SIPINA Induction Graph Method for Seismic Risk Prediction

B. Selma

***Abstract*—**The aim of this study is to test the feasibility of SIPINA method to predict the harmfulness parameters controlling the seismic response. The approach developed takes into consideration both the focal depth and the peak ground acceleration. The parameter to determine is displacement. The data used for the learning of this method and analysis nonlinear seismic are described and applied to a class of models damaged to some typical structures of the existing urban infrastructure of Jassy, Romania. The results obtained indicate an influence of the focal depth and the peak ground acceleration on the displacement.

***Keywords*—**SIPINA method, seism, focal depth, peak ground acceleration, displacement.

I. Introduction

CURRENTLY, it is not possible to predict the earthquakes in the short and medium term with accuracy. It is possible, however, to assess the whole of the socio-economic consequences of an earthquake that can affect a city or territory. This is what has been done in California and Japan, where are developed seismic scenarios: It is to define the characteristics of a possible earthquake (from the historical seismicity and the regional tectonic context). Evaluation of seismic vulnerability and prediction of harmfulness parameters controlling the seismic response using advanced methodologies has become a crucial element in reducing seismic effects.

The Artificial Intelligence (AI) is one of the branches of computer science. It is based, among others, on theories that have been fundamentally developed for conventional systems in the physics and biology. The application of AI methods contributes to the improvement of existing conventional methods. We used in this study SIPINA (Interactive System for Interrogation Process in a non-tree form) which is a generation decision trees algorithm from data. The most widespread and used algorithms in data mining field are CART [1] (classification and regression trees), ID3 [2], C4.5 [3], and SIPINA [4]; the notoriety of these methods is such that it seems difficult to do an article on the trees and induction graphs without quoting at least one of these methods [5].

SIPINA algorithm has been applied to a range of problems of prediction and is proven to be remarkably successful in all areas. Some research studies have been achieved using the SIPINA method in the engineering field [6]-[8], their application showed that are powerful tools of simulation and optimization.

Our contribution within the framework of this application is that we present a method of representation and simplification of graphs generated by the SIPINA method. In addition, we generate simple conjunctive rules starting from the optimal induction graph.

II. Description of the SIPINA Algorithm

With the help of the sample Ω_a, we start the symbolic treatment for the construction of the induction graph (SIPINA method) [9]:

1. Set the measure of uncertainty;
2. Set parameters: λ, μ, and the initial partition S_0;
3. Apply the SIPINA algorithm for going from partition S_i to S_{i+1} and generate induction graph;
4. Generate prediction rules [10] [11].

The parameters λ, μ, the partitions and all other notions used in this process are introduced by means of examples and defined in the following paragraphs.

We chose SIPINA method [4], that we consider as the best among the methods based induction graphs; the effectiveness of SIPINA is superior to the classical methods such as ID3 and C4.5 [12]. The tree methods (ID3, C4.5 ...) often lead to undesirable situations because they consider certain distributions on classes as equivalent when they are not. These disadvantages result from both the process which proceeds exclusively by division and criteria that are insensitive to sample size. The SIPINA method attempts to reduce the disadvantages of tree methods on the one hand by the introduction of the merge operation and on the other hand by the use of a sensitive measure to the workforce.

A. Search for the Partition S_1

- The initial partition S_0 is formed of the sample of base.
- The λ parameter is set in an automatic way.
- Research of the variable X_j leading to the best partition S_1, or maximizing the gain on the uncertainty.

$$\Delta I\left(S_0, S_{X_j}\right) = i_\lambda(S_0) - i_\lambda(S_{X_j})$$

B. Basic Operations for the Passage of the Partition S_i to S_{i+1}

- Bursting: a segment t of S_i is divided using a predictor X in I segments $t_h = t \cap [X = h]$ where: $S_{i+1} = S_i - \{t\} + \{t_1\} + \cdots \{t_I\}$
- Fusion: It merges the two segments t_q and t_r of S_i where: $S_{i+1} = S_i - \{t_q\} - \{t_r\} + t_q \cup t_r$.

Boumediene Selma is with the Department of Civil Engineering, Abdelhamid Ibn Badis University of Mostaganem, Algeria (e-mail: selma.boumediene@yahoo.fr).

- Eligible partition: S_{i+1} is eligible if the gain on the Uncertainty; $\Delta I(S_i, S_{i+1}) = i_\lambda(S_i) - i_\lambda(S_{i+1}) > 0$

C. Passage of the Partition S_i to S_{i+1}

Fusion: It merges the two segments of S_i leading to a partition S'_{i+1} maximizing the gain on the uncertainty $\Delta I(S_i, S'_{i+1})$. If gain > 0, we pose S_{i+1}= S'_{i+1} and returns a step of fusion. Otherwise, passage to the next phase.

Fusion-splitting: It built all the partitions obtained by the fusion of two segments of S_i. For each of these partitions, it searches the Predictor leading to the best breakdown of two segments merged. It retains the partition to gain on maximum uncertainty. If this partition is eligible, it defines S_{i+1}. And it returns to step Fusion. Otherwise, proceed to the next phase.

Splitting: For each segment of S_i, it search the best partition eligible obtained by bursting using a predictor. It retains that which leads to the best gain on the uncertainty. If this best Eligible partition exists, it defines S_{i+1} and it sets out again in phase 1. Otherwise, the process stops and S_i is optimal.

III. Application

The previously described method was applied for the analysis of the maximum relative displacement of structures. The collected data are presented in Table I [13], [14].

TABLE I
EARTHQUAKE DATA USED FOR ANALYSIS

Earthquake	H, [km] (focal depth)	PGA, [%g] (peak ground acceleration)	Maximum relative displacement, δ_{max}, of the structure, [cm]
București 1977	109	0.1987	4.41
Focșani 1986	133	0.2934	3.08
București 1990	89	0.1008	1.80
București 1977	109		12.32
Focșani 1986	133	0.5551	5.83
București 1990	89		9.92
București 1977	109		20.79
Focșani 1986	133	0.9369	9.84
București 1990	89		16.72
București 1977	109		23.87
Focșani 1986	133	1.0729	11.27
București 1990	89		19.14

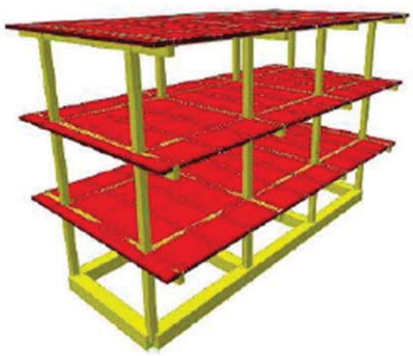


Fig. 1 Structural model used for analysis

The structural model, corresponding to a reinforced concrete structure is described in [14]. The time history analyses have been performed using SAP2000 software [15]. The results of the time history analyses for the types of loadings are listed in Table I in terms of maximum relative displacements.

As a case study, we considered the data for the structure in Table I. The data with the numerical attributes is given as follows in ARFF (Attribute-Relation File Format), a common format used to describe data for machine learning algorithms.

```
@RELATION earthquakes
@ATTRIBUTE h REAL
@ATTRIBUTE pga REAL
@ATTRIBUTE displacement REAL
@DATA
109, 0.1987, 4.41
133, 0.2934, 3.08
89, 0.1008, 1.80
109, 0.5551, 12.32
133, 0.5551, 5.83
89, 0.5551, 9.92
109, 0.9369, 20.79
133, 0.9369, 9.84
89, 0.9369, 16.72
109, 1.0729, 23.87
133, 1.0729, 11.27
89, 1.0729, 19.14
```

First, numerical data are discretized. Because some attributes can have a greater importance than others, we chose three intervals for the H attribute and five intervals for the remaining ones, in order to ensure a better precision. The pre-processed data are presented below:

```
@RELATION d-earthquakes
@ATTRIBUTE h {Low, Medium, High}
@ATTRIBUTE pga {VeryLow, Low, Medium, High, VeryHigh}
@ATTRIBUTE displacement {VeryLow, Low, Medium, High,
VeryHigh}
@DATA
Medium, Low, VeryLow
High, Low, VeryLow
Low, VeryLow, VeryLow
Medium, Medium, Medium
High, Medium, VeryLow
Low, Medium, Low
Medium, High, High
High, High, Low
Low, High, High
Medium, VeryHigh, VeryHigh
High, VeryHigh, Medium
Low, VeryHigh, High
```

IV. Results

We note that the decision graph is made up of succession of Fusion and splitting of summits.

Some fusions are not very interesting. This is the case when it is not followed by a splitting. That is simply to bring two predictions rules together leading to the same conclusion. We can eliminate them.

At the end of the symbolic treatment, we can generate the rules coming from the decision tree (graph). Let us consider the graph of Fig. 2 as if it was a final induction graph, without worrying about checking the details of all calculations that lead to this graph. At that point, we can deduce three prediction rules R1, R2 and R3 that have the form *if* condition *then* conclusion, where condition is a logical expression in disjunctive-conjunctive form and conclusion is the majority class in the node reached by the condition.

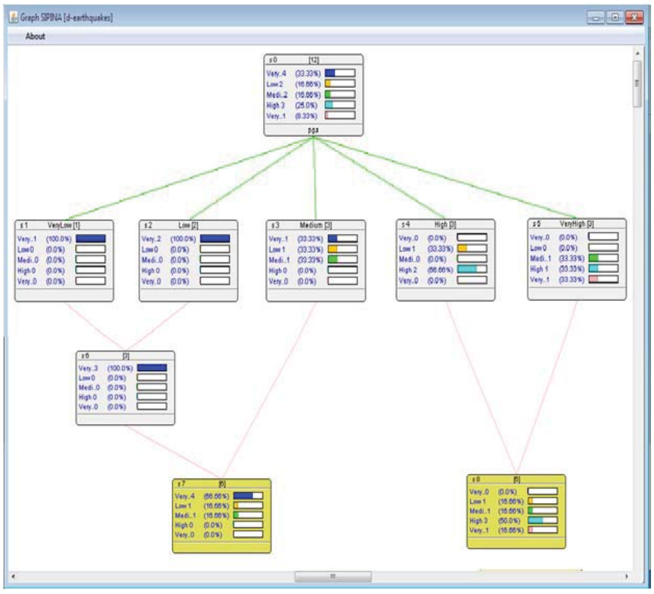


Fig. 2 Generated graph

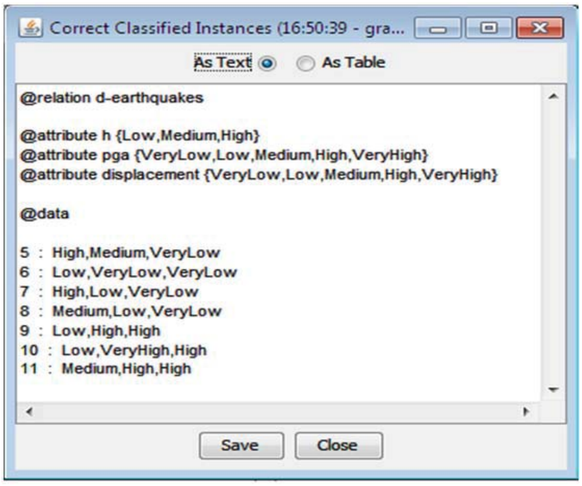
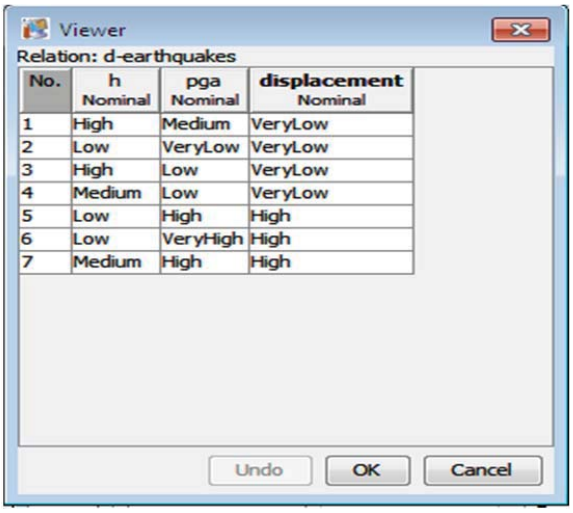


Fig. 3 Correct classified instances in text format



No.	h Nominal	pga Nominal	displacement Nominal
1	High	Medium	VeryLow
2	Low	VeryLow	VeryLow
3	High	Low	VeryLow
4	Medium	Low	VeryLow
5	Low	High	High
6	Low	VeryHigh	High
7	Medium	High	High

Fig. 4 Correct classified instances in table format

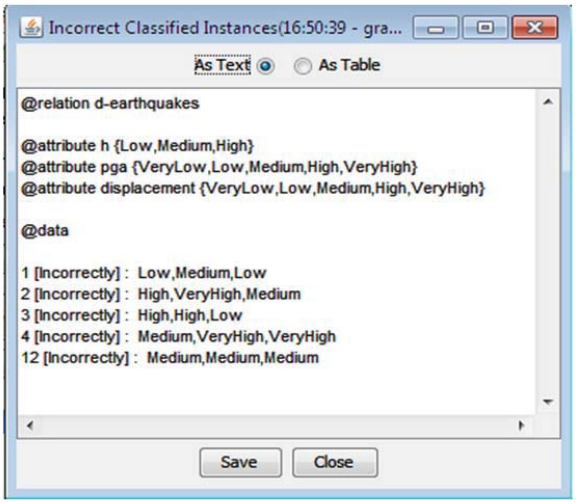
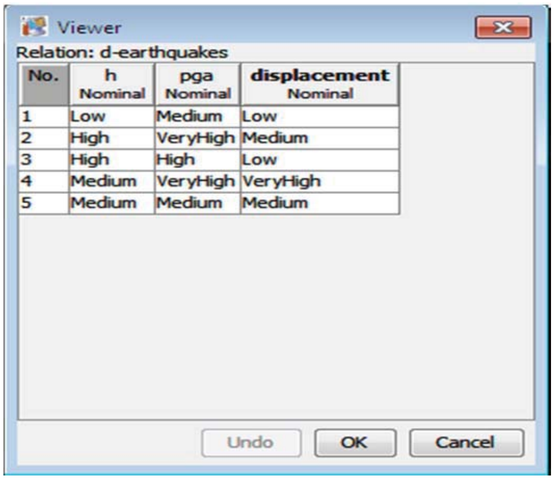


Fig. 5 Incorrect classified instances in text format



No.	h Nominal	pga Nominal	displacement Nominal
1	Low	Medium	Low
2	High	VeryHigh	Medium
3	High	High	Low
4	Medium	VeryHigh	VeryHigh
5	Medium	Medium	Medium

Fig. 6 Incorrect classified instances in table format

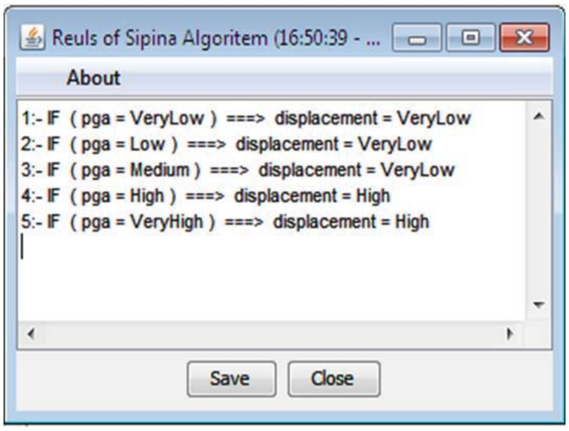


Fig. 7 Generated rules

V. Conclusion

In this article, we discussed the data mining in the field of civil engineering and in particular the earthquakes, providing models for the prediction of seismic risks by the Boolean modeling of seismic prediction rules. In the context of the analysis of earthquakes, the induction graph generated is a Boolean model that will allow us to see more closely the relations between the earthquake and structures exposed to the latter compared to the harmful parameters. The generated induction graph will facilitate the identification of seismic risks at the level of a given region in order to propose the best measures in support of precautions taken during the construction of structures.

References

[1] Breiman, L., Bagging Predictors. Technical Report No, 421, Department of statistics, University of California, 1994.
[2] Quinlan, J.R., Discovering rules by induction from large collections of examples, Expert Systems in the Microelectronic age, PP 168-201, 1979.
[3] Quinlan, J.R., Learning efficient classification procedures and their applications to chess endgames. In Machine Learning: An artificial Intelligence approach, Volume 1. Morgan Kaufmann, 1983.
[4] D.A. Zighed, J.P. Auray and G.Duru. sipina: Méthode et logiciel. Lacassagne,1992.
[5] Quinlan R., C4.5: Programs for Machine Learning. Morgan Kaufmann Publ., San Mateo, CA, 1993.
[6] M. Hammami, Y. Chahir, L. Chen, D. Zighed, "Détection des régions de couleur de peau dans l'image" revue RIA-ECA vol 17, Ed.Hermès, ISBN 2-7462-0631-5, Janvier 2003, pp.219-231.
[7] M. Hammami, "Modèle de peau et application à la classification d'images et au filtrage des sites web", Thesis , Ecole Centrale de Lyon, 2005.
[8] B. Atmani, B. Beldjilali, "Knowledge discovery in database: Induction graph and cellular automaton", Computing and informatics journal, Vol.26, 2007, pp.171-197.
[9] Zighed, D.A.: SIPINA for Windows, ver 2.5. Laboratory ERIC, University of Lyon, 1996.
[10] Rabaseda, S., Zighed, D.A.: Generation and Simplification of Rules in Graphs of Induction. Acts of the 25th Symposium of the Economic Structures, Econometrics and Data Processing, 1996, p. 7.
[11] Rakotomalala, R., Zighed, D.A.: Feschet, Characterization of Production Rules in a Process of Induction. Hermes Science Publication, Paris 1999.
[12] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, 1993.
[13] Atanasiu, G.M., Popa, B. F. (2005), Modeling of the seismic action for new concept of performance-based design, Proceedings of the 4th International Conference on Management of Technological Change, Chania, Greece, pp. 153-159.
[14] Popa, B. F. (2006), Elaborating Advanced Methodologies for Evaluating the Dynamic Performance of Reinforced Concrete Structures, PhD Thesis, Technical University of Iasi, Romania.
[15] SAP2000 vs. 9.1 (2005), Integrated Software for Structural Analysis and Design, Computer and Structures Inc., http://www.csiberkeley.com