Discovering Attribute Dependence in Databases by Integrating Symbolic Learning and Statistical Analysis Techniques

I. F. Imam, R. S. Michalski and L. Kerschberg

Center for Artificial Intelligence
George Mason University
Fairfax, VA, 22030

ABSTRACT

The paper presents a method for integrating two different types of data analysis: symbolic inductive learning and statistical methods. The method concerns the problem of discovering rules characterizing the dependence between a group of dependent attributes and a group of independent groups, e.g., decision attributes and measurable attributes. The method proceeds along two paths. One paths generates a set of rules characterizing this dependency using a symbolic inductive learning program (AQ15). In the obtained rules, each attribute is assigned an "importance score" which represents the ratio of the total number of examples that can be classified by this attribute to the maximum number of classified examples. Second path calculates the correlation coefficient between the decision and measurable attributes using the Chi-square test. The independent attributes with low "importance score" and low correlation are considered irrelevant, and removed from the data. The AQ15 is applied again to the data set spanned over the reduced set of attributes, in order to determine the simplest expression characterizing the dependency. The method has been experimentally applied to two real world problems: "Wind bracing" —to determine the dependence of the type of wind-bracing for tall buildings on the parameters of the building, and "Accident factors"—to determine the dependence between the age of construction workers and the types of accidents. The results have shown that the proposed approach of combining two different methods for relevant attribute determination gives better results than those obtained by applying either method separately.

Key words: Knowledge discovery, Databases, Machine learning.

1. Introduction

Methods for knowledge discovery in databases can be generally classified into two groups: statistical and conceptual. Statistical methods are analyze numerical data and creating statistical characterizations (Chan & Wong, 1991; Klosgen, 1992; Major & Riedinger, 1992). Conceptual methods (also called symbolic or logic-based) are primarily oriented toward analyzing (mostly) symbolic data and creating general "conceptual" descriptions in the form of decision rules, decision trees, or other symbolic representations (e.g., Kaufman, Michalski & Kerschberg, 1991; Gonzalez, et al, 1991; Bergadano, et al, 1991; Piatetsky-Shapiro, 1991; Manago & Kodratoff, 1991; Piatetsky-Shapiro & Matheus, 1992; Michalski, et al, 1992).

Most statistical methods involve computing statistical correlation among the attributes. Conceptual methods typically use a machine learning technique. They typically involve obtaining the user's goal (requesting the decision attributes), selecting a relevant subset of data, and then inducing rules or patterns characterizing the data.

This paper addresses the problem of integrating statistical and conceptual methods. It presents a new strategy for determining irrelevant attributes (for any given user's goal) in a large database. The method uses an inductive learning technique AQ15 (Michalski et. al, 1986) for generating conceptual descriptions and calculating the "importance score" for each input attribute. It also uses the Chi-square test to compute correlation among the input and output attributes.

The importance score for an attribute is the ratio of the total number of data tuples (examples linked to the given output attribute) that are covered by induced decision rules that contain this attribute to the maximum number of matched examples by any attribute. Attributes with an importance score less than a given threshold and with low Chi-square score are assumed to be irrelevant and are removed from the database. The modified database is used as an input to AQ15 program. The program determines decision rules characterizing the dependence between the reduced set of input attributes and the output attribute.

2. A Brief Description of the AQ15 Rule Learning Program

In order to make the paper self-contained, we briefly describe the system, AQ15 (Michalski et al., 1986) that learns decision rules for a given set of decision classes from examples. The method uses the AQ algorithm, which is an implementation of the STAR methodology (Michalski, 1983). The algorithm starts with a "seed" example of a given decision class, and generates a set of the most general conjunctive descriptions of the seed (decision rules). Such a set is called a star of this seed. Subsequently, the algorithm selects from the star a description that optimizes a criterion reflecting the needs of the problem domain. If the criterion is not defined, the default criterion is to select the description covers the largest number of positive examples and, with the second priority, involves the smallest number of attributes. If this description does not cover all examples of the class, a new seed is selected, from uncovered examples and the process continues until a complete class description is generated. The algorithm can work with just a few examples or with very many examples.

The learned descriptions are represented in the form of a set of decision rules (a "ruleset") expressed in an attributional logic calculus, called *variable-valued logic (VL1)* (Michalski, 1973). A distinctive feature of this representation is that it employs, in addition to standard logic operators, the internal disjunction operator (a disjunction of values of the same attribute), which can significantly simplify rules involving multivalued attributes.

AQ15 can generate decision rules that represent either *characteristic* or *discriminant* concept descriptions, or an intermediate form, depending on the settings of its parameters. A characteristic description states properties that are true for all objects in the concept. The simplest characteristic concept description is in the form of a single conjunctive rule. The most desirable is a rule with the *longest* condition part (stating as many common properties of objects of the given decision class as possible).

A discriminant description states only the properties that discriminate a given concept from a fixed set of other concepts. The most desirable is a rule with the *shortest* condition part

(Michalski, 1983). For example, to distinguish a given set of tables from a set of chairs, one may only need to indicate that tables "have large flat top." A characteristic description would include also properties such as "have four legs, have no back, have four corners, etc." Discriminant descriptions are typically much shorter than characteristic descriptions.

3. Discovering Attribute Dependence

An application of our methodology is done on two distinct databases, the wind bracing (Arciszewski, et al, 1992) and the accident data (Arciszewski, et al, 1992). Four different strategies were done on each set of data to show the effect of each strategy on the performance of the discovery. All the strategies are based on the inductive learning system AQ15 (Michalski, et al., 1986). The wind bracing data contains a decision attribute representing the structural worth of the buildings with four values (i.e. four decision classes) high (c1), medium (c2), low (c3) and infeasible (c4). The data has also seven attributes: number of stories (x1), bay length (x2), wind intensity (x3), number of joints (x4), number of bays (x5), number of vertical trusses (x6), and number of horizontal trusses (x7). The database consists of 336 examples. Only 220 random examples were used for learning a concept description and 116 for testing. The goal of the discovery is to learn knowledge about the important features which may affect the structural worth of any building.

The accident data divides examples of construction accidents according to the age of the workers. The attribute "age" has four possible values; young (c1), medium (c2), old (c3), and unknown (c4). There are 12 other attributes in the database: Race (x1), Marital Status (x2), Children (x3), Occupation (x4), Job Experience (x5), Hour of Day (x6), Season (x7), Accident type (x8), Work Period (x9), Injury Type (x10), Injured Part (x11), Return Work (x12). The accident database consists of 225 examples. A random sample of size 169 (75%) is chosen as a training set and 56 examples for testing. The goal of the discovery is to learn knowledge about the relation between the age of the workers and the character of accidents.

Figure 1 shows a plan of performing the four strategies. The first strategy uses AQ15 to learn rules directly from the database. The second strategy uses the chi-square test to determine the set of attributes relevant to the decision attribute(s) (the user interest). A subset of the database, which contains only the relevant attributes, is used by AQ15 to learn a new concept description in the form of rules. Third strategy uses the output of the first one together with the database to calculate the importance score for each measurable attribute. Any attribute has importance score greater than the specified threshold is considered as relevant attribute. A subset of the database which contains only the relevant attributes is used by AQ15 for learning description rules. In strategy four, each of the second (chi-square test) and the third (importance score) experiment provides a set of relevant attributes. A subset of the database which contains the union of both relevant attributes is used as an input to AQ15 for discovery.

To test the effectiveness of our methodologies, we use training examples in the discovery process, and testing examples in measuring the correctness of the discovered knowledge. These strategies used AQ15 to generate knowledge, in the form of discriminant rules, from examples (each example is equivalent to one tuple of the database). Discriminant rules are used to calculate the importance score because discriminant rules differentiate classes from each other, and contain only attributes relevant to this goal, while characteristic rules include all attributes, relevant or irrelevant in the generated rules. Thus characteristic rules would produce equal attribute importance score.

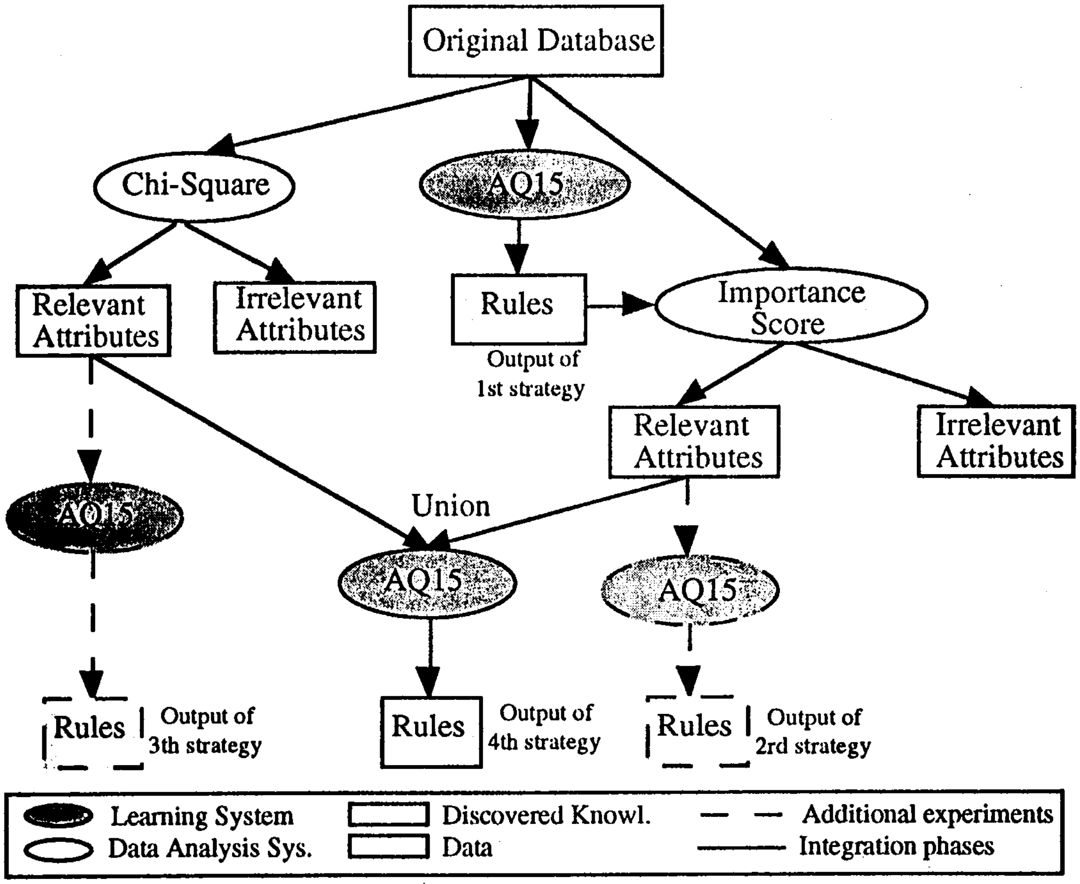

Figure 1: A plan of performing the different strategies.

3.1. Using AQ15 for discovery

In the first strategy, we ran AQ15 against the wind bracing data and the accident data, separately. First, The databases are translated into a symbolic format (examples) to be used by AQ15. We set the AQ15's parameters to generate a discriminant and disjoint rules. The effect of intersected rules on this methodology would be interesting work for the future. Also, a parameter was set to print with each rule the number of examples that it covers. After running AQ15, the produced rules are tested against the testing examples. The accuracy of testing the discovered rules against the testing data for the wind bracing data is 81.74%. The accuracy of testing the discovered rules against the testing data for the accident data is 79.11%.

3.2. Determining the relevant attributes to the decision attribute using Chi-square test

The second strategy searches first for the attributes relevant to the decision attribute, then it applies AQ15 to these relevant attributes. The strategy uses the Chi-square test to determine the correlation between the decision attribute and all of the other attributes. The strategy was done as follows: 1) create statistical tables for each attribute in the data against the decision attribute; 2) detect the irrelevant attributes to the decision attribute by calculating the Chi-square value using formula 1; 3) modify the database by removing the irrelevant attributes from the database; 4) run AQ15 on the data set and determine the testing accuracy.

An attribute x is strongly correlated to the decision attribute if the calculated Chi-square value is greater than the measured one for a significant percentage. We choose 97.5% and 75% significance for the wind bracing and accident data, respectively. In the case of the accident data, we reduced the significant interval to include more attributes (in the case of 90% significant interval, there was just two attributes and the accuracy was zero).

Figure 2 presents a relational table for the attribute x7 of the wind bracing data and the decision classes. Each number in the table aij represents the number of examples that belong to class ci and have x7 take value vj. For example, the number 12 in the first row and first column represents the number of examples that belong to class c1 and the value of x7 has value one.

Concept Values	Legal values of attribute x7				Total
	v1	v2	v3	v4	
c1	12	9	8	6	**35**
c2	20	21	22	27	**90**
c3	28	15	24	23	**90**
c4	1	1	2	1	**5**
Total	**61**	**46**	**56**	**57**	**220**

Figure 2: Distribution of the values of attribute (x7) of the wind bracing data in the training examples over each of the decision classes.

Formula 1 shows how the Chi-square significance can be calculated from Figure 2. The decision attribute has four values, c1, c2, c3 & c4. The attribute x7 has four values, v1, v2, v3 & v4

$$\textbf{Chi-square}\ (x_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} [\ (a_{ij} - c_{ij})^2 / c_{ij}\ ] \qquad (1)$$

where n is the number of decision classes, m is the number of values of xj, a_{ij} (is the number in row i and column j in the table), is the number of examples in class ci where the attribute x takes value vj, $c_{ij} = (T_{ci} * T_{vj}) / T$, T_{ci} and T_{vj} are the total values over the decision class ci and the total values over the value vj of the given attribute, respectively. T is the total number of examples. The calculated value of Chi-square for attribute x7 is 5.6.

The statistical degree of freedom for determining the Chi-square value is calculated as follows:

(Number of decision classes - 1) * (Number of legal values of the attribute - 1) (2)

Attribute x7 has a degree of freedom 9. For a confidence interval equal to 97.5%, the tabled Chi-square value is 19.02. As Chi-square(x7)=5.6 < 19.02, The attribute x7 is irrelevant. From the wind-bracing data, the Chi-square test shows that the set of relevant attributes Rw= {x1, x4, x5, x6}, and the set of irrelevant attributes Iw= { x2, x3, x7}. We removed the irrelevant attributes from the original database and applied the AQ15 program to the rest of the data. The discovered rules were simpler, but not as accurate as the rules learned from the entire original attribute set only using AQ15. The accuracy of testing the knowledge (learned over the reduced data set) against the testing data was 74.78% for the wind bracing data and 32.89% for the accident data.

3.3. Determining irrelevant attributes based on the analysis of decision rules

The third strategy is used to determine the irrelevant attributes by calculating the importance score for each attribute in the decision rules generated by the AQ15 program. In this strategy, we assumed that irrelevant attributes are those which occur in rules that match less than a specified percentage (threshold) of the maximum number of matched examples. In this research, this threshold is set arbitrarily (50% for the wind bracing data, and 40% for the accident data).

This strategy addresses also the issue of whether or not these irrelevant attributes increase the complexity of the discovery process, and/or reduce the predictive accuracy (or the degree of correctness) of the discovered knowledge.

Figure 3 shows three rules produced in the first strategy. Each attribute present in any rule is considered as an important to the rule and has a number of examples that are covered by this rule.

1 [x1 = 1][x4 = 1 v 3][x7 = 1..3]
{Total: 14; Examples covered: 2, 4, 11..14, 18, 19, 23, 26, 27, 29..31}
2 [x1&x4 = 1 v 3][x5 = 1][x7 = 1 v 2 v 4]
{Total: 11; Examples covered: 1, 3, 6, 9, 15, 17, 21, 28, 32..34}
3 [x1 = 1 v 2][x4 = 3][x5 = 2][x7 = 4]
{Total: 3; Examples covered: 10, 22, 24}

Figure 3: A sample of AQ15 rules, with list of the covered examples.

To print the total number of examples that match each rule, set the parameter "echo", in the input file to AQ15, to the appropriate mode which print after each rule the examples that are covered by this rule.

In this example, rule 1 covers 14 different examples, which means the attributes x1, x4 and x7 have a score of 14, each, from this rule. For these attributes, the score will increase by 11 from the second rule, while the attribute x5 takes an initial score 11.

The importance score of an attribute is the ratio of the total number of examples that are covered by any rule, (for all the decision classes) that including that attribute, to the maximum number of covered examples. If this ratio is less than the specified threshold, we consider that attribute irrelevant to the discovery process.

For example, suppose we have n classes c1, ..., cn, and m attributes A1, ..., Am. Assume also for each attribute Aj there are Ec_{ij} examples matching rules which contain Aj, and belong to class ci (i=1,..,n; j=1,.., m). The importance score for Aj is given by:

$$I\_S(Aj)= \left(\sum_{i=1}^{n} Ec_{ij} \right) / \left(\max_{j} \sum_{i=1}^{n} Ec_{ij} \right) \qquad (3)$$

We chosen a threshold value of 0.5 for the wind bracing data, and 0.4 for the accident data. Determining the optimum threshold is very important point for further research. In this strategy, the accident data with threshold 0.5 did not perform well.

Figure 4 shows for each attribute (of the wind bracing data), the number of examples that are covered by rules which include that attribute. Figure 5 shows the importance score for each attribute.

Figure 6 shows for each attribute (of the accident data), the number of examples that are covered by rules which include that attribute. Figure 7 shows the importance score for each attribute.

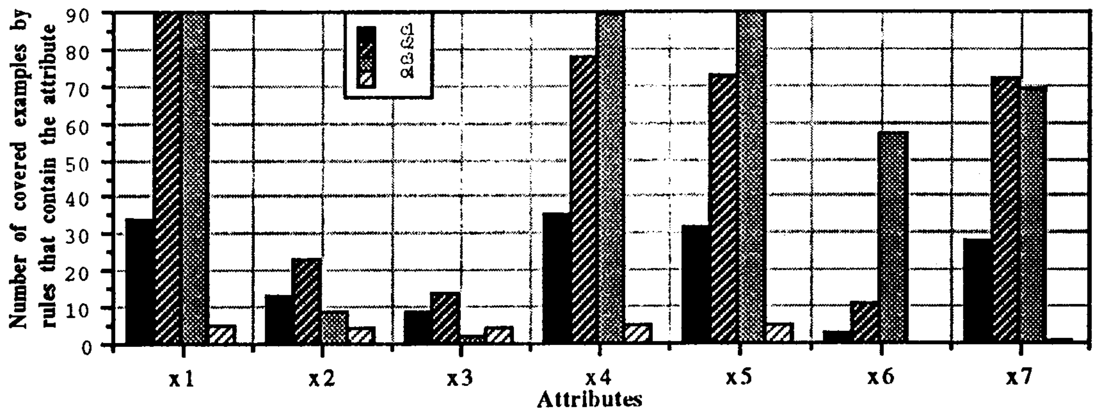

Figure 4: The frequency of the wind bracing attribute in AQ15 rules.

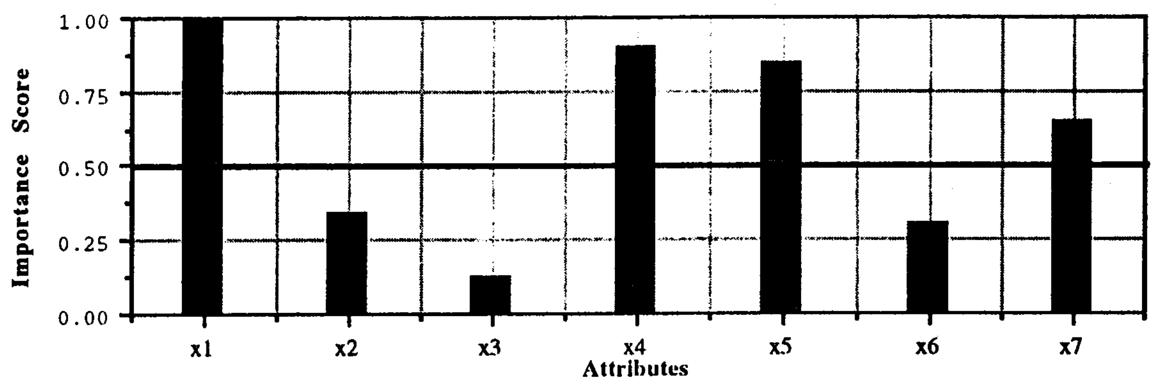

Figure 5: The importance score for the wind bracing attributes.

As we can see in Figure 5, the set of relevant attributes in the wind bracing database is Rw= {x1, x4, x5, x7}. From Figure 7, the set of relevant attributes in the accident database is Ra={x2, x4, x5, x6, x7, x8, x11}. Consequently, the sets of irrelevant attributes are Iw= {x2, x3, x6} and Ia= {x3, x9, x10, x12}. We apply the AQ15 program to discover knowledge from the data that contains only the relevant attribute for both problems. The produced rules from both problems (wind bracing and accident) were simpler and accurate than the discovered rules in the first two strategies. The accuracy of the discovered knowledge was 83.48% and 80.44, for the wind bracing and accident problems, respectively. Full results are shown in figure 8.

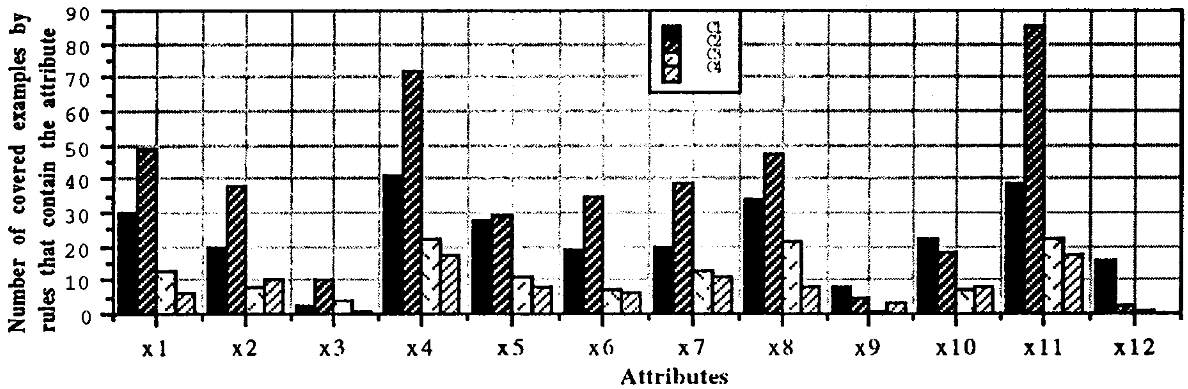

Figure 6: The frequency of the accident attributes in AQ15 rules.

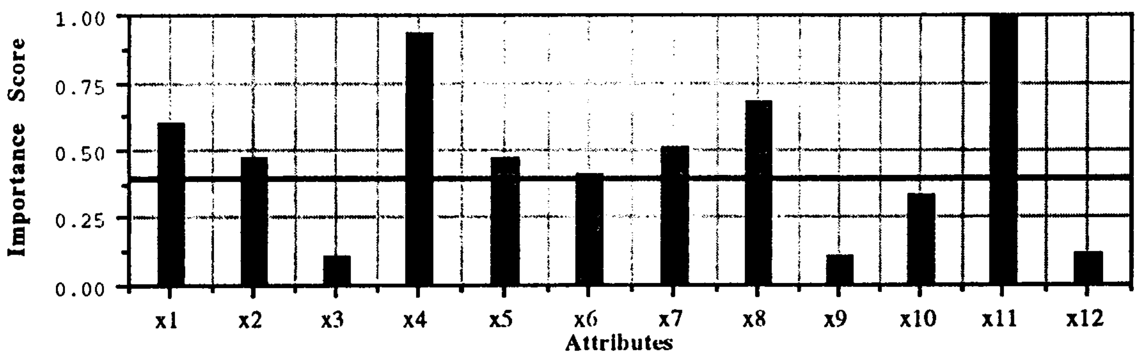

Figure 7: The importance score for the accident attributes.

The accident data answers a very interesting question which is "can we get this results just by adjusting the importance score?". The answer is no!. For example, (as we will see in the next section) the integration of the two methods adds x3 from the chi-square relevant attributes to the importance score set of relevant attributes which increases the accuracy. If we want to add x3 using the importance score, we have to add all the other attributes (because x3 has the minimum importance score).

3.4. Discovering knowledge by integrating the Chi-square and the importance score techniques

This section introduces the integration of the Chi-square and the importance score together with AQ15 for the discovery process. The second strategy, Chi-square, searches for relevant attributes to the decision attribute or the user interest, but the third strategy, importance score, searches for relevant attributes to the discovery algorithm. The integration strategy searches for relevant attributes to both the decision attribute and the discovery algorithm. As we can see from the previous strategies, It is inefficient either to remove an attribute that somehow is important to the discovery algorithm nor to keep an attribute that is totally irrelevant to the decision attribute (the user interest).

The proposed method locates the irrelevant set of attributes (to the decision attribute) to both using the Chi-square test, and it determines the set of attributes that are irrelevant to the discovery Algorithm AQ15 using the importance score. The intersection of these two sets gives the optimal set of irrelevant attributes. On the other hand, the intersection of the relevant sets from the chi-square test and the importance score gives the optimal set of relevant attributes (to the user interest), with the specified threshold and confidence interval.

After performing the wind-bracing experiment, the set of irrelevant attributes to the decision attribute (from chi-square) was $Iw_c = \{x2, x3, x7\}$, and the set of irrelevant attributes to the discovery algorithm (from importance score) was $Iw_{is} = \{x2, x3, x6\}$. The optimal set of irrelevant attributes is given by the intersection of these sets $O_i = Iw_c \cap Iw_{is} = \{x2, x3\}$. The union of the relevant attributes from both methods is $U_r = Rw_c \cup Rw_{is} = \{x1, x4, x5, x6, x7\}$. The optimal set of relevant attribute is $O_r = Rw_c \cap Rw_{is} = \{x1, x4, x5\}$. Applying AQ15 on the database with all the attributes in U_r gave us the best accuracy in the four strategies 86.09%.

Applying this strategy on the accident data, the set of irrelevant attributes to the decision attribute (from chi-square) was $Ia_c = \{x4, x6, x7, x8, x9, x10, x11, x12\}$, and the set of irrelevant attributes to the discovery algorithm (from importance score) was $Ia_{is} = \{x3, x9, x10, x12\}$. The optimal set of irrelevant attributes is given by the intersection of these sets $O_i = Ia_c \cap Ia_{is} = \{x9, x10, x12\}$. The union of the relevant attributes from both methods is $U_r = Ra_c \cup Ra_{is} = \{x1, x2, x3, x4, x5, x6, x7, x8, x11\}$ (where $Ra_{is} = \{x1, x2, x4, x5, x6, x7, x8, x11\}$). The optimal set of relevant attribute is $O_r = Ra_c \cap Ra_{is} = \{x1, x2, x5\}$. Applying AQ15 on the database with all the attributes in U_r gave us the best accuracy 82.67%.

4. Discussion of Results

The goal of integrating statistical and symbolic techniques is to simplify the data by focusing on the relevant attributes, and in the same time to improve the discovered knowledge. The accident data is very noisy. Using chi-square for determining the relevant attributes for such noisy data give poor results. The Chi-square test with 95% confidence interval determined all but x1 and x2 where relevant. Most likely, the discovery or classification results produced by two attributes will fail. From these experiments, one can notice that, determining the relevant attributes to other attributes (statistically) is not as important as determining the relevant attributes through the discovery algorithm.

Figure 8 shows a summary of results from applying various method on the wind-bracing and the accident databases.

Discovered knowledge from the wind bracing data

Measuring the structural worth of buildings does not depend on the bay length nor the wind intensity. It depends highly only on the number of floors, the number of joints, and the number of bays. The expert supported these results.

Discovered knowledge from the accident data

The relation between the age of workers and the accident features does not depend on whether they can return to their work or not (x12), the type of injury (x10), nor the work period (x9). The worker age depends on the race (x1), depends on the job experience (x2), and depends on the marital status (x5). These relevant attributes and the attribute "age" were classified by the expert as personal information.

Experiment	Database			
	Accident		Wind Bracing	
	# of attributes	Accuracy	# of attributes	Accuracy
AQ15 Only	12	79.11%	7	81.74%
AQ15 & Chi-square	4	32.89%	4	74.78%
AQ15 & Importance Score	8	80.44%	4	83.48%
AQ15, Chi-square & Importance Score	9	82.67	5	86.09%

Figure 8: A summary of results on the comparing different inductive programs in terms of the number of attributes and the prediction accuracy.

5. Summary

This paper presented a new method for determining irrelevant attributes in databases, and thus improving the process of knowledge discovery. The methodology combines a symbolic inductive learner (AQ15), with two statistical approaches (Chi-square and importance score). It uses the Chi-square test to determine the correlation coefficient between the decision attribute (which is defined by the user), and the measurable attributes. It also determines irrelevant attributes by calculating the "importance score" of each attribute and testing it against a defined threshold. This score is calculated from the output rules of AQ15 over the entire database. The methodology removes from the database any attribute that has an importance score less than the threshold, and an insignificant correlation (from Chi-square). The experiments demonstrated that combining the three methods leads to the improvement of the final rules characterizing patterns in data. The rules were simpler and more accurate.

Future research should concern the development of a general framework for solving the problem of dependence among the attributes (specially, with very large databases). Another interesting problem would be to study the proposed method in the context of using other learning techniques, such as decision trees, neural networks and genetic algorithms.

ACKNOWLEDGMENTS

The authors thank Eric Bloedorn, Mike Hieb, and Janusz Wnek for their careful review and useful comments.

This research was conducted in the Center for Artificial Intelligence at George Mason University. The Center's research is supported in part by the National Science Foundation under grant No. IRI-9020266, in part by the Defense Advanced Research Projects Agency under the grant No. N00014-91-J-1854, administered by the Office of Naval Research, and the grant No. F49620-92-J-0549, administered by the Air Force Office of Scientific Research, and in part by the Office of Naval Research under grant No. N00014-91-J-1351.

REFERENCES

Arciszewski, T, Bloedorn, E., Michalski, R., Mustafa, M., and Wnek, J., "Constructive Induction in Structural Design", Report of Machine Learning and Inference Labratory, MLI-92-7, Center for AI, George Mason Un., 1992.

Chan, K. C., and Wong, A. K., "A Statistical Technique for Extracting Classificatory Knowledge from Databases", *Knowledge Discovery In Databases*, Piatetsky-Shapiro, G., Frawley, W., (Eds.), AAAI Press, 1991.

Frawley W., Piatetsky-Shapiro G., and Matheus C., "Knowledge Discovery in Databases: An Overview", *AI Magazine, pp. 57-70,* Fall, 1992.

Gonzalez, A.J., Myler, H.R., Towhidnejad, M., McKenzie, F.D., Kladke, R.R., and Laureano, R., "Automated Extraction of Knowledge from Computer-Aided Design Databases", *Knowledge Discovery In Databases*, Piatetsky-Shapiro, G., Frawley, W., (Eds.), AAAI Press, 1991.

Manago, M. and Kodratoff, Y., "Induction of Decision Trees from Complex Structured Data", *Knowledge Discovery In Databases,* Piatetsky-Shapiro, G., Frawley, W., (Eds.), AAAI Press, 1991.

Kaufman, K.A., Michalski, R.S., and Kerschberg, L., "Mining for Knowledge in Databases: Goals and General Description of the INLEN System", *Knowledge Discovery In Databases,* Piatetsky-Shapiro, G., Frawley, W., (Eds.), AAAI Press, 1991.

Klosgen, W., "Problems for Knowledge Discovery in Databases and Their Treatment in the Statistics Interpreter Explora", *International Journal of Intelligent Systems*, Yager, R.R. (Ed.), pp. 649-673, Vol. 7, No. 7, Sep., 1992.

Major, J.A., and Riedinger, D.R., "EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud", *International Journal of Intelligent Systems*, Yager, R.R. (Ed.), pp. 591-607, Vol. 7, No. 7, Sep., 1992.

Michalski, R.S., "AQVAL/1-Computer Implementation of a Variable-Valued Logic System VL1 and Examples of its Application to Pattern Recognition" in *Proceeding of the First International Joint Conference on Pattern Recognition*, Washington, DC, pp. 3-17, October 30-November, 1973.

Michalski, R.S., "A Theory and Methodology of Inductive Learning", *Artificial Intelligence*, Vol. 20, pp. 111-116, 1983.

Michalski, R.S., Mozetic, I., Hong, J., and Lavrac, N., "The Multi-Purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains," *Proceedings of AAAI-86*, pp. 1041-1045, Philadelphia, PA: , 1986.

Michalski, R.S., Kerschberg, L., Kaufman, K.A., and Ribeiro, J.S., "Mining for Knowledge in Databases: The INLEN Architecture, Initial Implementation and First Results, "*Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies*, Vol. 1, No. 1, August 1992.

Piatetsky-Shapiro, G., "Discovery, Analysis, and Presentation of Strong Rules", *Knowledge Discovery In Databases*, Piatetsky-Shapiro, G., Frawley, W., (Eds.), AAAI Press, 1991.

Piatetsky-Shapiro, G., and Matheus, C.J., " Knowledge Discovery Workbench for Exploring Business Databases", *International Journal of Intelligent Systems*, Yager, R.R. (Ed.), pp. 675-685, Vol. 7, No. 7, Sep., 1992.