
SPE 68818

# A Rule Induction Algorithm for Application to Petrophysical, Seismic, Geological and Reservoir Data

Clayton V. Deutsch[1], YuLong Xie*[1], and A. Stan Cullick[2]
[1] University of Alberta, Edmonton, Alberta
[2] Landmark Graphics, Austin, Texas



This paper was prepared for presentation at the SPE Western Regional Meeting held in Bakersfield, California, 26–30 March 2001.



## Abstract

This paper introduces an algorithm for rule induction intended to provide new insights, improve the reliability and expedite the utilization of large petrophysical and geologic databases. Very large petrophysical, geophysical, and geological databases contain multiple data types, which must be interpreted for application in subsurface modeling. This paper presents a significant advance in discovering complex and even nontrivial data relationships from such databases.

Geoscientists are often challenged to predict subsurface lithologies and properties from multivariate relationships within large databases of core, wireline, and seismic data. Many data analysis techniques are used including histograms, parametric and non-parametric regression, *n*-dimensional histograms, cluster analysis, discrimininant analysis, principal components analysis. This paper introduces a new algorithm that seeks to discover "rule-like" relationships within the data that can be used to make predictions. The method is loosely derived from a data mining technology of classification.

Concepts of data attribute distinguishability and importance are introduced to assess the value of the data and the outcomes to predictability. The new theory, implementation details, and an application are presented. Current petrophysical, seismic, and geostatistical analysis benefit from the rule induction algorithm presented. Improved reservoir characterization and forecasting result.

## Background

The field of data mining[2,7,8,11,17,18] has grown in recent years to deal with large databases available in different industries, in particular, the financial and medical fields. Data mining is the identification or discovery of patterns in data. There are several different types of data mining. These include classification, clustering (segmentation), association, and sequence discovery. The main focus of classification is supervised induction, that is, inference of rules and relationships from large databases. The aim is to extract knowledge from data, so that results not directly in the training data set can be predicted. The training data helps to distinguish predefined classes. Neural networks[3,9,13,14,27], decision trees[5,6,16,19,26,29] and "if-then-else" rules are classification techniques. A disadvantage of neural networks is that it is difficult to provide a good rationale for the predictions made, that is, the rules are not always clear.

Data mining is an interdisciplinary field bringing together techniques from statistics, machine learning, artificial intelligence, pattern recognition, database, and visualization technologies. The methods used in data mining are not fundamentally different from older quantitative model-building techniques, but are natural extensions and generalizations of such methods. There are many applications of various data mining techniques to petroleum characterization[1,4,12,15,28].

A rule-based algorithm is intended to provide understandable rule-like relationships in the data. A rule is a prevailing quality or state. Induction is an instance of reasoning from a part to a whole. Rules indicate the degree of association between variables, map data into predefined classes, and identify a finite set of categories or clusters to describe the data. The rules support specific tasks and are generated by repeated application of a certain technique, or more generally an algorithm, on the data. Rough Sets[17,20-25,30,31] are specialized methods for inducing rules. The essential idea of rough sets is to express uncertain knowledge through an approximation space, which is constructed as certain sets.

Many other methods, including regression analysis, assume that there is a functional form between the predictor and response variables. These smooth out variations and are difficult to apply to multivariate nonlinear responses. Discriminant analysis separates samples into groups based on relationships in the training data. The relations must be linear combinations of variables that are made explicit. *N*-

* Currently in Pacific Northwest National Laboratory, Richland, WA, USA and On leave from Xiangtan University, Xiangtan, Hunan, P. R. China

dimensional histograms are used to delineate a relationship between a response and multiple predictors, which preserves the uncertainty in the relation by reading the value of the response directly from the set of predictors. Principal components analysis is a popular technique to discover the source of variation within the data but again results are expressed as linear combinations of multiple variables. All of these techniques can "fit" the data but might not be good predictors and do not provide insight into relationships. Clustering is a method to partition the database into segments where each segment member shares similar qualities. Clustering techniques may include optimization algorithms to determine the maximum similarity among members within each group and a minimum similarity among members across the groups.

Obvious applications of rule induction in petroleum geological data mining include (1) predicting reservoir facies from multiple wireline logs, (2) predicting reservoir permeability from wireline logs, (3) predicting reservoir facies from multiple seismic attributes, and (4) predicting stratigraphic geometries and spatial architecture from quantified analog, outcrop, seismic, and numerical stratigraphic data.

## Rule Induction Algorithm

In rough set theory, concepts of *coverage* and *accuracy* are introduced to describe the uncertainty in probabilistic rules. In practice, every object in the data table may lead to a rule. The challenge is to identify significant rules, that is, accurate rules with high frequency of occurrence. Data mining has no *a priori* model, that is, no functional relationship between the data is assumed.

A feature of geological data is that most rules are probabilistic and not deterministic. Another aspect of geological data is that there are relatively few attributes from a classical data mining perspective. These two considerations were used to develop the concepts and terms described below.

### Data Table.

Table 1 (all Tables and Figures have been put at the back of the paper) shows a schematic *data table* for rule induction. The table has *M* observations, *N* condition attributes. Without loss of generality, it is assumed that there is a single decision attribute. Multiple decision attributes merely increase the number of rules and the configuration space. Initially, the condition attributes and decision attribute can have different values or categories, which must be transformed into discrete codes. The codes are based on statistical analysis of the data for each attribute. For example, the continuous variable of "gamma-ray count" may be transformed into 3 discrete codes of *high, average,* and *low.*

The number of categorical values that condition attribute *i, ca*$_i$, can take is $n_i, (i = 1,..., N)$, $n_o$ is the number of categorical value the decision attribute *de* can take. The number of discrete or categorical values the decision attribute and some condition attributes can take is fixed, or will be determined for a problem, say $n_1$= 3, but the codes themselves can be different; both {0,1,2} and {5,9,11} are valid codes for the three categorical values of attribute 1.

### Configuration and Data Coverges.

A *configuration* is a unique combination of the values of condition attributes. The total number of possible configurations is: $N_N = \prod_{i=1}^{N} n_i$ , where $n_i, (i = 1,..., N)$ is the number of categorical values of condition attribute *i*. Note that some configurations may not appear in the training data table. Further, define $N_i \left(= \prod_{k=1}^{i} n_k\right)$ as the cumulative number of configuration up to condition attribute $i, i = 1,..., N$ . Configuration index $j, j = 1,..., N_N$, is uniquely determined by the values $l_i, i = 1,..., N$ taken by each condition attribute, where $j = \sum_{i=1}^{N} (l_i - 1) \times \frac{N_N}{N_i} + 1$, and $l_i \in 1,..., n_i$ . The condition attribute values $l_i, i = 1,..., N$ from a configuration index *j* may be retrieved easily.

Every configuration may lead to potential rules in the system. Since the decision attribute can take $n_o$ outcomes, there are $N_N \times n_o$ potential rules for the system.

The occurrence of configuration *j* with outcome *o* is counted as $C_{j,o}$, ( $j, j = 1,..., N_N$ and $o, o = 1,..., n_o$ ) from the data table. If $C_{j,o} = 0,$ then there are no observations in the data table corresponding to that configuration and outcome pair. The number of observations *M* is:

$$M = \sum_{j=1}^{N_N} \sum_{o=1}^{n_o} C_{j,o}$$

The *configuration coverage* $C_j$ is defined as the total number of observations associated with configuration *j* , *i.e*.,

$$C_j = \sum_{o=1}^{n_o} C_{j,o}, \quad j = 1,..., N_N$$

The *decision coverage* $D_o$ is specified as the number of observations associated with outcome *o, i.e*.,

$$D_o = \sum_{j=1}^{N_N} C_{j,o}, \quad o = 1,..., n_o$$

The *relative (configuration) coverage* of configuration *j*, $\hat{C}_j$ is a measure used to determine closeness to sufficiency of observations in obtaining a reliable rule. $\hat{C}_j$ is undefined for $C_j = 0,$ equal to zero for $C_j = 1,$ and asymptotic to one as the *coverage* increases:

$$\hat{C}_j = 1 - \frac{1}{C_j}, \quad j = 1,..., N_N$$

The definition of *relative (configuration) coverage* could be different.

The *relative (decision) coverage* of a decision value, $\hat{D}_{j,o}$, is a measure of the proportion of the occurrence frequency of a configuration and decision outcome pair out of the number of observations associated with outcome *o, i.e.,*

$$\hat{D}_{j,o} = \frac{C_{j,o}}{D_o}, \quad j = 1,..., N_N, o = 1,..., n_o$$

These relative coverages will be used to adjust the importance of rules. For example, more weight will be placed on decision values with small global proportions to avoid cases where $C_{j,o}$ is small, but is close to $D_o$. In such a case, this configuration is very important to that decision.

**Conditional Probability and Accuracy.**

The *conditional probability* of each outcome value, $o, (o = 1,..., n_o)$, for configuration $j, (j = 1,..., N_N)$ is defined as:

$$p_{o|j} = \frac{C_{j,o}}{C_j}$$

Since there are $n_o$ possible decision attribute categories, a probability of $1/n_o$ implies no information. Any conditional probability different from $1/n_o$ entails preference in the decision category, *i.e.*, some knowledge. Specifically, as $p_{o|j}$ nears 1, configuration *j* implies decision category *o*. A closeness of $p_{o|j}$ to 0 means configuration *j* does not lead to decision category *o*. Conditional probabilities close to 1 or 0 contain equally important information for rule induction. The former leads to a positive rule relating configuration *j* to a specific decision category *o* and the latter leads to negative rules relating configuration *j* to some other decision category. For a data system, a $N_N$ by $n_o$ conditional probability table will be established.

The knowledge conveys by the conditional probability leads us to define a measure of *accuracy* as follows:

$$a_{o|j} = \begin{cases} p_{o|j} \bullet n_o - 1, & \text{if } p_{o|j} \le \dfrac{1}{n_o} \\ \dfrac{(p_{o|j} \bullet n_o - 1)}{n_o - 1}, & \text{otherwise} \end{cases}$$

The accuracy $a_{o|j} = 0$ if $p_{o|j} = 1/n_o$ (no knowledge); the accuracy $a_{o|j} = 1$ when we have $p_{o|j}$ equals to 1 (perfect positive rule); and, $a_{o|j} = -1$ when we have $p_{o|j}$ equals to 0 (perfect negative rule). A graph of accuracy versus probability is shown on Fig. 1.

**Significance.**

The *significance* is defined as a combined measure of the *accuracy* and *coverage*, which is used to provide a relative ranking to the rules. The significance is the product of the *accuracy* and the *relative configuration coverage, i.e.*,

$$S_{j,o} = \hat{C}_j \bullet a_{o|j}$$

This equation holds for negative accuracy. For positive accuracy, the significance is further modified to account for the *relative decision coverage, i.e.*,

$$S_{j,o} = \hat{C}_j \bullet a_{o|j} \bullet \hat{D}_{j,o}$$

As mentioned above, $\hat{D}_{j,o}$, is used to account the proportion of occurrence of an outcome with a specific configuration to the overall occurrence of that outcome, which serves like a normalization factor considering the unequally occurrence of decision attribute categories. Positive accuracy results from occurrence exceeding the average occurrence and the scaling using $\hat{D}_{j,o}$ will bring up the importance for rules associated with low occurrence outcome categories. Negative accuracy always results from occurrence lower than average level, the scaling factor is not applied for avoiding to bring the significance towards zero. For example, a zero occurrence $C_{j,o}$ leads to negative one in accuracy (perfect negative rule) and large negative significance value, applying $\hat{D}_{j,o}$ will revert it to zero significance.

**Rule Table.**

Rules are the conditional probabilities for those configuration and decision outcome pairs with large positive or negative significance. Positive *rules* are extracted from the configuration and decision outcome pairs ranked from positive 1 to zero by positive *significance*, and negative rules are extracted from the configuration and decision outcome pairs ranked from negative one to zero *significance*. A practitioner can decide whether to assign a threshold for significance. Positive rules provide evidence for the decision outcome and negative rules provide evidence against the decision. If a configuration and decision outcome pair's significance does not meet the criteria, the rule table entry is "missing" for lack of data.

We expect many "missing" entries for rule induction with real data. The "missing" entries will be eliminated or reduced based on compatible rules generated from subsets of data variables.

**Distinguishability.**

In a data mining exercise, there is an implicit assumption that all decision outcomes are distinguishable from the data. This assumption is not always true and a measure of *distinguishability* can assist in determining the decision that is distinct on the basis of the data table.

Once the rule induction is finished, one can construct a data table like the one shown in Table 2. For each observation in the training data set, a predicted probability is assigned by looking for the right configuration *j* and reading the corresponding conditional probability $p_{o|j}$.

The data table may be divided into $n_o$ classes with an indicator function:

$$ind(m;o) = \begin{cases} 1 & \text{if } o = \tilde{o}_m, \quad m = 1,..., M; o = 1,..., n_o \\ 0 & \text{otherwise} \end{cases}$$

where $\tilde{o}_m$ is the true outcome category for observation *m*.

The summation of $ind(m;o)$ is a counting of all outcomes related to decision value *o*.:

$$D_o = \sum_{m=1}^{M} ind(m;o)$$

The average predicated probability associated with prediction of *o* is:

$$\hat{p}_o = \frac{1}{D_o} \sum_{m=1}^{M} ind(m;o) \bullet p_{m,o}, \quad o = 1,...,n_o$$

where $p_{m,o}$ is the predicted probability value of observation *m* having decision category *o*.

As $\hat{p}_o$ approaches 1, prediction is good. As $\hat{p}_o$ goes to $p_o = \frac{D_o}{M}$, *i.e.*, the global proportion for decision outcome *o*, there is no information provided by the rules.

Therefore, the relative information value for prediction outcome *o* is defined as: $I_o = \frac{\hat{p}_o - p_o}{p_o}$

A composite information value can be defined as the average over all outcomes:

$$\bar{I} = \frac{1}{n_o} \sum_{o=1}^{n_o} I_o$$

The expected predicated probability and information measure is used to determine whether outcomes *o* and *o'* are distinct.

Let's consider merging two different decision outcomes *o* and *o'*, then we have $n_o - 1$ new decision outcomes and the predicted probability table (like Table 2) will be updated. Updating the predicted probability is straightforward. We derive $n_o - 1$ probability values from the previous $n_o$ values. One of them will be $p_{o|j} + p_{o'|j}$ when the actual outcome turns out to be *o* or *o'*.

Following the same procedure, new $I_o$ and the $\bar{I}$ values can be calculated. If we define $\bar{I}_{o:o'}$ as the information value with *o* and *o'* merged, the change in $\bar{I}$ is:

$$\Delta\bar{I}_{o:o'} = \bar{I}_{o:o'} - \bar{I}$$

$\Delta\bar{I}_{o:o'} > 0$ means that *o* and *o'* merged leads to an improved prediction. Thus, one should consider whether to treat outcomes *o* and *o'* together, *i.e.*, as indistinguishable, since the given training data table cannot differentiate them. One could seek to obtain additional data on a different data type that would distinguish those outcomes. $\Delta\bar{I}_{o:o'} < 0$ means that merging the outcomes leads to a poorer prediction. Thus, the outcomes should be treated as distinct.

*Lumping* or *merging* could be considered for all possible pairs of outcome values, and the $\Delta\bar{I}$ can be tabulated, plotted or ranked from high to low. The high positive values are candidates for merging while the low positive and negative values should be kept separate. The $\Delta\bar{I}_{o:o'}$ table can be visualized as a grayscale map. For example, Fig. 2 shows a map of the matrix pairing nine decision outcomes from the facies assignment example. The bigger the $\Delta\bar{I}$ value, the higher the distinguishability of outcome pair and the darker the color. The diagonal cells are the pairing of outcomes with themselves and have $\Delta\bar{I}$ values of 0. As shown in the Fig., outcomes 7, 8, 9 are very distinguishable from other outcomes. The upper left corner area with positive $\Delta\bar{I}$ values indicate that outcomes 1-2, 2-3, and 2-4 are not as distinguishable and might be candidates for combination.

**Data or Condition Attribute Value.**

The same measure of information can be used to consider the value (importance) of condition attributes or sets of condition attributes. All procedures described above are based on the entire data set with all condition attributes, but the procedures can be applied to any data subset with partial condition attributes.

For a *N* condition attribute data set, there are $2^N - 1 (= \sum_{l=1}^{N} C_N^l)$, (*l* is the number of condition attributes in the subset and $l = 1,...,N$ ) subsets. For each subset, a new data set is derived and the algorithm is applied. A new set of configurations of condition attributes is obtained and $p_{o|j}$ values are calculated for every decision category *o* for a given configuration *j*. For each individual decision category, the expected predication probability of the occurrence of decision category *o* when given the data entry having decision category *o* is also calculated as well as the information measure. The information value for each subset is, for example, $\bar{I}_{\{1,2,3,...,N\}} = \bar{I}$, for all attributes, $\bar{I}_{\{2,3,...,N\}}$ for leaving out condition attribute 1, $\bar{I}_{\{2\}}$ for just attribute 2, *etc*. These results can be plotted as a graph or tabulated in a table that ranks the value of the condition attributes (subsets). The largest values, or maximums of information in the subsets with the same number of condition attributes, would normally tend to increase as the number of attributes in the subsets increases. However, noisy data or attributes that do not yield additional information value could reduce $\bar{I}$.

**Completion of Rule Table.**

The rule (conditional probability) tables built from the entire data set are not complete since there may be no observations for some condition attribute configurations and there are some rules have been screened out due to small significance. A complete rule table is necessary for practical usage of rule induction algorithm and "missing" entries in the rule table should be filled in. Filling "missing" entries in the rule table will be accomplished based on significant rule derived from data subsets.

There are many subsets with various number of condition attributes, therefore, it is necessary to order the subsets in filling procedure. This ordering will be based on their information value. The ordered subsets will be considered

until all "missing" entries are filled in. The entries filled from subsets are flagged to indicate the subset useed.

The utilization of the rules is two-fold: (1) gain insight and make general observations from the rule table, and (2) use the conditional probability associated with each configuration and decision outcome pair in the conditional probability table to model the predictions probabilistically. These applications will be illustrated with two examples.

The probabilities are used in modeling. Of course, there is opportunity to reduce uncertainty by incorporating additional data sources in the modeling phase.

## Implementation

There are issues related to (1) the physical meaning and subjective understanding of the rules, (2) the usage of results, (3) the meaning of a probabilistic prediction, and (4) data integration. The significance has value between -1 to + 1. In principle, the potential rules with largest absolute significance values will be chosen as important rules, but there exists ambiguity of how to define big and small, *e.g*., is 0.9 close enough? When there are –1 and/or +1 in the significance list, those rules will be chosen without question. The others will depend on judgment. Fortunately, we do not care much about rules in the middle of the list.

The rule induction is only as good as the training data. But in contrast to other techniques, *e.g*. neural networks, each rule is clear and understandable to a practitioner as to its applicability. Another issue is the selection of categorical classes by application of thresholds to a continuous variable. This could be performed by optimization.

The rule induction procedure consists of the following steps: (1) prepare the data table and determine optimal classes for the conditioning attributes (discretization), (2) enumerate all possible configurations and outcomes, (3) calculate the *conditional probability, configuration coverage, decision coverage, accuracy*, and *significance*, (4) present a table for each outcome $o = 1,...,n_o$ with the configurations sorted from low to high *significance*, (5) apply cutoff criteria on significance to build a rule table, (6) evaluate information value for the condition attributes and/or decision outcome, and (7) fill up the "missing" entries in the rule table based on subsets of the conditioning attributes.

## Example: Synthetic Fluvial Channel System

In fluvial or deepwater depositional setting, the sandy facies occur as sand channels with associated levee and crevasse deposits. For simplicity without loss generality, we only consider sand-filled channels embedded with a matrix of shale in this example. The corresponding effective permeability depends on the orientation, sinuosity, width, thickness and geometry of the channels. The relationship between these geometric features and the effective permeability are *rules* to be extracted for future use. For this purpose, channel facies models are created with a variety of channel parameters such as *orientation, thickness*, and *sinuosity*, which are called condition attributes for the rule induction system. Flow simulation is conducted for each generated model and the effective permeability in *X* direction is taken as quality measure (decision categories) of the model.

The synthesized data set consists of the geometrical features of channel sands and the associated effective permeability. An object modeling program, *fluvsim* in *GSLIB*[10], was used to generate multiple facies models. The simulation domain is a 100 by 100 two-dimensional area. Three triangular distributions with non-overlapped value ranges are set for each of the three parameters, *i.e*., orientation (***O***), sinuosity (***S***), and width (***W***) (width/thickness ratio). Totally, there are 27 combinations (sets) of the parameter distributions. For each set of parameter distributions, parameters were randomly drawn and 50 facies models were generated. Thus, there are a total of 1350(=50×27) different facies models created. For each facies model, permeability values of 100 and 1 were assigned to channel sand and shale, respectively and a flow simulation was conducted with no-flow boundaries, using the *GSLIB*[10] program *flowsim*. The effective permeability in the *X* direction was taken as a measure of the quality or productivity of each facies model.

The data generated were organized as a data table with 1350 rows, each representing one facies model, and 4 columns, where the first three denotes the values of three condition attributes, *i.e.*, the *orientation* (***O***), *sinuosity* (***S***), and *width* (***W***) (*width/thickness ratio*) of the channels, and the fourth column was the value of the decision attribute, *i.e.*, the *effective permeability* (***K***) in *X* direction.

Fig. 3 shows one of 50 facies models for two of 27 parameter distributions. Fig. 4 shows the composite histogram of effective permeability of the 1350 facies models and the histograms of effective permeability of 50 facies models from the two parameter distributions shown in Fig. 3, respectively.

The rule induction method requires that the data be *binned* into categorical variables. It is reasonable and straightforward to classify the three geometrical parameters (condition attributes) into three categories consistent with the three distributions. The three categorical values of *orientation* (***O***) are defined as 2 (good: azimuth angle of $0 \pm 10$, which is aligned with the *X* direction), 1 (medium: azimuth angle of $45 \pm 10$) and 0 (bad: azimuth angle of $90 \pm 10$, which is perpendicular to the *X* direction). The three categorical values for *sinuosity* (***S***) are denoted as 2 (high: high deviation $30 \pm 5$ and short length of $20 \pm 5$ units), 1 (medium: medium deviation $20 \pm 5$ and medium length of $35 \pm 5$ units), 0 (low: low deviation $10 \pm 5$ and long length of $50 \pm 5$ units). The three categorical values for *width* (***W***) are assigned as 2 (large: $25 \pm 5$), 1 (medium: $15 \pm 5$) and 0 (small: $5 - 10$). Columns 2 to 4 of Table 3 list the codes of the parameters used in the generation of facies models.

The continuous variable of effective permeability has to be discretized into categorical values. The method of *binning* the permeability data could affect the final rule induction results. For simplicity, three categories were determined for the effective permeability by just inspecting the permeability histograms. Summary statistics of the effective permeability

for the 50 realizations of each configuration are listed in columns 5 and 6 of Table 3.

The three categories determined for effective permeability are: 2 (high: $k \geq 14$), 1 (medium: $4 < k < 14$), and 0 (low: $k \leq 4$). The number of instances and proportions of the observations on each of the three decision categories for each configuration are tabulated in the last three columns of Table 3. By inspecting Table 3, we can expect certain rules like those listed in Table 4. Note that the "No Matter ***W***" in Table 4 indicates that the width can have any of the categorical values.

## Results and Discussion

The algorithm was first applied to the 1350 by 4 synthesized channel sand/effective permeability data table and 27 configurations of conditioning attributes were extracted. Table 5 lists the configurations of condition attributes, the occurrence frequency $C_{j,o}$, and the significance $S_{j,o}$ of decision category $o$ when given configuration $j$. By construction, all 27 possible configurations of condition attributes in this example have equal, non-zero sample coverage, *i.e.*, $C_j = 50$. This will not necessarily be the case for most data systems in practice.

The full set of rules is sorted by significance for each decision category, which is listed in Table 6. The positive rules shown in Table 7 are taken from the top portion of Table 6, which has the highest positive significance values. The negative rules as tabulated in Table 8 are taken from the bottom portion of Table 6, which have the lowest significance values.

The rule set is used to estimate the permeability class of training objects and Table 9 lists the accuracy rate of the classification. Even though this accuracy rate derived from the training set is not a good measure to evaluate the predictability of the modeling, it still can provide some indication of the prediction model. The rules are quite reliable and predictable.

For this synthesized data set, the training data covers all configurations by construction. Therefore, the rule set is complete from the set with all condition attributes and there are no blanks in the rule sets to be filled. Fig. 5 shows the information value of all 7 subsets of this synthesized data set and the full set with all three condition attributes has the highest information value. In the situation of the full set does not cover all configurations of condition attributes, such information values will be used to rank the subsets for retrieving corresponding rules from the subsets. Fig. 6 shows the information value change when decision classes are lumped/merged pair-wise for the full condition attribute set. The diagonal elements set the basis for comparison that corresponds to a situation without decision-class lumping. Off-diagonal elements show the change in the information values when the attributes are lumped. Darker color than the diagonal elements indicate an increase in the value of information and those lighter colors denote a decrease in the information value.

## Discussion

The proposed algorithm works well for the example shown here and for other examples that are not shown because of space considerations. The most important positive and negative rules are retrieved successfully. The proposed *significance* definition combines measures of *accuracy* and *coverage* and serves as a quality measure of rules. Also, the significance identifies positive and negative rules, similar to the positive region and negative region in rough sets.

The proposed rule induction technique is suited to geological data where most attributes are significant. The proposed significance measure can be used in combination with other rule induction techniques and serves as a ranking measure to identify the most important rules. In general, however, the algorithm will need to be extended to include attribute reduction.

Careful examination of a data table may lead an experienced person to infer similar, if not the same rules; however, there are many advantages to automatic rule induction. The procedure works for very large datatables with many attributes, it is repeatable, and avoids personal biases. It could also lead to nonintuitive, but meaningful data relations. The effort in making predictions can be greatly reduced.

## References

1. Balch RS, Stubbs BS, Weiss WW, and Wo S (1999) Using artificial intelligence to correct multiple seismic attributes to reservoir properties. In: SPE Annual Tech Conference and Exhibition. Houston, TX, October 5-8 1999, SPE 56733
2. Berry MJA, Linoff G (1997) Data Mining Techniques, For Marketing, Sales, and Customer Support. Wiley Computer Publishing, John Wiley & Sons Inc., New York
3. Bishop CM (1995) Neural Networks for Pattern Recognition. Oxford University Press, New York
4. Bradley S, Fayyad U, Mangasarian OL (1999) Mathematical programming for data mining formulations and challenges. INORMS J. on Computing 11:217-238
5. Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and Regression Tree. Chapman and Hall, New York
6. Buntine W (1992) Learning classification trees. Statistics and Computing, 63-73
7. Cabcna P, Hadjinian P, Stadler R, Verhees J, Zanasi A (1997) Discovering Data Mining from Concept to Implementation. Prentice Hall, New York
8. Two Crows Corporation (1999) Introduction to Data Mining and Knowledge discovery
9. Curram SP, Mingers J (1994) Neural networks, decision tree induction and discriminant analysis: an empirical comparison. Journal of the Operational Research Society 45:440-450
10. Deutsch CV, Journel AG (1998) GSLIB: Geostatistical Software Library: an User's Guide. Oxford University Press, New York, 2nd Ed.
11. Fayyad U, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (1996) Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge, MA

12. Hook JR, Nieto JA, Kalkomey CT, Ellis D (1994) Facies and permeability prediction from wireline logs and core: a north sea case study. In: 35th Annual SPWLA Logging Symposium,Tulsa
13. Kohonen T (1989) Self-Organization and Association Memory (3rd edition). Springer-Verlag
14. Kohonen T (1995) Self-Organization Maps. Springer-Verlag, Heidelberg
15. Lee SH, Datta-Gupta A (1999) Electrofacies characterization and permeability predictions in carbonate reservoirs: roles of multivariate analysis and nonparametric regression. In: SPE Annual Tech Conference and Exhibition, Houston, TX, October 5-8, SPE 56658
16. Lira TS. Loh Y, Shih YS (1999) A comparison of prediction accuracy, complexity and training time of thirty-three old and new classification algorithms. The Machine Learning Journal, 1-27
17. Lin TY, Cercone N (1998) Rough Sets and Data Mining: Analysis for Imprecise Data. KluwerAcademic Publisher
18. Menon S, Sharda R (1999) Data mining update: new modes to persue old objectives. ORMS Today, 26-29
19. Murthy SK, Kasif S, Salzberg S (1994) A system for induction of oblique decision trees. Journal of Artificial Intelligence Research, 2:1-33
20. Orlowska E (1998) Incomplete information: Rough set analysis. In: Vol 13, Studies in Fuzziness and soft computing, p620. Physica Verlag
21. Pal SK, Skowron A (1999) Rough Fuzzy Hybridization: A New Trend in Decision-Making. Springer Verlag, New York
22. Pawlak Z (1991) Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publisher, Dordrecht
23. Polkowski L, Skowron A (1998) Rough sets and current trends in computing. In: First international Conference, RSCTC'98, p601, Warsaw, Poland, Springer
24. Polkowski L, Skowron A (1998) Rough sets in knowledge discovery, i: Methodology and Application. In: Vol 18, Studies in Fuzziness and Soft Computing, p570. Physica Verlag
25. Polkowski L, Skowron A (1998) Rough sets in knowledge discovery, ii: Applications, Case Studies, and Software Ssystems. In: Vol 19, Studies in Fuzziness and Soft Computing, page 601. Physica Verlag
26. Quinlan JR (1986) Induction of decision tree. Machine Learning, 1:81-106
27. Ripley BD (1996) Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge
28. Scheevel JR, Payrazyan K (1999) Principal component analysis applied to 3D seismic data for reservoir property estimation. In: SPE Annual Tech Conference and Exhibition, Houston, TX, October 5-8. SPE 56734
29. Shih YS (1999) Families of splitting criteria for classification trees. Statistics and Computing
30. Slowinski R, Slowinski R (1992) Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publisher, Dordrecht
31. Ziarko WP (1993) Rough sets, Fuzzy Sets, and Knowledge Discovery. In: Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD '95), Banff, Alberta, Canada, 12-15 October. Springer-Verlag

| | $ca_1$ | $ca_2$ | ... | $ca_{N-1}$ | $ca_N$ | *de* |
|---|---|---|---|---|---|---|
| $o_1$ | 0 | 2 | ... | 1 | 2 | 1 |
| $o_2$ | 2 | 1 | ... | 1 | 2 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| $o_{M-1}$ | 2 | 2 | ... | 0 | 1 | 0 |
| $o_M$ | 1 | 0 | ... | 2 | 2 | 2 |

**Table 1**. A typical data table for rule induction *ca,* condition attribute; *de*, decision attribute; *o*, observation; *N*, number of condition attributes; *M*, number of observations

| *Observation (m)* | *Predicted probability from the rules* | *real outcome* $(\tilde{o}_m)$ |
|---|---|---|
| *1* | $p_{1.1}, p_{1.2}, ..., p_{1.n_c}$ | $\tilde{o}_1$ |
| *2* | $p_{2.1}, p_{2.2}, ..., p_{2.n_c}$ | $\tilde{o}_2$ |
| *...* | *...* | *...* |
| *m* | $p_{m.1}, p_{m.2}, ..., p_{m.n_c}$ | $\tilde{o}_{M-1}$ |
| *...* | *...* | *...* |
| *M-1* | $p_{M-1.1}, p_{M-1.2}, ..., p_{M-1.n_c}$ | $\tilde{o}_{M-1}$ |
| *M* | $p_{M.1}, p_{M.2}, ..., p_{M.n_c}$ | $\tilde{o}_M$ |

**Table 2**. Predicted probability from the rules

| *No*[a] | *O* | *S* | *W*[b] | *mean K* | *Std K* | *Klow* | *Kmed* | *Khigh*[c] |
|---|---|---|---|---|---|---|---|---|
| *1* | *0* | *0* | *0* | *3.90* | *1.53* | *32 (64)* | *18 (36)* | *0 (0)* |
| *2* | *0* | *0* | *1* | *2.22* | *0.48* | *49 (98)* | *1 (2)* | *0 (0)* |
| *3* | *0* | *0* | *2* | *1.91* | *0.31* | *50(100)* | *0 (0)* | *0 (0)* |
| *4* | *0* | *1* | *0* | *8.50* | *2.57* | *4 (8)* | *46 (92)* | *0 (0)* |
| *5* | *0* | *1* | *1* | *5.14* | *2.53* | *23 (46)* | *26 (52)* | *1 (2)* |
| *6* | *0* | *1* | *2* | *4.29* | *2.87* | *34 (68)* | *16 (32)* | *0 (0)* |
| *7* | *0* | *2* | *0* | *9.40* | *2.30* | *1 (2)* | *49 (98)* | *0 (0)* |
| *8* | *0* | *2* | *1* | *9.69* | *4.26* | *3 (6)* | *39 (78)* | *8 (16* |
| *9* | *0* | *2* | *2* | *8.16* | *4.75* | *14 (28)* | *30 (60)* | *6 (12* |
| *10* | *1* | *0* | *0* | *8.47* | *2.75* | *2 (4)* | *47 (94)* | *1 (2)* |
| *11* | *1* | *0* | *1* | *6.94* | *5.05* | *23 (46)* | *21 (42)* | *6 (12* |
| *12* | *1* | *0* | *2* | *4.55* | *4.01* | *34 (68)* | *12 (24)* | *4 (8)* |
| *13* | *1* | *1* | *0* | *8.21* | *1.95* | *0 (0)* | *50 (100)* | *0 (0)* |
| *14* | *1* | *1* | *1* | *7.88* | *3.28* | *6 (12)* | *42 (84)* | *2 (4)* |
| *15* | *1* | *1* | *2* | *5.28* | *3.49* | *25 (50)* | *23 (46)* | *2 (4)* |
| *16* | *1* | *2* | *0* | *5.02* | *1.00* | *5 (10)* | *45 (90)* | *0 (0)* |
| *17* | *1* | *2* | *1* | *5.66* | *1.76* | *6 (12)* | *44 (88)* | *0 (0)* |
| *18* | *1* | *2* | *2* | *5.97* | *2.52* | *15 (30)* | *35 (70)* | *0 (0)* |
| *19* | *2* | *0* | *0* | *17.46* | *2.17* | *0 (0)* | *2 (4)* | *48 (96* |
| *20* | *2* | *0* | *1* | *18.74* | *4.05* | *0 (0)* | *9 (18)* | *41 (82* |
| *21* | *2* | *0* | *2* | *21.19* | *5.80* | *0 (0)* | *6 (12)* | *44 (88* |
| *22* | *2* | *1* | *0* | *8.66* | *1.44* | *0 (0)* | *50 (100)* | *0 (0)* |
| *23* | *2* | *1* | *1* | *8.54* | *2.64* | *3 (6)* | *46 (92)* | *1 (2)* |
| *24* | *2* | *1* | *2* | *9.08* | *4.17* | *8 (16)* | *34 (68)* | *8 (16* |
| *25* | *2* | *2* | *0* | *3.97* | *0.77* | *29 (58)* | *21 (42)* | *0 (0)* |
| *26* | *2* | *2* | *1* | *3.66* | *1.06* | *35 (70)* | *15 (30)* | *0 (0)* |
| *27* | *2* | *2* | *2* | *3.23* | *0.85* | *42 (84)* | *8 (16)* | *0 (0)* |

a: index of configuration; b: O, Orientation; S, Sinuosity; W, width; c: KLow, k≤4; KMedium, 4<k<14; KHigh, k≥14;

**Table 3**. Parameter used in facies models generation and the statistic of the effective permeability values

| | | | | | | |
|---|---|---|---|---|---|---|
| *Good O (2)* | *and* | *low S (0)* | *and* | *No matter W* | *leads to* | *high K (2)* |
| *Bad O (0)* | *and* | *low S (0)* | *and* | *No matter W* | *leads to* | *low K (0)* |
| *Good O (2)* | *and* | *High S (2)* | *and* | *No matter W* | *leads to* | *low K (0)* |
| *Medium O* | | | | | *mostly leads to* | *medium K* |

**Table 4**. Some expected rules for the fluvial system

| No | Condition Attribute | | | Occurrence $C_{i,o}$ | | | Significance $S_{i,o}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | O | S | W | Klow | Kmed | Khigh | Klow | Kmed | Khigh |
| 1 | 0 | 0 | 0 | 32 | 18 | 0 | 0.640 | 0.000 | -1.000 |
| 2 | 0 | 0 | 1 | 49 | 1 | 0 | 0.980 | -0.980 | -1.000 |
| 3 | 0 | 0 | 2 | 50 | 0 | 0 | 1.000 | -1.000 | -1.000 |
| 4 | 0 | 1 | 0 | 4 | 46 | 0 | 0.000 | 0.920 | -1.000 |
| 5 | 0 | 1 | 1 | 23 | 26 | 1 | 0.000 | 0.000 | -0.980 |
| 6 | 0 | 1 | 2 | 34 | 16 | 0 | 0.680 | 0.000 | -1.000 |
| 7 | 0 | 2 | 0 | 1 | 49 | 0 | -0.980 | 0.980 | -1.000 |
| 8 | 0 | 2 | 1 | 3 | 39 | 8 | 0.000 | 0.780 | 0.000 |
| 9 | 0 | 2 | 2 | 14 | 30 | 6 | 0.000 | 0.600 | 0.000 |
| 10 | 1 | 0 | 0 | 2 | 47 | 1 | 0.000 | 0.940 | -0.980 |
| 11 | 1 | 0 | 1 | 23 | 21 | 6 | 0.000 | 0.000 | 0.000 |
| 12 | 1 | 0 | 2 | 34 | 12 | 4 | 0.680 | 0.000 | 0.000 |
| 13 | 1 | 1 | 0 | 0 | 50 | 0 | -1.000 | 1.000 | -1.000 |
| 14 | 1 | 1 | 1 | 6 | 42 | 2 | 0.000 | 0.840 | 0.000 |
| 15 | 1 | 1 | 2 | 25 | 23 | 2 | 0.500 | 0.000 | 0.000 |
| 16 | 1 | 2 | 0 | 5 | 45 | 0 | 0.000 | 0.900 | -1.000 |
| 17 | 1 | 2 | 1 | 6 | 44 | 0 | 0.000 | 0.880 | -1.000 |
| 18 | 1 | 2 | 2 | 15 | 35 | 0 | 0.000 | 0.700 | -1.000 |
| 19 | 2 | 0 | 0 | 0 | 2 | 48 | -1.000 | 0.000 | 0.960 |
| 20 | 2 | 0 | 1 | 0 | 9 | 41 | -1.000 | 0.000 | 0.820 |
| 21 | 2 | 0 | 2 | 0 | 6 | 44 | -1.000 | 0.000 | 0.880 |
| 22 | 2 | 1 | 0 | 0 | 50 | 0 | -1.000 | 1.000 | -1.000 |
| 23 | 2 | 1 | 1 | 3 | 46 | 1 | 0.000 | 0.820 | -0.980 |
| 24 | 2 | 1 | 2 | 8 | 34 | 8 | 0.000 | 0.680 | 0.000 |
| 25 | 2 | 2 | 0 | 29 | 21 | 0 | 0.580 | 0.000 | -1.000 |
| 26 | 2 | 2 | 1 | 35 | 15 | 0 | 0.700 | 0.000 | -1.000 |
| 27 | 2 | 2 | 2 | 42 | 8 | 0 | 0.840 | 0.000 | -1.000 |

**Table 5**. Potential rules for fluvial channel sand data set

| index | ca | | | de | $s_{j,o}$ | index | ca | | | de | $s_{j,o}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | O | S | W | | | | O | S | W | | |
| 19 | 2 | 0 | 0 | 2 | 0.262 | 27 | 2 | 2 | 2 | 1 | -0.52 |
| 21 | 2 | 0 | 2 | 2 | 0.209 | 11 | 1 | 0 | 1 | 2 | -0.64 |
| 20 | 2 | 0 | 1 | 2 | 0.174 | 14 | 1 | 1 | 1 | 0 | -0.64 |
| 3 | 0 | 0 | 2 | 0 | 0.112 | 9 | 0 | 2 | 2 | 2 | -0.64 |
| 2 | 0 | 0 | 1 | 0 | 0.107 | 17 | 1 | 2 | 1 | 0 | -0.64 |
| 27 | 2 | 2 | 2 | 0 | 0.072 | 21 | 2 | 0 | 2 | 1 | -0.64 |
| 22 | 2 | 1 | 0 | 1 | 0.068 | 16 | 1 | 2 | 0 | 0 | -0.7 |
| 13 | 1 | 1 | 0 | 1 | 0.068 | 4 | 0 | 1 | 0 | 0 | -0.76 |
| 7 | 0 | 2 | 0 | 1 | 0.064 | 12 | 1 | 0 | 2 | 2 | -0.76 |
| 10 | 1 | 0 | 0 | 1 | 0.058 | 23 | 2 | 1 | 1 | 0 | -0.82 |
| 23 | 2 | 1 | 1 | 1 | 0.055 | 8 | 0 | 2 | 1 | 0 | -0.82 |
| 4 | 0 | 1 | 0 | 1 | 0.055 | 15 | 1 | 1 | 2 | 2 | -0.88 |
| 16 | 1 | 2 | 0 | 1 | 0.052 | 19 | 2 | 0 | 0 | 1 | -0.88 |
| 17 | 1 | 2 | 1 | 1 | 0.049 | 10 | 1 | 0 | 0 | 0 | -0.88 |
| 26 | 2 | 2 | 1 | 0 | 0.043 | 14 | 1 | 1 | 1 | 2 | -0.88 |
| 14 | 1 | 1 | 1 | 1 | 0.043 | 7 | 0 | 2 | 0 | 0 | -0.94 |
| 12 | 1 | 0 | 2 | 0 | 0.039 | 5 | 0 | 1 | 1 | 2 | -0.94 |
| 6 | 0 | 1 | 2 | 0 | 0.039 | 10 | 1 | 0 | 0 | 2 | -0.94 |
| 8 | 0 | 2 | 1 | 1 | 0.035 | 2 | 0 | 0 | 1 | 1 | -0.94 |
| 1 | 0 | 0 | 0 | 0 | 0.033 | 23 | 2 | 1 | 1 | 2 | -0.94 |
| 18 | 1 | 2 | 2 | 1 | 0.026 | 25 | 2 | 2 | 0 | 2 | -1.0 |
| 25 | 2 | 2 | 0 | 0 | 0.024 | 3 | 0 | 0 | 2 | 2 | -1.0 |
| 24 | 2 | 1 | 2 | 1 | 0.024 | 26 | 2 | 2 | 1 | 2 | -1.0 |
| 9 | 0 | 2 | 2 | 1 | 0.016 | 21 | 2 | 0 | 2 | 0 | -1.0 |
| 15 | 1 | 1 | 2 | 0 | 0.014 | 7 | 0 | 2 | 0 | 2 | -1.0 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *5* | *0* | *1* | *1* | *1* | *0.009* | *16* | *1* | *2* | *0* | *2* | *-1.0* |
| *11* | *1* | *0* | *1* | *0* | *0.009* | *13* | *1* | *1* | *0* | *2* | *-1.0* |
| *5* | *0* | *1* | *1* | *0* | *0.009* | *13* | *1* | *1* | *0* | *0* | *-1.0* |
| 15 | 1 | 1 | 2 | 1 | 0.005 | 17 | 1 | 2 | 1 | 2 | -1.0 |
| *25* | *2* | *2* | *0* | *1* | *0.003* | *19* | *2* | *0* | *0* | *0* | *-1.0* |
| *11* | *1* | *0* | *1* | *1* | *0.003* | *4* | *0* | *1* | *0* | *2* | *-1.0* |
| *1* | *0* | *0* | *0* | *1* | *0.001* | *2* | *0* | *0* | *1* | *2* | *-1.0* |
| *6* | *0* | *1* | *2* | *1* | *-0.04* | *1* | *0* | *0* | *0* | *2* | *-1.0* |
| *18* | *1* | *2* | *2* | *0* | *-0.1* | *22* | *2* | *1* | *0* | *2* | *-1.0* |
| *26* | *2* | *2* | *1* | *1* | *-0.1* | *20* | *2* | *0* | *1* | *0* | *-1.0* |
| *9* | *0* | *2* | *2* | *0* | *-0.16* | *18* | *1* | *2* | *2* | *2* | *-1.0* |
| *12* | *1* | *0* | *2* | *1* | *-0.28* | *6* | *0* | *1* | *2* | *2* | *-1.0* |
| *20* | *2* | *0* | *1* | *1* | *-0.46* | *27* | *2* | *2* | *2* | *2* | *-1.0* |
| *24* | *2* | *1* | *2* | *2* | *-0.52* | *3* | *0* | *0* | *2* | *1* | *-1.0* |
| *8* | *0* | *2* | *1* | *2* | *-0.52* | *22* | *2* | *1* | *0* | *0* | *-1.0* |
| *24* | *2* | *1* | *2* | *0* | *-0.52* | | | | | | |

**Table 6**. Full set of rules sorted according to significance value

| | | | | | | |
|---|---|---|---|---|---|---|
| *Good O (2)* | *and* | *low S (0)* | *and* | *No matter W* | *leads to* | *high K (2)* |
| *Bad O (0)* | *and* | *low S (0)* | *and* | *large W (2)* | *leads to* | *low K (0)* |
| *Bad O (0)* | *and* | *low S (0)* | *and* | *large W (1)* | *leads to* | *low K (0)* |
| *Good O (2)* | *and* | *high S (2)* | *and* | *large W 2)* | *leads to* | *low K (0)* |
| *Good O (2)* | *and* | *medium S (1)* | *and* | *no matter W* | *leads to* | *medium K* |
| *Medium O (1)* | *and* | *no matter S* | *and* | *small W (0)* | *leads to* | *medium K* |
| *Bad O (0)* | *and* | *high S (2)* | *and* | *small W (0)* | *leads to* | *medium K* |

**Table 7.** The derived positive rules

| | | | | | | |
|---|---|---|---|---|---|---|
| *Good O (2)* | *and* | *medium S (1)* | *and* | *small W (0)* | *never leads* | *low K (0)* |
| *Bad O (0)* | *and* | *small S (0)* | *and* | *large W (2)* | *never leads* | *low K (0)* |
| *Good O (2)* | *and* | *high S (2)* | *and* | *large W (2)* | *never leads* | *high K (2)* |
| *Bad O (0)* | *and* | *medium S (1)* | *and* | *large W (2)* | *never leads* | *high K (2)* |
| *Medium O (1)* | *and* | *high S (2)* | *and* | *large W (2)* | *never leads* | *high K (2)* |
| *Good O (2)* | *and* | *small S (0)* | *and* | *medium W* | *never leads* | *low K (0)* |

**Table 8.** The derived negative rules

| | *1* | *2* | *3* | *accuracy* |
|---|---|---|---|---|
| *1* | *353* | *90* | *0* | *79.68%* |
| *2* | *135* | *583* | *17* | *79.32%* |
| *3* | *12* | *27* | *133* | *77.33%* |

**Table 9.** Results of classification

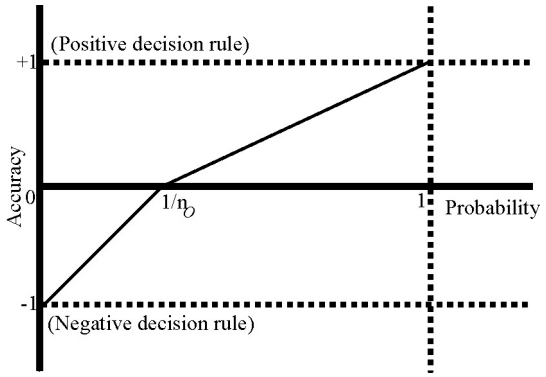


**Fig. 1 -** Relationship between accuracy and conditional probability

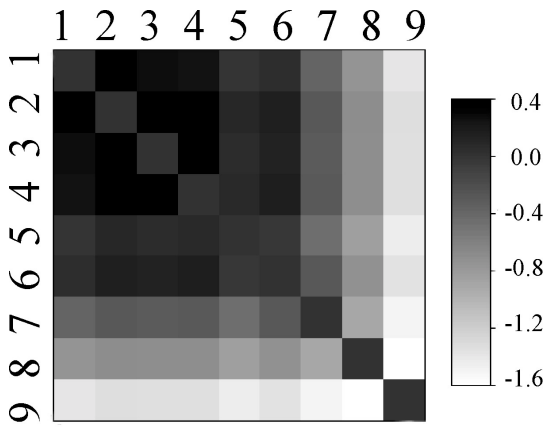


**Fig. 2 -** Information changes for lumped/merged decision categories

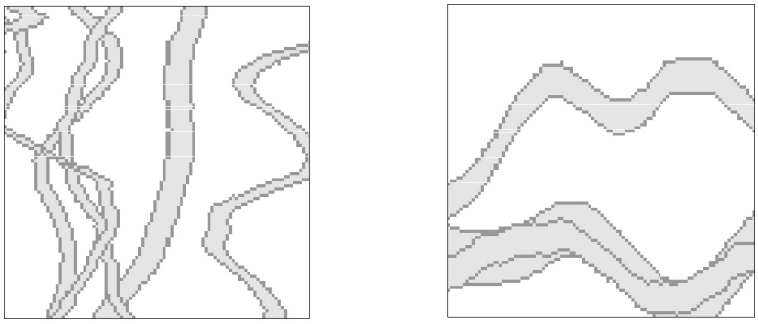


**Fig. 3 -** Geological model of one realization from parameter distribution 1 and 20.

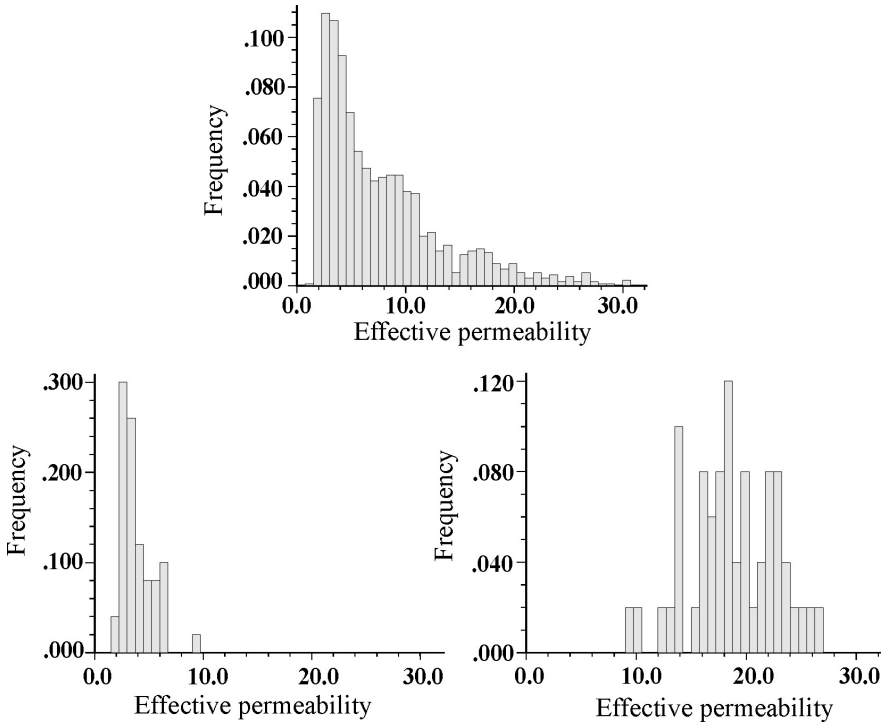


**Fig. 4 -** *Left,* Composite histogram of efective permeability of overall 1350 facies model; histograms of effective permeability of 50 realizations for parameter distribution 1 (*middle*) and 20 (*right*)

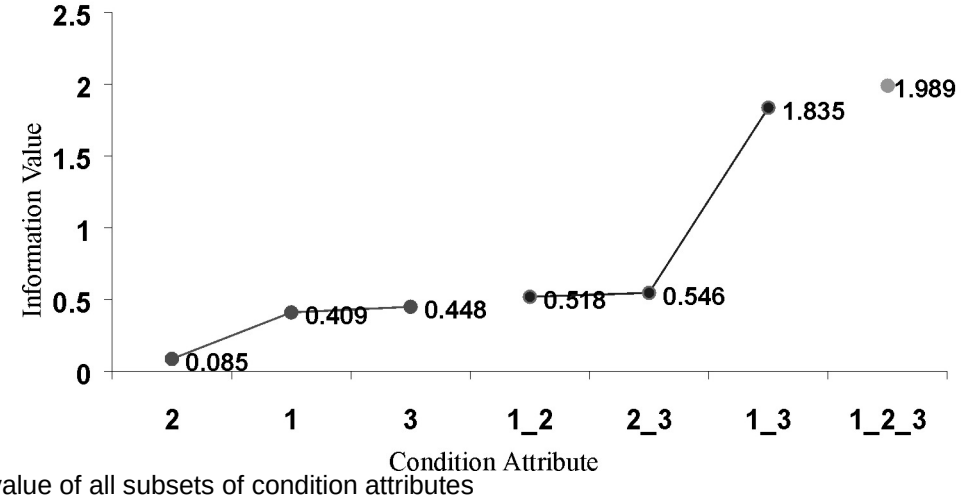


**Fig. 5 -** Information value of all subsets of condition attributes

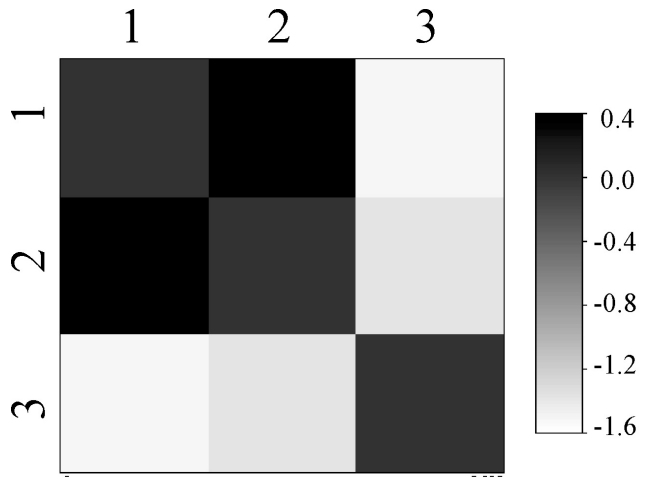


**Fig. 6 -** Information changes for lumped/merged decision categories