# An ID3-improved approach of for optimum rule mining through granular computing search algorithm

**Hadis Samadi Alinia**
Dept. of Surveying and Geomatics Eng.
University of Tehran, Tehran, Iran.
alinia@ut.ac.ir

**Mahmoud Reza Delavar**
Dept. of Surveying and Geomatics Eng
Center of Excellence in Geomatics Eng and Disaster Management
University of Tehran, Tehran, Iran
medelavar@ut.ac.ir

**Yiyu Yao**
Dept. of Computer Science
University of Regina, Saskatchewan, Canada
yyao@cs.uregina.ca

*Abstract*— **Rule induction is an area of machine learning in which formal rules are extracted from a set of observations or training dataset. Inducted rules can be expressed as a final result of the decision tree in which each branch represents a possible scenario of decision and its outcome. Existing decision learning algorithms like Iterative Dichotomiser (ID3) is an attribute centered method which may introduce unnecessary attributes in the classification rules. To overcome the problem, coverage and confidence measures are applied to select the most promising attribute-value at each step. The proposed approach is granule centered in which, instead of focusing on the selection of a suitable partition, i.e., a family of granules is defined, at each step, by values of an attribute. This paper is concentrated on the selection of a single granule. The decision tree learning algorithm ID3 and granular network are successfully applied for information table of test dataset of seismic vulnerability of urban areas in Tehran, capital of Iran.**

***Keywords: ID3 decision tree; Granule network; Uncertainty; Consistent classification; Seismic vulnerability***

## I. INTRODUCTION

Classification is one of the main tasks in machine learning, data mining, and pattern recognition (Mitchell, 1997). One of the tasks of knowledge discovery and data mining is to search for knowledge, patterns, and regularities derivable from data stored in a database (Yao, 2001). Data mining uses inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules.

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (Quinlan, 1982). It is an algorithm to construct a decision tree in which, the greedy search algorithm is used to test each attribute at every tree node through the given sets. Entropy as a metric information gain, measures the amount of information in an attribute to select one which is the most informative in each node. An important feature of ID3 algorithms is that when splitting a node, an attribute is chosen based on only information about this node, but not on any other nodes at the same level. The consequence is that in the decision tree, different nodes at the same level may use different attributes, and moreover, the same attribute with all possible values may be used at different levels. The use of local optimal criteria makes it use unnecessary attributes in this process of construction of the decision tree and also makes it difficult to judge the overall quality of the partial decision tree during its construction process.

The lack of this local optimization approach can be solved by using a global optimization method which chooses a attribute-value in favour of all nodes at the same level. A granular computing approach (Yao and Yao, 2002) is proposed as a granule network to extend the existing classification algorithms. It is a label of theories, methodologies, techniques, and tools that makes use of granules in the process of problem solving. In this approach, many measures have also been proposed and studied to quantify various aspects of rules, such as confidence, uncertainty, applicability, quality, accuracy, usefulness and interestingness (Yao, 2000).

To demonstrate the potential of granular computing to other existing problem solving approaches, we consider the problem of seismic vulnerability. Seismic vulnerability classification for Tehran, capital of Iran, where several known and unknown active faults located on and huge earthquakes will permeate human settlement there, is important. Production of seismic vulnerability map could help local and national disaster management organizations to create and implement a plan to promote awareness of seismic vulnerability and implement damage reduction measures in Tehran (Alinia and Delavar, 2009).

Production of seismic vulnerability map generally depends on various criteria. So, for knowledge discovery of seismic vulnerability of urban areas, induction rules from a training dataset of expert's opinion is proposed.

The scheme of this work starts with the design of the granule decision tree and ID3 from an information table of training dataset of 20 urban areas to induct classification rules. Then, the extracted rules are applied to the dataset of 50 test sites in Tehran, to evaluate the precision of classification with respect to an expert opinion. Accuracy of the two approaches is compared. This paper is in continuous of last paper about introducing of application of granule algorithm for seismic vulnerability classification. In this paper, both ID3 and granule network algorithms are implemented and results are compared.

## II. CLASSIFICATION ALGORITHMS

Two essential tasks in machine learning and data mining are the representation of objects and the identification of forms and types of knowledge to be mined. An attribute-value language provides a simple and useful tool for dealing with the two tasks.

Objects are represented in terms of their values on a set of attributes. More specifically, information about objects is summarized in an information table defined by (Pawlak, 1991):

$$S = (U, At, L, \{V_a | a \in At\}, \{I_a | a \in At\})$$

where U is a finite non-empty set of objects,
At is a finite non-empty set of attributes,
L is a language defined by using attributes in At,
Va is a non-empty set of values of $a \in At$,
$I_a$: U → Va is an information function that maps an object of U to exactly one possible value of attribute a in Va.

In the language L, an atomic formula is given by a = v, where L is a language defined using attributes in At, $V_a$ is a non-empty set of values. For $a \in At$ and $v \in V_a$, formulas can be formed by logical operators such as negations, the conjunction and disjunction. If $\phi$ and $\psi$ are formulas, then $\sim \phi, \phi \wedge \psi$ and $\phi \vee \psi$ can be calculated.

If $\phi$ is a formula, the set, is $m_s(\phi) = \{x \in U | x| = \phi\}$, is called the meaning of the formula $\phi$ in S which is interpreted using subsets of objects. Generally, a concept is described jointly by its intension and extension. $\phi$ is intension of concept ($\phi$, m ($\phi$)), and m ($\phi$) is the extension of concept ($\phi$,m($\phi$)).

In a supervised classification, each object is associated with a unique and predefined class label. Suppose an information table is used to describe a set of objects. Without loss of generality, we assume that there is a unique attribute class taking class labels as its value. The set of attributes is expressed as At = F ∪{class}, where F is the set of attributes used to describe the objects. The goal is to find classification rules of the form, $\phi \Rightarrow$class $=c_i$ , where $\phi$ is a formula over F and $c_i$ is a class label.

### A. ID3 Algorithm

ID3 algorithm is a partition-based algorithm in which child nodes are pairwisely disjoint with each other and naturally cover their parent granule. The algorithm selects the most promising attributes to split the examined granules at each level, and each granule is labelled by one of the possible values of the selected attribute. If the attribute perfectly classifies the training sets then ID3 stops; otherwise it recursively operates on the n, the number of possible values of an attribute, partitioned subsets to get their "best" attribute. A statistical property, called information gain, is used for determining how well a given attribute separates training examples into targeted classes. The entropy of information is computed by Equation (1) (Shannon, 1948):

$$E_n = \sum p_n \left(\sum -p_i \log_2 p_i\right) \quad (1)$$

where $p_n$: *Probability for all data*
$p_i$: *probability for each case*$(i)$

At step 0, $\sum p_n = 1$, $E_n = I(i) = \sum_i -p_i \log_2 p_i$

Each node of the ID3 decision tree is labelled by an attribute, and each branch is labelled by a value of the parent attribute.

### B. Granular Computing Approach

Granular computing ('GrC') was first introduced by T.Y. Lin, in 1997. Yao proposed a method of construction of a granule network to practice machine learning problem. In a granule network, each node is labeled by smaller granules labeled by an atomic formula. In addition, the smaller granule is obtained by selecting those objects of the larger granule that satisfy the atomic formula. The family of the smallest granules thus forms a conjunctively definable covering of the universe. Opposite of ID3 algorithm, proper granulation is capable of removing some redundant and irrelevant information and at the same time facilitates getting rid of over fitting problem (Yuchun and Zhang, 2005).

This algorithm enables user to classify objects with overlap concepts. At each step, only the most suitable granule defined by an attribute-value pair is selected, instead of a partition. Granulation of a universe involves dividing the universe into subsets or grouping individual objects into clusters. A granule is a subset of the universe. A family of granules that contains every object in the universe is called a granulation of the universe.

Partition and covering are two simple and commonly used granulations of universe. A partition consists of disjoint subsets of the universe, and a covering consists of possibly overlap subsets. Partitions are a special type of coverings. The main difference between a partition solution and a covering solution is that an object is only classified by one rule in the partition based solution, while an object may be classified by more than one rule. In this paper, seismic vulnerability granule network as an example demonstrates the covering-based solution.

In granular computing, information granules are first constructed and computations are subsequently carried out with the granules (Yao, 2000).

#### 1) Induction of classification rules

For classification tasks, it is assumed that each object is associated with a unique class label. Objects can be divided into classes which form a granulation of the universe.

To construct a granule network it is required that the universe is divided into grouping or partitions of the same class with atomic formula of the class label. Then, all partitions of the atomic formula of attribute-value and the meaning of the formula are constructed. To select the most promising formula and its related granule at each step, some quantitative measures are used to estimate the quality of a rule:

##### a) Measurements on single granule

Generality: generality indicates the relative size of the granule. A granule defined by the formula is more general if

it covers more instances of the universe. The quantity may be viewed as the probability of a randomly selected object satisfying the formula (Yao and Yao, 2002).

*b) Measurements on relationship between granules*

Confidence: Confidence or absolute support is defined as the fraction of instances that are correctly classified by the rule among the instances for which it makes any prediction. Thus, it is a measure of the correctness or the precision of the inference. As Equation (2) illustrates, The quantities can be computed by fraction of number of samples that satisfies the THEN part of the rule, to the number of samples that satisfy only the IF part of association rule (Yao and Yao, 2002). If the quantity of confidence of a rule is kept high then less number of association rules will be mined but their prediction accuracy will be quite high (Zhao et al., 2007).

$$\text{Confidence (class} = c_i|a = v) = p\,(\text{class} = c_i|a = v) \qquad (2)$$

Coverage: coverage is a measure of the applicability or recall of the inference. It indicates fraction of data in a class correctly classified by the rule (Yao and Yao, 2002). The quantities can be computed by fraction of number of samples that satisfy the THEN part of the rule, to the size of training data with the same class label as the rule consequent. Equation (3) indicates how coverage can be computed (Zhao et al., 2007):

$$\text{Coverage } (a = v|\text{class} = c_i) = p\,(a = v|\text{class} = c_i) \qquad (3)$$

In the granule network algorithm, a rule with confidence equal to one and high coverage nearest to one by considering any redundancy is selected as a decision rule. It is obvious that more information with high quality about a concept will be inferred using granules if the granule tree covers both a high absolute support and a high coverage.

Conditional entropy provides a measure that is inversely related to the strength of the inference. This measurement which depends on the confidence is a most commonly used measure for selecting attribute-value in the construction of decision tree for classification (Quinlan, 1982). If an object satisfies the formula of attribute-value, one can identify one equivalent class in which the object belongs with no uncertainty. In this case, confidence of the formula for at least one equivalent class is 1. Conditional entropy is defined by Equation (4) (Zhao et al., 2007):

$$\text{Entropy } (a = v) = -\sum_{c_i \in V_{class}} p(class = c_i\,|a = v)\log\,(class = c_i|a = v) \qquad (4)$$

## III. CASE STUDY

Induction of automatic classification rules for seismic vulnerability classification of Tehran using granular computing algorithm includes eight important steps which are investigated in this paper as follows:

1: Load a dataset for classification.

2: Set U as the root node of granule tree at initial stage.

3: Construct the family of basic concepts with respect to atomic formulas:

$$BC\,(U) = \{(a = v, m\,(a = v)) \mid a \in C,\ v \in V_a\}$$

4: Set the granule network to GN = ({U},∅), which is a graph consisting of only one node and no arc.

5: Set the activity status of U.

6: Select the BC = (a = v, m (a = v)) with maximum value of fitness with respect to U.

7: While the set of inactive nodes is not a non-redundant covering solution of the consistent classification problem, do:

7-1: Select the active node N with the maximum value of activity.

7-2: Select the basic concept BC = (a = v, m (a = v)) with maximum value of fitness with respect to N.

7-3: Modify the granule network GN by adding the granule N ∩ m (a = v) as a new node, connecting it to N by arc, and labeling it by a = v.

7-4: Set the activity status of the new node.

7-5: Update the activity status of N.

8: Export a granule tree and its corresponding classification rules.

The most important issue in constructing the granule decision tree is characterizing status of activity of a node at each level. A granule is non-active if it has two conditions: (i) the granule is a subset of a unique class and (ii) union of all granules at low level and non-redundant cover the solution of the root granule. An active granule will be further divided through efficient measures. After union of all inactive granules constructing a covering solution of universe set, construction of decision granule tree would be stopped.

The study area is located between 51 23' N, 51° 33' N Longitude and 34° 46' E, 35° 49' E Latitude. Five effective factors in assessment of seismic vulnerability of the areas including slope, percentages of weak buildings with 4 floors and less, percentage of more than 4 floor buildings, percentage of buildings built before 1966 and percentage of buildings built between 1966 and 1988 were considered. Data are obtained from Statistical Center of Iran. All spatial and non-spatial data were converted to ArcGIS database format. Because of less space of this paper, 7 rows out of 20 rows of training dataset of Seismic information table is selected and is illustrated in Table (1):

TABLE I. SEISMIC INFORMATION TABLE

| object | slope | Build less4 | Bef-66 | Bet 66-88 | Build more4 | class |
|---|---|---|---|---|---|---|
| U1 | Very high | high | low | high | low | 5 |
| U2 | Very high | moderate | low | High | low | 5 |
| U3 | low | moderate | moderate | moderate | low | 3 |
| U4 | Very high | low | low | Very high | low | 4 |
| U5 | low | low | low | high | low | 2 |
| U6 | low | low | low | moderate | low | 1 |
| U7 | Very high | high | low | Very high | low | 5 |

Column of the decision class are the grade of seismic vulnerability filled by some seismic experts. Five classes for discerning levels of seismic vulnerability between the groups of urban blocks are considered. These values range from very high vulnerable, high vulnerable, moderate vulnerable, low vulnerable and very low vulnerable. In order to ease the process, it is necessary to brief the label of the classes including very low vulnerable=1, low vulnerable=2, moderate vulnerable=3, high vulnerable=4 and very high vulnerable=5.

### *A. Decision Tree*

By applying the developed granule network algorithm and ID3 algorithm in ICS (Zhao et al., 2007) on training dataset, the granule decision tree and ID3 decision tree are constructed. Figures (1) and (2) illustrate the decision trees:

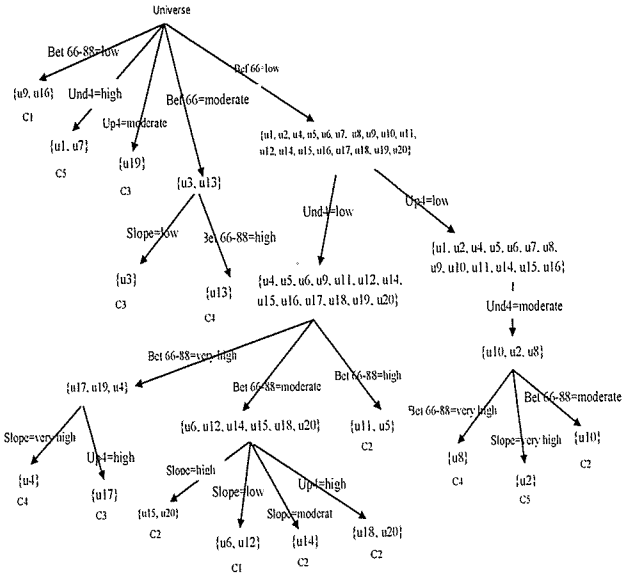


Figure 1. Seismic vulnerability granule network

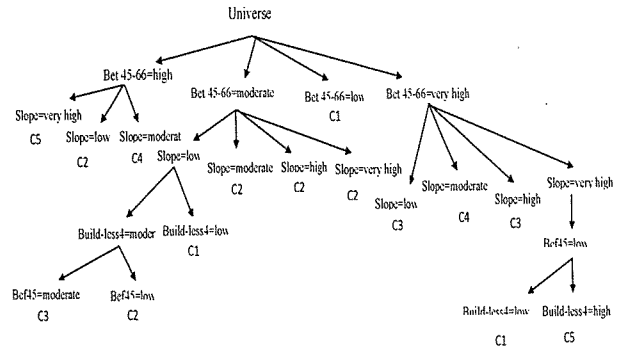


Figure 2. Seismic vulnerability ID3 decision tree

15 rules are extracted by the two approaches, which were implied to 50 test sites in Tehran, then the degrees of seismic vulnerability of the urban areas resulted from the tree were compared with an expert's judgment. The results verified that 48% of the urban areas are in complete agreement using the granular decision tree while 38% of the areas are consistent with that of ID3 decision.

## IV. CONCLUSION

This paper has proposed a new approach to induction of the classification rules for assessing the physical vulnerability of Tehran's urban areas against earthquake. In this work granular computing network was successfully employed for induction rules with high accuracy. The results indicates a covering based solution, searches larger spaces and is a global optimization in which at each step, only the most suitable granule defined by an attribute-value pair is selected. This leads to more accurate results compared to classical decision tree like ID3.

## REFERENCES

Alinia, S.H. and Delavar, M.R. (2010). Granular computing model for solving uncertain classification problems of seismic vulnerability in: *Spatial data quality from process to decision*, pp.132-133.

Mitchell, T.M.(1997). *Machine learning*. New York: McGraw-Hill.

Yao, Y.Y. (2001). On Modelling data mining with granular computing, *Proceedings of COMPSAC*, pp.638-643.

Yao, J.T. and Yao, Y.Y. (2002). Induction of classification rules by granular computing, proceedings of the third international conference on rough sets and current trends in computing, *Lecture notes in artificial intelligence*, 331-338.

Yao, Y.Y. (2000). Granular computing: basic issues and possible solutions, *Proceedings of the 5th joint conference on information sciences*, 186-189.

Pawlak, Z. (1991). *Rough sets: theoretical aspects of reasoning about data*, Dordrecht: Kluwer Academic Publishers.

Yuchun, T., Bo, J., and Zhang, Y.-Q.(2005). Granular support vector machines with association rules mining for protein homology prediction, Artificial Intelligence in Medicine. *Computational Intelligence Techniques in Bioinformatics*. 35(1-2), 121–134.

Zhao, Y., Yao, Y.Y. and Yan, M. (2007). ICS: an interactive classification system, *Proceedings of the 20th Canadian conference on artificial intelligence* (CAI'07), 134-145.

Quinlan, J. R.(1982). learning efficient classification procedures and their applications to chess end-games, in Michalski. R. S.. Carbonel, G. and Muchell. T.( (Eds). *Machine learning; An artificial intelligence approach* (pp. 391 to 411).

Shannon, C.E. (1948). The mathematical theory of communication. *The Bell System Technical Journal*. 27 (3/4), 373–423.