DISCOVERING CAUSAL RULES IN RELATIONAL DATABASES

FLORIANA ESPOSITO, DONATO MALERBA, VINCENZA RIPA, and GIOVANNI SEMERARO
Dipartimento di Informatica, Università degli Studi di Bari, Bari, Italy

This article explores the combined application of inductive learning algorithms and causal inference techniques to the problem of discovering causal rules among the attributes of a relational database. Given some relational data, each field can be considered as a random variable, and a hybrid graph can be built by detecting conditional independencies among variables. The induced graph represents genuine and potential causal relations, as well as spurious associations. When the variables are discrete or have been discretized to test conditional independencies, supervised induction algorithms can be used to learn causal rules, that is, conditional statements in which causes appear as antecedents and effects as consequences. The approach is illustrated by means of some experiments conducted on different data sets.

As the amount of data stored in databases grows exponentially, the scientific community feels a greater need of sophisticated tools for analyzing and summarizing data. Since every data summarization carries the potential for generalization, which is a widely investigated topic in the area of machine learning, it is not surprising that a variety of learning algorithms have been embedded into systems for knowledge discovery in databases (Cercone & Tsuchiya, 1993; Piatetsky-Shapiro & Frawley, 1991). Such algorithms can find different types of patterns in the data, such as concept descriptions, taxonomies, and qualitative and quantitative laws (Matheus et al., 1993). However, the problem of finding data dependencies and, more specifically, causal dependencies has received greater attention in statistics (Asher, 1983; Glymour et al., 1987) than in the machine learning community.

Several approaches to *probabilistic* causal inference have been proposed in the literature (see Spirtes et al. (1993) for a large collection of algorithms). All of them work on attributional representations and find a graph whose nodes are the considered attributes or *variables,* while its edges represent causal dependencies among the variables. Henceforth, such a dag will be called *causal structure* or *causal model.* Any causal model is actually a syntactic object from which it is possible to derive some equations, or *constraints,* which are said to be *implied* by the model itself. The idea that is common to the best known approaches is that of searching for the causal structure that *implies* a set of constraints that best fits the set of constraints that hold in the data set (Esposito et al., 1994). Originally, Glymour et al. (1987) considered two types of constraints on correlations, namely, *partial* and *tetrad* equations, which

Address correspondence to Floriana Esposito, Dipartimento di Informatica, Università degli Studi di Bari, via Orabona 4, 70126 Bari, Italy. E-mail: esposito@vm.csata.it

Applied Artificial Intelligence, 11:71–83, 1997

0883-9514/97 $12.00 + .00

involve triplets and foursomes of correlation coefficients, respectively. Tetrad equations proved to be very useful for the study of the causal structure governing the set of *latent,* or unmeasured, variables that might affect those that are measured. A few years later, Spirtes et al. (1990) and Verma and Pearl (1990) showed how constraints defined by conditional independencies can be profitably exploited to infer causal structures. The advantage of conditional independencies over partial or tetrad equations is that there are several statistical tests for conditional independence of two variables of any scale (nominal, ordinal, interval, and real), while correlations can be computed only for continuous variables. Moreover, tetrad equations apply only to linear models, that is, causal structures where each effect can be defined as a linear combination of its direct causes. For this reason, where the causal discovery problem concerns real-world databases with mixed-mode data (continuous and discrete), constraints defined by conditional independencies appear to be more suitable.

In this article we describe how CAUDISCO (CAUsal DISCOvery), a system inspired by Pearl and Verma's (1991) theory of inductive causation, discovers causal rules in relational databases. Given a set of nominal or continuous variables, the system first finds significant independencies in the data and then looks for the best causal model explaining them. The output is a *partially oriented inducing path graph* (or *hybrid graph*), that is, a directed acyclic graph (dag) whose edges can be of four distinct types: nondirected (—), directed ($\Rightarrow$), partially directed ($\rightarrow$), and bi-directed ($\leftrightarrow$). The last three types of edges represent different causal relations, namely, genuine and potential causal relations and spurious associations between two variables. In a spurious association the existence of a latent variable that affects the two variables is hypothesized, but nothing is postulated about the causal relations among latent variables.

The causal structure inferred by CAUDISCO explains dependencies among observed variables but provides no information on how causes can influence their effects. In this article we also cope with the problem of learning *causal rules,* that is, conditional statements in which causes appear as antecedents and effects as consequences. In particular, given a causal structure, we define one or more inductive learning problems and we apply a well-known learning system to produce the causal rules. Some experimental results on artificial data and a database available in the University of California, Irvine (UCI) machine learning repository are reported in a later section.

USING CONDITIONAL INDEPENDENCIES FOR INFERRING CAUSAL STRUCTURES

Let $x, y, z_1, z_2, \ldots, z_n$ be $n + 2$ distinct variables taken from a set O of random variables with a joint probability distribution P. Then x and y are said to be conditionally independent in the context $z_1, z_2, \ldots, z_n$ if

$$P(x, y|z_1, z_2, \ldots, z_n = P(x|z_1, z_2, \ldots, z_n)\, P(y|z_1, z_2, \ldots, z_n)$$

Henceforth, we will write $I(x, y|z_1, z_2, \ldots, z_n)$ when x and y are conditionally independent in the context $\{z_1, z_2, \ldots, z_n\}$ of size n, and $\neg I(x, y| z_1, z_2, \ldots, z_n)$ when they are not.

For our purposes, the random variables in O are fields of one or more relations in a database. Since their value is known, they are called *observed* or *measured* in order to distinguish them from the *latent* variables, which represent some external factors that may affect the observed variables but are not included in O. Here we are deliberately neglecting the problem of missing values that strongly affect real databases.

Under *Markov* and *faithfulness* conditions (Spirtes et al., 1993), there exists a perfect correspondence between conditional independencies that hold for a probability distribution P over O and the independence relations determined by a dag over O. More precisely, given a dag, the set of all conditional independencies implied by the Markov condition over the dag can be characterized by a graphical criterion, called *d-separation* (Geiger & Pearl, 1989). Conversely, given a probability distribution P over O, for which the faithfulness condition holds, independence relations implied by the Markov condition are the only conditional independence relations true in P. When both conditions are satisfied, the dag is said to be a *perfect I-map* of a dependency model (Pearl, 1988). Consequently, a set of conditional independencies uniquely determines a dag G, whose edges may have a causal interpretation as reported below.

Pearl and Verma (1991) distinguish three types of causal relations on the grounds of conditional independencies: potential causal influence, genuine causal influence, and spurious association.

Definition 1: Potential Causal Influence

A variable x has a potential causal influence on another variable y if

1. x and y are dependent in every context [for each context S: $\neg I(x, y|S)$], and
2. there exists a variable z and a context S such that
   a. x and z are independent given S [$I(x, z|S)$]
   b. z and y are dependent given S [$\neg I(z, y|S)$].

Henceforth, a potential causal influence of x on y will be indicated by a partially directed edge, that is, $x \rightarrow y$. If there are no latent variables, then the only two situations that can occur are those reported in Figure 1, where the double arrow ($\leftrightarrow$) denotes a spurious association (see definition 3). Indeed, if y caused x, then x and z would be dependent in contrast with condition a. However, what guarantee is there that x and y have no latent common cause? If we were sure that all the variables

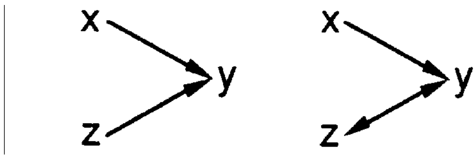


Figure 1. Two cases of potential causal influence.

involved in the phenomenon under study had been taken into account, then condition 1 would be sufficient to avoid this eventuality. Nevertheless, since we do not have this certainty, we can only postulate a *potential* cause of x on y, which does not exclude the possibility of a spurious association.

Definition 2: Genuine Causal Influence

A variable x has a genuine causal influence on another variable y if another variable z exists such that

1. either x and y are dependent in every context and there is a context S such that
   a. z has a potential causal influence on x
   b. z and y are dependent given S $[\neg I(z, y|S)]$
   c. z and y are independent given $S \cup \{x\}$ $[I(z, y|S \cup \{x\})]$
2. or x and y are in the transitive closure of rule 1.

The genuine causal association is represented by a directed edge, that is, $x \Rightarrow y$. Figure 2 shows the only two cases satisfying condition 1 when there are only three measured variables. Note that there is no way of inverting the direction of the edge from x to y without violating any of the conditions b and c. Condition 2 covers those cases in which there is a path from x to y whose edges represent genuine causal relationships oriented in the direction from x to y.

Definition 3: Spurious Association

Two variables x and y have a spurious association if they are dependent in some context S $[\neg I(x, y|S)]$ and two variables z_1 and z_2 exist such that

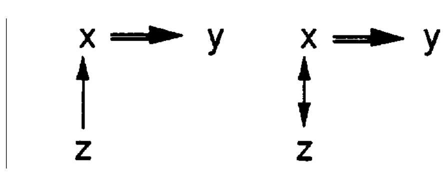


Figure 2. Two cases of genuine causal influence.

1. z_1 and x are dependent in S [$\neg I(z_1, x|S)$]
2. z_1 and y are independent in S [$I(z_1, y|S)$]
3. z_2 and x are independent in S [$I(z_2, x|S)$]
4. z_2 and y are dependent in S [$\neg I(z_2, y|S)$].

The spurious association between x and y is represented by a bi-directed edge, that is, $x \leftrightarrow y$. If there are no latent variables, and x, y, z_1,and z_2 are the only observed variables, then the causal model in Figure 3 represents the only possible situation that can occur. In this model, conditions 1 and 2 prevent x from causing y, while conditions 3 and 4 prevent y from causing x. Thus the dependence between x and y can only be explained by the presence of a latent variable affecting both x and y. In this example the context S is the empty set.

It is interesting to note that in these definitions, a nontemporal semantics of causation is given. This is not a limit but an advantage for those systems that aim to discover knowledge in archival data (Zytkow & Baker, 1991), since in most cases, information on the time in which a variable is measured is not available. In fact, although some database management systems put time stamps on data, they generally record the time at which data are entered into the database and not the time at which the measurement is performed.

Given a set of conditional independencies $I(O)$ found for each pair of observed variables, CAUDISCO operates as follows.

1. Variables that are dependent in any context are connected by an edge (—).
2. The definition of potential causal influence is applied to connected edges.
3. The transitive property of potential causal influence is applied.
4. The definition of genuine causal influence is used in order to mark some directed edges properly.

The output produced by CAUDISCO is able to explain the effect of a latent variable on a pair of observed variables, that is, the system is appropriate for those phenomena in which the set of observed variables O is not causally sufficient in a population (Spirtes et al., 1993). CAUDISCO is based on Pearl and Verma's (1991) IC-algorithm, but in addition, it can manage an initial causal model as well as the background knowledge on either the temporal order of variables or the maximum

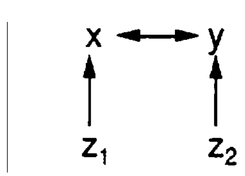


Figure 3. A case of spurious association. A latent variable affects both x and y.

size of contexts. As a matter of fact, the background knowledge is very useful to reduce the computational complexity of the algorithm, since the number of tests to be performed is at worst exponential.

Independence tests performed by CAUDISCO depend on the type of variable. For *continuous* variables, $x_1, x_2, x_j, x_{j+1}, \ldots, x_n$, the Pearson partial correlation coefficient $\rho_{1,2.j,\ldots,n}$ is computed and a test on the hypothesis $\rho_{1,2.j,\ldots,n} = 0$ is performed. Indeed, under the assumption of normality for the distribution of variables $x_1, x_2, x_j, x_{j+1}, \ldots, x_n$, it can be proven that the condition $\rho_{1,2.j,\ldots,n} = 0$ implies $I(x_1, x_2 | x_j, x_{j+1}, \ldots, x_n)$ (Anderson, 1984). The partial correlation coefficient can be estimated by $r_{1,2.j,\ldots,n}$, which is recursively defined as follows:

$$r_{1,2.j,\ldots,n} = \frac{r_{1,2.j+1,\ldots,n} - r_{1,j.j+1,\ldots,n} \times r_{2,j.j+1,\ldots,n}}{\sqrt{1 - r^2_{1,j.j+1,\ldots,n}}\ \sqrt{1 - r^2_{2,j.j+1,\ldots,n}}}$$

The base of the recursion is given by $r_{1,2}$, which is the unbiased estimate of the correlation coefficient $\rho_{1,2}$. Anderson provides a parametric test for the hypothesis $r_{1,2.j,\ldots,n} = 0$. Indeed, Fisher's z statistics,

$$z = \frac{1}{2} \log \frac{1 + r_{1,2.j,\ldots,n}}{1 - r_{1,2.j,\ldots,n}}$$

has an approximate normal distribution for sample size greater than 25.

In contrast to TETRAD II (Spirtes et al., 1993), the independence test for *categorical* variables is based on the χ^2 test (Hogg & Tanis, 1977). More precisely, a contingency table is built for each n-tuple of values of the variables in the context, and a χ^2 test is performed for each table. If the tests are positive for all tables, then a conditional independence is found. This approach has the disadvantage that a large sample is necessary in order to obtain significant results. However, there is no universally accepted distribution-free test for conditional independence of any context size involving categorical variables.

Finally, the same χ^2 test is also performed for integer-valued variables, after their domains have been discretized, according to an entropic criterion (Wong & Chiu, 1987). When the test involves variables of mixed-mode (continuous and categorical/integer), the system discretizes the continuous variables as well, and applies the χ^2 test. In the future, we plan to extend such tests to nonparametric association measures specifically designed for ordinal variables, such as Kendall's τ and Spearman's ρ (Gibbons, 1993).

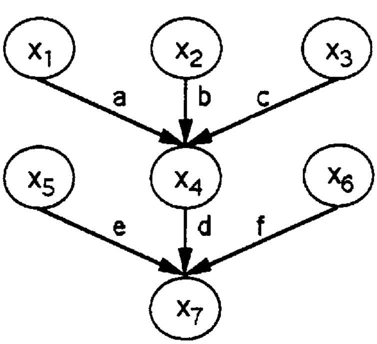
EXPERIMENTAL RESULTS

As observed in the previous section, the discovery of causal knowledge in a database strongly depends on the number of cases available. In order to test the effect of the sample size on the causal structure discovered by the system, we organized an experiment as follows. A database is generated from the causal model in Figure 4, in which all variables are observed, continuous and normally distributed. Furthermore, the type of dependence is linear; thus the joint probability distributions of all variables are still normal. These are ideal conditions that help us to understand which is the minimal sample size required by CAUDISCO in order to recover the whole causal structure. The significance level fixed for each test is $\alpha = 0.05$.

In Table 1, some statistics on the different trials are reported, namely, sample size, maximum size of the context, time needed for the learning phase on a Sparc Station 2, number of edges found, number of exact edges found, number of genuine causal influences detected by the algorithm, number of potential causes found, number of spurious association edges, number of edges left undirected, and number of tests performed. Each trial is identified by a number.

It is interesting to note that with 500 samples (trial 3), the causal structure is almost completely recovered. Indeed, CAUDISCO outputs two spurious associations, namely, $x_3 \leftrightarrow x_4$ and $x_3 \leftrightarrow x_6$, since it fails to detect the independence $I(x_3, x_6|\varnothing)$. However, with 1000 samples, the system is already able to recover the complete structure without mistakes. In trials 6–9, we perturbed the condition of normality by introducing two exponentially distributed variables. For trials 6 and 7, in which the perturbed variables have mean and standard deviation equal to $1/\lambda = 1/4$, we observe a worsening with respect to the other two trials with $\lambda = 2$. Indeed, in trials 8 and 9 the system is still able to recover the complete graph, while in trial 6 it does not detect any dependence between x_6 and x_7, and in trial 7 it fails to find that $I(x_1, x_7|\{x_4\})$.

In the previous experiment, we considered the case of a linear causal model, where all variables are numeric. However, as pointed out in the introduction, most of the real-world applications present mixed-mode data. In the second experiment, we consider the problem of recovering the causal structure underlying mixed-mode synthetic data, generated from the model in Figure 5*a*.



variable	distribution
x1	N(0,1)
x2	N(0,1)
x3	N(0,1)
x4	Σ N(0,1)
x5	N(0,1)
x6	N(0,1)
x7	Σ N(0,1)

edge	value
a	0.5
b	0.5
c	0.5
d	1.0
e	0.5
f	0.5

Figure 4. Causal model for trials 1–9.

Table 1. Experimental results of trials 1–9

	Trial number								
	1	2	3	4	5	6	7	8	9
Sample size	50	100	500	1000	10000	1000	10000	1000	10000
Context size	5	5	5	5	5	5	5	5	5
Search time (s)	1	2	4	5	37	4	37	5	37
Edges found	2	4	7	6	6	5	7	6	6
Exact edges	1	2	4	6	6	5	6	6	6
Genuine causes	0	0	1	1	1	1	1	1	1
Potential causes	1	2	3	5	5	4	5	5	5
Spurious associations	0	1	2	0	0	0	0	0	0
Undirected edges	0	0	0	0	0	0	0	0	0
Tests	90	204	247	234	230	188	244	243	216

The problem concerns the estimation of reliability of an insured driver, given the following six variables:

- Age (A): integer values are uniformly distributed in the range [18, 80].
- Years elapsed since attainment of driver's license (T_a): distributed in the range [0, A–18] as follows:

$$P(T_a) = n + mx \qquad 0 \leq x \leq T_{\max}$$

$$P(T_a) = 0 \qquad \text{otherwise}$$

where $T_{\max} = A - 18$, $n = 1/5T_{\max}$, and $m = 8/5\ T^2_{\max}$.

- Type of (Italian) driver's license (T): four nominal values, B, C, D, E, distributed as follows: $P(B) = 0.57$, $P(C) = 0.285$, $P(D) = P(E) = 0.143$.
- Time elapsed since renewal of the driver's license (T_r): values are uniformly distributed in the range [0, 4] for drivers over 60 or drivers with license E. In all other cases, the license is renewed every 10 years.

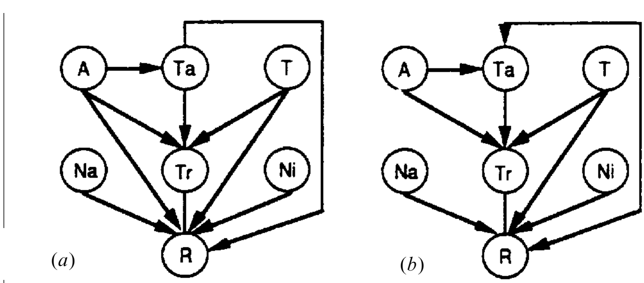


Figure 5. (*a*) Causal structure defined for the insurance domain. (*b*) Inferred causal structure.

- Number of accidents over the last 5 years (N_a): integer values are uniformly distributed in [0, 10].
- Number of infractions of road regulations in the last 5 years (N_i): integer values are uniformly distributed in the range [0, 20].

The reliability of the insured driver (R) can take seven distinct values: very high (0–12), high (13–20), medium-high (21–28), medium-low (29–36), low (36–44), very low (45–52), and unacceptable (>53). Values reported in parentheses are scores associated to each reliability level. Scores are computed as the weighted sum of the previous six variables.

From a sample of ten thousand cases, CAUDISCO is able to recover the causal structure in Figure 5*b*. Time needed depends on the maximum context size considered, and ranges from 27 s for size zero to 117 s for size five. The variables A, T_a, T_r, N_a, and N_i have been discretized in order to perform the independence tests.

Obviously, the dag reported in Figure 5*b* can only explain which variable causes which, but it provides no information on how some variables affect the reliability, or how age and type of license may affect the time elapsed since the last license renewal. As shown by Malerba et al. (1994), it is possible to define multiple learning problems from a causal model. In this case, there are two problems:

1. Find R given A, T_r, N_a, T, and N_i.
2. Find T_r given A, T_a, and T.

We used a decision tree induction system, namely, C4.5 (Quinlan, 1993), in order to generate causal rules for both learning problems. Such rules explain how some causes and effects are related to each other. For the first learning problem, 324 rules were generated, with the number of conditions in the body ranging from two to four. The maximum percentage of covered examples (*support*) for such rules is 3.12%, while the lowest predictive accuracy (*confidence*) is 31.4%. Some of the most interesting rules are reported below:

Support	Confidence	Rule
1.63%	73.7%	$[8 \le N_a \le 9] \wedge [9 \le N_i \le 12] \rightarrow [R = \text{very\_low}]$
0.24%	94.4%	$[4 \le N_a \le 6] \wedge [12 \le N_i \le 14] \wedge [5 \le T_r \le 7] \rightarrow [R = \text{low}]$
3.12%	72.0%	$[4 \le N_a \le 6] \wedge [0 \le N_i \le 2] \rightarrow [R = \text{medium\_high}]$
0.35%	47.0%	$[0 \le N_a \le 1] \wedge [0 \le N_i \le 2] \wedge [T_r = 0] \rightarrow [R = \text{very\_high}]$

where the semantics of the right arrow ($\rightarrow$) is that of causal implication.

From the second learning problem, we obtained 33 rules, with the number of conditions in the body ranging from one to two. This time the maximum support is 10.47%, but the rules are generally less predictive because of the discretization made on the target attribute T_r. An instance of a rule produced by C4.5 for the second problem is the following:

Support	Confidence	Rule
5.97%	43.7%	$[6 \leq T_a \leq 11] \wedge [T = B] \rightarrow [7 \leq T_r \leq 9]$

All these rules can be stored in a rule base, so that the knowledge discovery system can readily answer possible *intelligent queries* made by the final user, such as

> Find all rules with support x and confidence y, such that the type of driving license is causally related to the reliability of the driver

or

> Find all rules with support x and confidence y that explain why the driver has a high reliability.

The experiment on insurance helped to illustrate the methodology as well as its potential applications to the field of knowledge discovery in databases. However, our conclusions were necessarily limited because of the laboratory-sized experiment.

More interesting and meaningful results have been obtained from the data set Pima Indian Diabetes available in the UCI machine learning repository. In this case, 768 female patients were taken into consideration, some of whom showed signs of diabetes according to World Health Organization criteria. The nine variables involved in this study were all numeric; thus we tested conditional independence by means of Pearson's correlation coefficient. We aimed to discover whether and to what extent age, diastolic blood pressure, and other measurable factors are causally related to the presence of diabetes.

By looking at the causal model found by the system (see Figure 6), it becomes evident that two of the nine variables, namely, diastolic blood pressure and diabetes pedigree function, are not considered causally relevant for the classification problem. The plasma glucose concentration after 2 hours in an oral glucose tolerance test is spuriously associated with the class, while age, body mass index, and number of

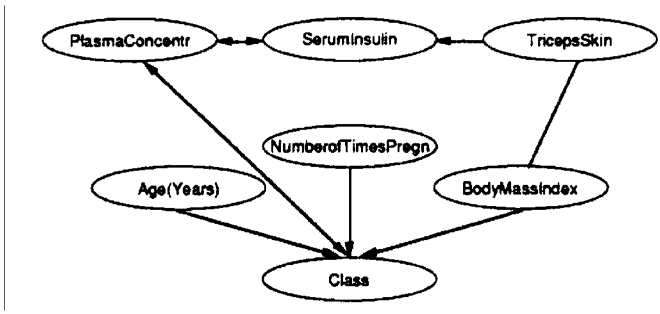


Figure 6. Causal structure found for the UCI data Pima Indian Diabetes.

pregnancies are three potential causes of diabetes. These results are coherent with well-known causal models studied in epidemiology for some forms of diabetes.

Once the causal structure has been recovered, it is possible to induce the rules for predicting the attribute class, given those attributes that are causally relevant, namely, age, plasma concentration, body mass index, and number of pregnancies. The 33 rules produced by C4.5 are quite simple and accurate. In the following, some of them are reported:

Support	Confidence	Rule
12.6%	90.3%	[0 ≤ PlasmaConcentr ≤ 88] → [Class = neg]
0.4%	67.5%	[29 ≤ Age ≤ 34] ∧ [6 ≤ NumberOfTimesPregn ≤ 8] → [Class = pos]

Such rules provide detailed information on the causal relations between these five variables and can be useful in abductive reasoning in intelligent systems. It is interesting to note that, by running C4.5 alone on the whole data set, we obtained 30 rules, 5 of which involved the attributes discarded by the causal analysis, namely, diastolic blood pressure and diabetes pedigree function. Moreover, the following rule was generated:

$$[30.5 \leq \text{SerumInsulin} \leq 81.5] \rightarrow [\text{Class} = \text{neg}]$$

which is equally odd, since statistical tests failed to detect any direct dependence between SerumInsulin and diabetes.

CONCLUSIONS

In this article a system for the discovery of causal rules in relational databases has been presented. The discovery process is two-phased: initially, a causal structure is inferred from the data, then a set of inductive learning problems is generated in order to learn the causal rules.

The first phase is based on the study of conditional independencies observed in the data. Conditional independencies define a set of constraints that must be satisfied by the induced causal structure. Such constraints can easily be established for both categorical and continuous variables, since they simply require a χ^2 test or a normal test for Fisher's z statistics, respectively. Currently, we are also investigating the application of some distribution free independence tests for ordinal variables, as well.

In the second phase we used the system C4.5 to generate a set of rules for each relevant causal dependence. This approach opens new perspectives in the area of inductive learning, since the rules we induce have a causal semantics, while rules generally produced by machine learning systems are based on the simpler semantics of logical entailment. When a causal explanation of a phenomenon is needed, it is possible to use causal inference systems as preprocessors, which select causally relevant variables for a given target attribute. This view can be extended to the problem of learning multiple dependent concepts, in which case the causal graph can be used to define the order in which target concepts must be learned (Malerba et al., 1996).

In the future, CAUDISCO will be embedded in a prototypical system for intelligent queries, which concern possible dependencies among attributes of one or more relations of a database. In contrast to ordinary queries whose answers are a set of tuples, the result returned to an intelligent query is a rule induced from data. Therefore intelligent queries activate inductive learning processes working on data generated by different views of the database. Such views are automatically generated in order to take into account different ways to explain dependencies among attributes in the query.

Finally, experiments described in this article present two limitations that we plan to overcome in the future. First, the learning system we used is appropriate only for classification problems, while in several cases, we have regression problems. For instance, the second learning problem generated in the insurance domain required the induction of a regression tree for T_r, as the time elapsed since the renewal of the driver's license is a numerical variable. We partially solved the problem by considering the same discretization of T_r produced by CAUDISCO, but we plan to use a regression tree induction system in future studies. Furthermore, when variables are spuriously associated, the direction of causal dependencies is not uniquely determined; thus any choice we make while defining a classification problem is arbitrary. Actually, in this case it would be plausible to apply unsupervised learning techniques to the discovery of latent classes for spuriously associated variables.

REFERENCES

Anderson, T. W. 1984. *An introduction to multivariate statistical analysis.* New York: John Wiley.

Asher, H. B. 1983. *Causal modelling.* Newbury Park, Calif.: Sage Publications.

Cercone, N., and M. Tsuchiya (eds.). 1993. Special issue on "Learning and discovery in knowledge-based databases." *IEEE Transactions on Knowledge and Data Engineering* 5(6).

Esposito, F., D. Malerba, and G. Semeraro. 1994. Discovering probabilistic causal relationships: A comparison between two methods. In P. Cheeseman and R. W. Oldford (eds.), *Selecting models from data.* Lecture Notes in Statistics 89, 233–242. Berlin, Germany: Springer-Verlag.

Geiger, D., and J. Pearl. 1989. Logical and algorithmic properties of conditional independence and qualitative independence. Technical Report CSD 870056, R-97-IL. Cognitive Systems Laboratory, University of California, Los Angeles.

Gibbons, J. D. 1993. *Nonparametric measures of association.* Newbury Park, Calif.: Sage Publications.

Glymour, C., R. Scheines, P. Spirtes, and K. Kelly. 1987. *Discovering causal structure.* Orlando, Fla.: Academic.

Hogg, R. V., and E. A. Tanis. 1977. *Probability and statistical inference.* New York: Macmillan.

Malerba, D., G. Semeraro, and F. Esposito. 1994. An analytic and empirical comparison of two methods for discovering probabilistic causal relationships. In F. Bergadano and L. De Raedt (eds.), *Machine learning: ECML-94.* Lecture Notes in Artificial Intelligence 784, pp. 198–216. Berlin, Germany: Springer-Verlag.

Malerba, D., G. Semeraro, and F. Esposito. 1996. A multistrategy approach to learning multiple dependent concepts. In C. Taylor and R. Nakhaeizadeh (eds.), *Statistics and machine learning: The interface.* 87–106.London, England: Wiley.

Matheus, J. C., P. K. Chan, and G. Piatetsky-Shapiro. 1993. Systems for knowledge discovery in databases. *IEEE Transactions on Knowledge and Data Engineering* 5(6):903–913.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, Calif.: Morgan Kaufmann.

Pearl, J., and T. S. Verma. 1991. A theory of inferred causation. In J. A. Allen, R. Fikes, and E. Sandewall (eds.), *Principles of knowledge representation and reasoning: Proceedings of the second international conference,* pp. 441–452. San Mateo, Calif.: Morgan Kaufmann.

Piatetsky-Shapiro, G., and W. J. Frawley (eds.). 1991. *Knowledge discovery in databases.* Cambridge, Mass.: AAAI/MIT.

Quinlan, J. R. 1993. *C4.5: Programs for machine learning.* San Mateo, Calif.: Morgan Kaufmann.

Spirtes, P., C. Glymour, and R. Scheines. 1990. Causality from probability. In *Proceedings of the international conference on advanced computing for the social sciences, Williamsburg, Va.*

Spirtes, P., C. Glymour, and R. Scheines. 1993. *Causation, prediction and search.* Lecture Notes in Statistics 81. Berlin, Germany: Springer-Verlag.

Verma, T. S., and J. Pearl. 1990. Equivalence and synthesis of causal models. In J. A. Allen, R. Fikes, E. Sandewall (eds.), *Proceedings of the sixth international conference on uncertainty in artificial intelligence,* pp. 220–227. Mountain View, Calif.: Association for Uncertainty in AI, Inc.

Wong, A. K. C., and D. K. Y. Chiu. 1987. Synthesizing statistical knowledge from incomplete mixed-mode data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9(6):796–805.

Zytkow, J. M., and J. Baker. 1991. Interactive mining of regularities in databases. In G. Piatetsky-Shapiro and W. J. Frawley (eds.), *Knowledge discovery in databases,* pp. 31–53. Cambridge, Mass.: AAAI/MIT.

PAGE 84 BLANK