of single point values. These statistically induced degrees of beliefs have an advantage over the subjective probabilities since they are based on objective information about the world, information which could in principle be obtained through experience/observation.

Two basic rules *specificity* and *generality*, are defined to resolve conflicts effectively in the process of reasoning. First, like many other inheritance reasoners, the model uses a specificity rule as a tool to resolve conflicts. However, specificity is interpreted differently from the viewpoint of set cardinality. Second, generality rule, a complementary criteria of specificity, is defined based on the degree of biasedness of the observations in predicate classes. As a complementary rule of specificity, the generality rule is defined as a measure of the degree of typicality of instances of each class. The basic idea behind the generality rule is that the set of instances in the class must correctly represent the general characteristic of the class they belong to. We studied two conditions of which the instances in a class must avoid in order to maintain the minimum degree of generality: immature class and biased class.

2. Probabilistic Defeasible Assertion

Exception handling has been identified as the major issue of many inheritance reasoners. For example, the defeasible inference that Clyde is gray since he is an elephant and elephants are typically gray. As has been discussed in the literature, such defeasible assertions can not be modeled as universally quantified assertions since there are some elephants that are not gray. However, even though these exceptional cases falsify a universal quantifier, they do not invalidate the entire defeasible assertions. Without a well defined semantics for the defeasible assertions, it is very difficult, maybe even impossible, to develop an inheritance reasoner which will provide justifiable inferences in all cases. It is clearly impossible for an inheritance reasoner to generate correct or reasonable answers in all cases. However, even if we may occasionally be wrong we would still like to have some global justification for all of the inferences that the system generates.

The proposed system interprets the defeasible assertions as being statistical assertions. In order to determine the degree of belief of assertion, a good approach is to take advantage of known facts. A fundamental assumption of this probability is, "The probability of an indicative conditional of the form *if A is the case then B is* is that the probability of *if A then B* should be equal to the ratio of the probability of *A and B* to the probability of *A*(ratio of conjunction of antecedent and consequent to antecedent)." For example, in assertion "Most of birds fly", we interpret this assertion based on statistical evidence. In other words, among the instances of birds, how many of them can fly. Roughly speaking, if there are X objects with property α, and Y of these have property β, then for any term t, $P(\beta(t)|\alpha(t)) = Y/X$, unless we have some additional information about the term t. For formulas like "Fly(Clyde)," if 90% of all birds fly and Clyde is known to be a bird, the inductive assumption would attach a degree of belief of 0.9 to the formula "Fly(Clyde)" by assuming that Clyde was a randomly selected bird.

Statistical semantics was introduced by Bacchus [7]. He introduced statistical majority probabilities. In his study, if more than half the objects among L_1 satisfy L_2, a rule $L_1 \rightarrow L_2$(L_1 is usually L_2) is defined. It is possible that the probabilities of some rules coincidentally become larger than a constant ε and thus they become defeasible rules. Another problem occurs when the set cardinality of the precedent class is very small. When the cardinality of a class domain is very small, the probability of getting an anomalous conclusion is very high. Geffner has studied another probabilistic reasoning, called ε-semantics reasoning [8]. He has considered probabilistic versions of inheritance reasoning. However, he used probabilities infinitesimally close to 1 and 0. Obviously, there are very rare cases in which the properties are related via infinitesimal probabilities.

3. Evidential Probabilities

To make the degree of belief of default assertions more precise, we adopt an explicit negation system. Most reasoning systems regard unknown states as false states for the sake of efficiency. We claim that *false* states are different from *unknown* states. Various methods for implicit negation have been proposed such as unique-name assumption, domain closure, closed world assumption, and predicate completion [9]. The main issue common in these strategies is the efficiency of representing knowledge, sacrificing expressive power for the sake of efficiency. In this paper, we use *explicit negation* to differentiate unknown from false. Notice that, as we mentioned earlier, the semantics of the system is totally based on current evidence. Therefore, if we do not differentiate false from unknown, we start with incorrect evidence.

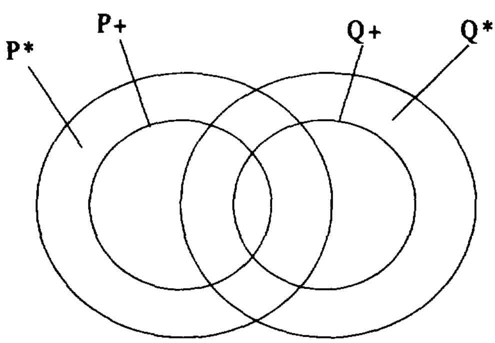
Each rule in the system comes with its corresponding probability, which is generated automatically. In most current reasoning systems, an expert assigns his/her subjective probability to each rule. One of the big problems in this approach is the impreciseness of the value of the probability assigned. There is also an issue of whether it is reasonable to describe probability by a single point rather than a range. While an agent might agree that the probability of an event lies within a given range, say between 1/3 and 1/4, he might not be prepared to say that it is precisely 0.287. In the proposed system, the probabilities are represented as intervals instead of point values. Introducing unknown states allows us to define the likelihood of a proposition A as a subinterval of the unit interval [0,1]. The lower bound of this interval is the degree of *support* of the proposition, S(A), and the upper bound is its degree of *plausibility*, Pl(A). The likelihood of an assertion A is written as [S(A), Pl(A)]. The support of A is meant to describe a lower bound on the degree of belief of an agent that A is actually the case. The corresponding upper bound is called plausibility. Intuitively, we view the interval [S(A), Pl(A)] as providing lower and upper bounds on the "likelihood" of A. The detailed method of calculating these probabilities is explained in the following subsection.

3.1 Support and Plausibility

This section introduces the basic framework of the reasoner. The system consists of three components: a set of objects, a set of predicates, and a set of edges. We have constant symbols which represent the members of the object, predicate symbols, and two types of edges. When an object belongs to a predicate, there is a strong edge from the object to the predicate. Similarly, if there is a strong or defeasible rule between two predicates, there is a strong edge or defeasible edge, respectively, between the two predicates. Reasoner Γ is being represented as follows.

$$\Gamma = < \boldsymbol{D}, \boldsymbol{P}, \boldsymbol{E} >$$

where $\boldsymbol{D}$ shows the domain of universe, $\boldsymbol{P}$ shows the set of predicates, and $\boldsymbol{E}$ are the edges of the network. As we mentioned, there are two types of edges : strong edges and defeasible edges. A strong edge will have the form $x \Rightarrow y$, or $x \not\Rightarrow y$, where y is a class. If x is an object, such an assertion would be interpreted as an ordinary atomic statement : for instance, they are analogous to $y(x)$ or $\neg y(x)$ in logic. They might represent statements like "Tweety is a bird" or "Tom isn't a Bird." If x is a class, these assertions would be interpreted as generic statements. For example, $x \rightarrow y$ and $x \not\rightarrow y$ might represent the statements "Birds fly" and "Penguins don't fly," respectively. In addition, as we mentioned earlier, the edges are accompanied by probabilities such that the general form of an edge is $P_1 \rightarrow P_2$:[$S(P_1 \quad P_2)$, $Pl(P_1 \quad P_2)$]. Values $S(P_1 \rightarrow P_2)$ and $Pl(P_1 \rightarrow P_2)$ represent support probability and plausibility probability, respectively (In this paper, we use $\Pr(P_1 \rightarrow P_2)$ instead of $\Pr(P_2 \mid P_1)$. Especially, when $S(P_1 \rightarrow P_2) = Pl(P_1 \rightarrow P_2) = 1$, we say $P_1 \Rightarrow P_2$).



(Fig. 1) Set notations for probability definitions

To differentiate unknown state from false state, we adopt the following definitions. Each predicate has three subsets: P^+, P^-, and P^*. P^+ contains objects known to have properties of P, P^- contains objects known to have properties of $\neg$P, and P^* contains objects which are inconclusive about P. Graphical notations of these sets are shown in (Fig. 1). A semantic structure for a network is an assignment of a triple of sets of objects (P^+, P^-, P^*) for each node in $\boldsymbol{P}$. The support and plausibility are given as

$$S(P \rightarrow Q) = \min\{Pr(Q|P)\} = \frac{\min\{\Pr(Q \wedge P)\}}{\max\{\Pr(P)\}}$$

$$\mathrm{Pl}(P \rightarrow Q) = \max\{Pr(Q|P)\} = \frac{\max\{\Pr(Q \wedge P)\}}{\min\{\Pr(P)\}}$$

Therefore, the support and plausibility are formally defined as follows($\|\cdot\|$ represents the cardinality of a set).

Definition 1

$$S(P \rightarrow Q) = \frac{\|P^+ \cap Q^+\|}{\|P^+\| + \|P^* \cap Q^-\| + \|P^* \cap Q^*\|}$$

$$Pl(P \rightarrow Q) = \frac{\|(P^+ \cup P^*) \cap (Q^+ \cup Q^*)\|}{\|P^+\| + \|P^* \cap Q^+\| + \|P^* \cap Q^*\|}$$

4. Detecting the Existence of Defeasible Rules

One of the most prominent features of the proposed reasoner is the separation of detecting a rule and calculating the probability of the rule. In this section, we show how the system detects the existence of rules and how their corresponding probabilities are computed. Suppose there are two predicate A, B. We will see how the existence of the defeasible rule between A and B is detected and how the probability of the detected rule is calculated.

We shall examine the chi-square(χ^2) method of testing the hypothesis that the two predicates are dependent on each other. In most reasoners, the existence of a defeasible rule is decided (1) by an expert or rational agent, or (2) if the ratio of consequent and antecedent is greater than a certain constant, they define it as a defeasible rule. We use a different approach to define dependency. First we differentiate the existence of the defeasible rule from

A^+	180 (50)	15 (119.4)	5 (30.6)	200
A^-	65 (190)	550 (453.7)	145 (116.3)	760
A^*	5 (10)	32 (23.9)	3 (6.1)	40
Total	250	597	153	1,000

(Fig. 2) χ^2 table for detecting defeasible rule

the probability of the rule. That is, even though the success probability between two predicates is high, it is still possible that they do not have a causal relationship. In the proposed approach, the existence of defeasible rule is decided by statistical analysis. A chi-square test is usually used to test the hypothesis that the observations agree or disagree with the theoretical frequencies [10]. In our case, it can be used to test whether the domain sets of two predicates are inter-dependent or not. The statistic we will use is

$$\chi^2 = \sum_{i,j \in \{+,-,*\}} \frac{(f_{ij} - F_{ij})^2}{F_{ij}}$$

where f_{ij} is the observed frequencies, and F_{ij} is the theoretical frequencies. For example, in Figure 2, the theoretical frequency of $A^+ \cap B^+$ is given as $250 \cdot \frac{200}{1000} = 50$. Other theoretical frequencies can be calculated in a similar way. To determine the dependency between A and B, the total χ^2 value of (Fig. 2) is computed as 566.58 and is compared with the value of $\chi^2_{[4, 0.95]}$, where $\chi^2_{[4, 0.95]}$ is the χ^2 value with degree of freedom being 4 and level of significance begin 0.95. As the value of $\chi^2_{[4, 0.95]}$ is given as 9.49, we can conclude that A is not independent of B. The other issue concerning the detecting rule is how to decide the direction of the detected rules. We employ the following strategy in the system.

$$P \rightarrow Q : P^+ \cap Q^+ \geq P^+ \cap Q^-$$
$$Q \rightarrow P : \textit{otherwise}$$

According to Definition 1, the probability interval of the above rule is given as follows.

$$[S(A \rightarrow B),\ Pl(A \rightarrow B)] =$$
$$[\frac{180}{200+32+3}, \frac{180+5+5+3}{200+5+3}] = [0.76,\ 0.92]$$

Sometimes, there are cases that the predicate A is a subset of the predicate B, which means there exists a strong rule between A and B. In this case, $A^+ \cap B^-$ and $A^+ \cap B^*$ in the above table are empty, and we can easily figure out that the system produces a strong rule between A and B.

The semantics for each defeasible or strict rule are given in the following. Suppose c denotes an instance and P, Q denote predicates.

1. c $\Rightarrow$ P is true iff $P^+(c)$
2. c $\not\Rightarrow$ P is true iff $P^-(c)$
3. P $\rightarrow$ Q : [S(P$\rightarrow$Q), Pl(P$\rightarrow$Q)], iff $\chi(P,Q) > \chi^2_{[4,0.95]}$ and $P^+ \cap Q^+ \geq P^+ \cap Q^-$
4. P $\not\rightarrow$ Q : [S(P$\not\rightarrow$Q), Pl(P$\not\rightarrow$Q)], iff $\chi(P,Q) > \chi^2_{[4,0.95]}$ and $P^+ \cap Q^+ \geq P^+ \cap Q$
5. P $\Rightarrow$ Q is true iff P $\rightarrow$ Q and $P^+ \subseteq Q^+$
6. P $\not\Rightarrow$ Q is true iff P $\not\rightarrow$ Q and $P^+ \cap Q^+ = \emptyset$

5. Inferences

Inferences are performed based on the hierarchical structure generated from the data. Inheritance reasoning in this case can be regarded as a classification problem given an available data set. Our claim in this system is that the most fundamental tools in classification reasoning are specificity and generality. Specificity has already been proposed in many inheritance reasoning systems while generality is a new inference tool introduced in this paper.

5.1 Specificity

Many inheritance reasoners use specificity as a tool for selecting natural preference criterion. This preference criterion is based on simple intuition: the more knowledge that is used to generate the degree of belief, the better is the degree of belief. Sentence α represents more knowledge than β if $\alpha \Rightarrow \beta$ is deducible from the knowledge base. For instance, the fact that "Royal_Elephant(Clyde) ⇒ Elephant(Clyde)" indicates that the knowledge Royal_Elephant(Clyde) should be preferred when including a degree of belief in Gray_Thing(Clyde). The idea of specificity was first given by Touretzky in the form of inferential distance [11]. Since then, many variations of specificity have appeared in the literature. The basic idea behind specificity-based algorithms including shortest distance is that the more specific the information is, the more precise the result is.

The system also adopts specificity as one of the tools for resolving conflicts. While the definition of specificity in other inheritance reasoners is based on network path or standard nonmonotonic logic, we are interpreting the specificity in terms of set relationships among classes. Suppose we have two conflicting reference classes, say A and B, and class A is a subset of class B(A is more specific than B), we choose A as the reference class. This is basically the same as the way in which the conventional inheritance reasoners solve conflicts. Now, the question is what if A has nothing to do with B ? In this case, B is irrelevant to class A and we have conflicts. In the system, it creates a derived class which is the conjunction of A and B, and that derived class becomes the target reference class, which is obviously more specific than either A or B. By doing this, to get the most specific reference class, the system does specification using set conjunction as many times as it can.

However, specificity alone can not solve all conflict problems. Let us pay a second visit to the above example. Assume that the class Elephant has only 3 instances, and all these instances are known to be not gray. It is not safe to say that Elephants are usually not gray. Where does this problem come from ? The reason is that the size of the class Elephant is so small that it can not represent correctly the characteristic of the class in the real world. The smaller the number of instances for a class is, the less likely the instances can represent the property of the real class correctly. Thus, if the reference class, whether original or derived, does not have enough instances, it becomes an immature class, and thus be prohibited to be a reference class because it does not have enough statistical information. The following section investigates this problem in more detail.

5.2 Generality

Generality is a complementary criteria of specificity. The basic idea behind the generality concept is that the more information we have, the more correct estimates we can have. Speaking in terms of class, when we have more sample instances, we have higher quality guesses. The instances of the class must have the generality which can represent the general characteristic of the class they belong to. We propose two conditions which the samples in a class must satisfy to maintain the minimal degree of generality of a class. In case a class has a very small number of samples, it is quite possible that these samples do not correctly represent the general characteristic of the class, thus lacking generality. Even though the cardinality is large enough to avoid the immature class problem, we have biased class problem if the sample is not evenly collected from its subpredicates. These requirements are discussed in more detail in the following.

For example, suppose we have a rule "Koreans are tall," "Seoul residents are not tall," and "Kim is a Seoul resident." If the instances of Seoul about property *tall* are very few, we can not conclude that Jim is not tall based on the Seoul class. When the cardinality of Seoul class is very small, it is quite possible that this small sample does not represent correctly the characteristic of Seoul residents. Then, how large should the class size be ? We need a quantitative study for deciding the cardinality of classes.

We should consider two kinds of predicates, leaf nodes and intermediate nodes, since each of these has different characteristics of minimum cardinality.

First, for leaf predicates, each class, say A, is divided into three subgroups(A^+, A^-, and A^*). Among them, we can consider a set $A^+ \cup A^-$ as a sample of class A. Considering the predicate A as a variable, we can easily see that the distribution of A follows a binary distribution. Statistically speaking, the problem can be regarded as selecting an appropriate sample size in estimating the proportion of A. In this type of predicate, we have dichotomous values, 'yes' or 'no', and thus we only need to estimate the proportion of the first value.

We will consider the following situation such that restricting to an acceptable level the probability that the difference between population mean P of predicate A and sample mean p of A is greater than a specified value. Let n denote the sample size of the population. Note that p is defined as $\frac{\|A^+\|}{\|A^+ \cup A^-\|}$, and n is denoted as $\|A^+ \cup A^-\|$ where $\|\cdot\|$ means cardinality function. N is the cardinality of domain of the universe, which is denoted as $\|\boldsymbol{D}\|$.

If the permissible error in the estimate of the population value of the mean is d and the degree of assurance desired is $1-\alpha$, then the following inequality holds.

$$\Pr(|P-p|>d) \le \alpha$$

According to *the central limit theorem* [12], p follows the normal distribution, i.e., $p \sim N(P,(1-f)\sigma^2(i)/n)$. Using t-distribution analysis, the above equation is rearranged to conclude

$$n \ge N[1+N(\frac{d}{t_\alpha S})^2]^{-1}$$

In particular, when we have a very large number of objects in our domain, the approximate size of a sample can be decided as follows.

$$n > \frac{t_\alpha^2 p(1-p)}{d^2}$$

Here we say that n is the minimum cardinality of A to avoid the immature class. If the class cardinality is less than the above number, the corresponding class is considered to be an immature class. One drawback of this approach is that when the size of the original population is not large enough, it needs a correct value of the size of the population, which we have to estimate.

Secondly, we will investigate the intermediate predicates. Intermediate predicates are defined to have a set of subpredicates. If there is a strong edge from predicate P to predicate Q, we define the predicate P as the subpredicate of Q. For an intermediate predicate Q and a set of subpredicates of Q, say P_1, P_2, ... P_k, these subpredicates partition the domain of universe into k disjoint subsets. In this case, predicate Q is called the parent predicate. If we consider each of these subpredicates as a category value of discrete variable, the distribution of predicate Q follows a multinomial distribution. For a parent predicate, computing the minimal cardinality involves a multinomial distribution, which means that we need to consider the statistical distributions of all subpredicates simultaneously. To estimate the minimal cardinality of the intermediate predicate, we need a correct value of the size of each subpredicate. Suppose $\|P_j^{'+}\|$ and $\|P_j^{'-}\|$ represent the cardinality of P_j^+ and P_j^-, respectively, in the real population. $\|Q_j^{'+}\|$ and $\|Q_j^{'-}\|$ are defined in a similar way. Let w_j and W_j be defined, respectively, as $w_j=\|P_j^+\|/\|Q_j^+\|$ and $W_j=\|P_j^{'+}\|/\|Q_j^{'+}\|$. The proportion of each subgroup w_j can be represented simultaneously, for a predicate Q and jth value assignment in Q. The question is how to derive the smallest cardinality n for a random sample from a multinomial population so that the probability would be at least $1-\alpha$ and that all of the estimated proportions would simultaneously be within specified distances of the true population proportions. This constraint can be expressed by the following equation.

$$\Pr(\bigcap_{j=1}^{k}|W_j-w_j|>d_j) \ge \alpha$$

For an individual parameter P_j, the probability that the sample value in the table lies outside the specified interval is

$$\alpha = \Pr(|Z_j| \geq d_j\sqrt{n}/\sqrt{w_j(1-w_j)}) = 2(1-\Phi(z_j))$$

where Z_j is a standard normal random variable, Φ is the cumulative standard normal distribution, and $z_j = d_j\sqrt{n}/\sqrt{w_j(1-w_j)}$.

6. Algorithm

We will now define the reasoning algorithm and discuss the basic ideas which motivated this particular method. Before we describe the algorithm, a set of rules are specified in the following. The probability for a strict rule is omitted for the sake of simplicity. By definition, strict rules($\Rightarrow$) contain the probability of [1,1]. If we use only **[AX0]-[AX5]**, we can do monotonic deductive inference because rules **[AX0]-[AX5]** are sound. However, monotonic deductive inference alone cannot do a large amount of useful nonmonotonic reasoning. The seven reasoning axioms proposed in this paper are described in (Fig. 3).

The algorithm takes network information and a focus node s as input, and generates conclusions with probability for all of the destination nodes reachable from the focus node. As we have mentioned, immature or biased class is apt to generate anomalous results. Therefore, these classes are deleted in advance from the network. The algorithm then proceeds by first finding a set of nodes(N_s) which the focus node belongs to, and delete nodes, within N_s, which s doesn't belong to. Now for each node P in N_s, calculate all possible paths reachable from P using **[AX0]-[AX6]**. The approach we present in the algorithm has the characteristic of credulous extension. Credulous extension of an inheritance hierarchy Γ with respect to a node s is a maximal unambiguous a-connected subhierarchy of Γ with respect to a. When there is a path $s \Rightarrow E$ or $s \Rightarrow \cdots \Rightarrow E$, then this path substitutes all paths from s to E. If there are two positive(negative) paths which share the same intermediate and end node, delete the redundant paths(resp. negative). Given two paths

$$s \Rightarrow A_1 \overset{+}{\underset{*}{}} A_2 \overset{+}{\underset{*}{}} \cdots A_k \overset{+}{\underset{*}{}} E$$

and

$$s \Rightarrow B_1 \overset{+}{\underset{*}{}} B_2 \overset{+}{\underset{*}{}} \cdots B_l \overset{-}{\underset{*}{}} E$$

repeat the procedure described in (Fig. 3) until there is only one path left for each destination node

[AX0] If $\{c \Rightarrow P_1,\ P_1 \Rightarrow P_2\},\ c \Rightarrow P_2$
[AX1] If $\{c \Rightarrow P_1,\ P_1 \not\Rightarrow P_2\},\ c \not\Rightarrow P_2$
[AX2] If $\{P_1 \Rightarrow P_2,\ P_2 \Rightarrow P_3\},\ P_1 \Rightarrow P_3$
[AX3] If $\{P_1 \Rightarrow P_2,\ P_2 \not\Rightarrow P_3\},\ P_1 \not\Rightarrow P_3$
[AX4] If $\{P_1 \rightarrow P_2{:}[\alpha,\beta],\ P_2 \Rightarrow P_3\},\ P_1 \rightarrow P_3{:}[\alpha,\beta]$
[AX5] If $\{P_1 \rightarrow P_2{:}[\alpha,\beta],\ P_2 \not\Rightarrow P_3\},\ P_1 \not\rightarrow P_3{:}[\alpha,\beta]$
[AX6] If $\{P_1 \Rightarrow P_2,\ P_2 \rightarrow P_3\}$, $\{P_1 \rightarrow P_2,\ P_2 \rightarrow P_3\}$, $\{P_1 \Rightarrow P_2,\ P_2 \not\rightarrow P_3\}$ or $\{P_1 \rightarrow P_2,\ P_2 \not\rightarrow P_3\}$,

1. $P_1 \rightarrow P_3{:}[S(P_1 \rightarrow P_3), PI(P_1 \rightarrow P_3)]$,
   if $\chi^2(P_1, P_3) > \chi^2_{[4,0.95]}$ and $P_1^+ \cap P_3^+ \geq P_1^+ \cap P_3^-$
2. $P_1 \not\rightarrow P_3{:}[S(P_1 \not\rightarrow P_3), PI(P_1 \not\rightarrow P_3)]$,
   if $\chi^2(P_1, P_3) > \chi^2_{[4,0.95]}$ and $P_1^+ \cap P_3^+ < P_1^+ \cap P_3^-$
3. otherwise, there is no edge between P_1 and P_3

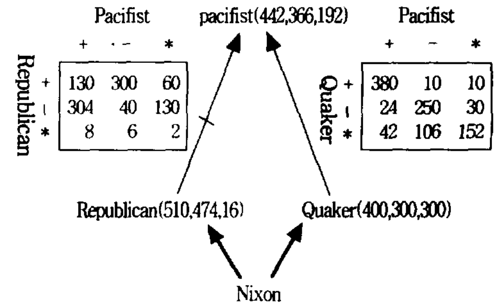
(Fig. 3) Reasoning axioms

($\overset{+}{*}$ represents {⇒ or →}). Now only one path is assigned to each destination node in the network. Furthermore, we have temporary node sets, which are formulas closed under union and conjunction. To calculate the probabilities of each path which are defined based on these temporary nodes, probability definitions in Definition 1 must be extended. For predicates A and B, the following set operations can be easily understood to accommodate the extended formula.

7. Examples

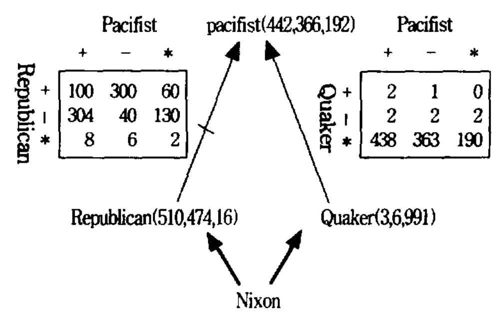
This section shows a couple of examples which show how the specificity or generality works in these cases. Nixon's diamond, shown in (Fig. 4), has been largely used to show the ambiguity of inheritance reasoning. Nixon is a Republican and at the same time a Quaker. Since $\chi^2(R,P)$ = 1161 is greater than $\chi^2_{[4,0.95]}$ = 9.49 and $R^+ \cap P^+ < R^+ \cap P^-$, there is a rule "R(Republican) is usually not P (Pacifist)" based on the semantics of rules. Similarly, a rule "Q(Quaker) is usually P(Pacifist)" exists since $\chi^2(Q,P)$ = 923 is greater than 9.49 and $Q^+ \cap P^+ < Q^+ \cap P^-$. Therefore, Republicans are usually not -Pacifist while Quakers are Pacifists. In this case, skeptical reasoning does not generate any conclusion because conflict occurs at Pacifist, while credulous reasoning provides two extensions: Nixon ⇒ Quaker → Pacifist and Nixon ⇒ Republican ↛ Pacifist, and does not give any preference among these extensions. In the proposed approach, the diagram will be transformed into one of the following paths.

1. Nixon → (Quaker ∩ Republican) → Pacifist, if
   - χ^2(Quaker ∩ Republican, Pacifist) > $\chi^2_{[4,0.95]}$
   - (Quaker$^-$ ∩ Republican$^+$ ∩ Pacifist$^+$) ≥ (Quaker$^+$ ∩ Republican$^-$ ∩ Pacifist)
2. Nixon → (Quaker ∩ Republican) ↛ Pacifist, if
   - χ^2(Quaker ∩ Republican, Pacifist) > $\chi^2_{[4,0.95]}$
   - (Quaker$^+$ ∩ Republican$^-$ ∩ Pacifist$^+$) < (Quaker$^+$ ∩ Republican$^+$ ∩ Pacifist$^-$)
3. There is no edge between Nixon and Pacifist, if
   - χ^2(Quaker ∩ Republican, Pacifist) ≤ $\chi^2_{[4,0.95]}$.



(Fig. 4) Nixon's diamond

We present another version of the Nixon example in (Fig. 5) to show the problem of immature class. As we have seen in (Fig. 4), "R(Republican) is usually not P(Pacifist)." Similarly, "Q(Quaker) is usually P(Pacifist)" because $\chi^2(Q,P)$>9.49 and $Q^+ \cap P^+ < Q^+ \cap P^-$. However, the set for Quaker is (3, 6, 991), which means only nine elements are in the class Quaker as samples. In this case, because it does not satisfy the minimum cardinality requirement, we just disregard the class Quaker because it is an immature class. Therefore, the system concludes that Nixon is not Pacifist.



(Fig. 5) Nixon's diamond with immature class

8. Conclusion

In this paper, we proposed a set-oriented statistical model for inheritance reasoner. The χ^2 method using contingency table is used to decide the existence of rules. Due to the introduction of unknown state, likelihood of each rule is given as an interval. Among the features of the general inheritance reasoner, we focused on the way of resolving conflicts. Two basic inference tools were introduced and explained to resolve conflicts in inheritance reasoning. Specificity, one of the most common inference tools, is interpreted differently from the viewpoint of set cardinality. As a complementary rule of specificity, the generality rule is defined as a measure of the degree of generality of the instances of each class. Generality plays the role of complementing the drawback of the specificity. The tradeoff between specificity and generality can prevent anomalous results caused by small cardinality and uneven distribution of instances.

References

[1] D. W. Etherington, *Reasoning with Incomplete Information*, Morgan Kaufmann, Los Altos, CA, 1988.

[2] H. Przymusinska and M. Gelfond, *Inheritance Hierarchies and Autoepistemic Logic*, Technical Report, Computer Science Department, University of Texas at El Paso, TX, 1988.

[3] B. A. Haugh, Tractable Theories of Multiple Defeasible Inheritance in Ordinary Nonmonotonic Logics, *Proceedings AAAI-88*, St. Paul, MN, 1988.

[4] T. Krishnaprasad, M. Kifer and D. S. Warren, On the Declarative Semantics of Inheritance Networks, in *Proceedings IJCAI-89*, Detroit, MI 1093-1103.

[5] F. Bacchus, *Representing and Reasoning about Probabilistic Knowledge*, MIT Press, 1990.

[6] L. A. Stein, Resolving Ambiguity in Nonmonotonic Inheritance Hierarchy, *Artificial Intelligence*, Vol.55, pp.259-310, 1992.

[7] F. Bacchus. A Modest, but semantically Well-Defined, Inheritance Reasoner, *Proceedings of 11th IJCAI*, 1989.

[8] H. Geffner, *Default Reasoning : Causal and Conditional Theories*, MIT Press, 1992.

[9] E. Davis, *Representation of Commonsense Knowledge*, Morgan Kaufman Publisher. 1990.

[10] H. Reichenbach, *Theory of Probability*, University of California Press, Berkeley and Los Angeles, CA, 1949.

[11] D. S. Touretzky, J. F. Horty, and R. H. Thomason, A clash of Intuitions : The current state of Nonmonotonic multiple inheritance systems, in *Proceedings IJCAI-87*, pp.476-482, 1987.

[12] V. Barnett, *Sample Survey Principles and Methods*, New York : Oxford University Press, 1991.

이 창 환

e-mail : chlee@cakra.dongguk.ac.kr

1982년 서울대학교 계산통계학과 졸업(학사)

1988년 서울대학교 계산통계학과 대학원 졸업(이학석사)

1994년 University of Connecticut 졸업(공학박사)

1994년~1995년 AT&T Bell Laboratories 연구원

1996년~현재 동국대학교 정보통신학과 조교수

관심분야 : 기계학습, 인공지능, 인공생명