Available online at www.sciencedirect.com
ScienceDirect
Procedia Environmental Sciences 26 (2015) 49 – 52
Procedia Environmental Sciences

Spatial Statistics 2015: Emerging Patterns

Predictive modelling of seismic hazard applying naïve Bayes and granular computing classifiers

Sheikhian, H.[a], M.R. Delavar[b] and A. Stein[3]

[a] MSc. Student, GIS Division, Dept. of Surveying and Geomatic Eng., College of Eng., University of Tehran, Tehran, sheikhain@ut.ac.ir
[b*] Center of Excellence in Geomatic Eng. in Disaster Management, Dept. of Surveying and Geomatic Eng., College of Eng., University of Tehran, Tehran, Iran (Corresponding author), mdelavar@ut.ac.ir
[c] Department of Earth Observation Science, University of Twente, The Netherlands, a.stein@utwente.nl

Abstract

Seismic damage mapping is an imperative part of urban risk assessment and reduction plans, especially in Tehran, the capital of Iran, where a large population lives in a potentially active seismic area. In this paper, a Bayesian statistical classification has been compared with the granular computing (GrC) algorithm for seismic physical vulnerability assessment. Both classifiers are verified by accuracy measurements. The results show that GrC had a better performance for seismic vulnerability assessment. In addition, the city of Tehran is judged to have a severe situation against possible earthquakes.

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee

Keywords Granular Computing; Bayesian Classification; seismic hazard assessment

1. Introduction

The city of Tehran is located within a seismic potential area which hosts a huge population living in almost old urban structure. Seismic damage mapping, as a multi-criteria decision making (MCDM) problem, involves various parameters and also expert judgments. This paper aims to produce seismic hazard maps by implementing Naïve Bayes (NB) and granular computing (GrC) classifiers.150 randomly selected samples of statistical urban units that are characterized by various seismic attributes ranked by experts against their degree of seismic physical

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns Committee
doi:10.1016/j.proenv.2015.05.022

vulnerability. Both classifiers are trained by the samples and used to classify the rest of the statistical units. Various accuracy measurements are applied and both models are verified to be used in seismic risk assessment. In addition, the produced vulnerability maps confirmed the severe situation of Tehran against possible earthquakes.

2. Naïve Bayes classifier

The Naïve Bayes (NB) classifier is a statistical classifier which implements Bayes' theorem with the assumption of strong independence between the input patterns attributes [1]. This classifier is fast, simple, accurate and robust to irrelevant attributes. It assigns prior probabilities to the input patterns and obtains the posterior probabilities by taking the evidence from training data. It assigns the class label with the highest probability to the input pattern [2]. Let C denote the class label for an instance of attribute-value observation X, and let c and x are particular values for C and X. Then the possibility of an instance x to have the class label c equals:

$$p(C=c \mid X=x) = p(C=c)\,p(X=x \mid C=c)\,. \tag{1}$$

Estimates for $p(C=c)$ and $p(X=x \mid C=c)$ are obtained from a training data set. If x is an unobserved instance, then (2) is used to obtain the class probability, based on the assumption of strong independence between the attributes:

$$p(C=c \mid X=x) = p(C=c)\prod p(X_i = x_i \mid C=c) \tag{2}$$

where X_i and x_i are values for attributes of a given instance and an instance of the training data set.

3. Granular Computing classifier

GrC algorithm operates on the granules of information when extracting classification rules. GrC groups input patterns based on attribute similarity and applies measurements on those granules to obtain the best set of rules [3, 4]. These measures include generality, absolute support, coverage and conditional entropy which are applied on the IF-THEN rules of the form of $\phi \rightarrow \psi$. Generality of a rule is the portion of instances satisfying rule condition [5]. Absolute support defines the conditional probability of a randomly selected object to contribute in ϕ if it satisfies ψ. The coverage of concept ϕ provided by concept ψ is the conditional probability that an object satisfying ϕ, also satisfies ψ. Entropy is the consistency of a specific formula ϕ based on formulas ψ. Table 1 explains the formulas for the mentioned parameters.

4. Experimental Results and conclusions

The city of Tehran has a population of more than 12 million people and a large amount of weakly constructed buildings. Based on the seismic characteristics and expert judgment, a data set containing 150 randomly selected samples of urban statistical units was formed. The selected seismic attributes include age of buildings built before 1966 when the first fortification regulation was established, and during 1966-1988 when a seismic resistance code has been implemented, the number of floors in the buildings, the ground slope and the earthquake intensity. Five degrees of vulnerability were defined by experts as class labels including very low, low, medium, high and very high seismic vulnerabilities. A correlation test confirmed the independence of the selected attributes, where all of the attribute pairs showed a correlation value below 0.5. Figure 1 displays the assigned vulnerability of the samples on the study area.

Table 1: Granular computing parameters

Parameter	Formula	Parameters
Generality	$G(\phi) = \frac{\lvert m(\phi)\rvert}{\lvert U\rvert}$	$\lvert m(\phi)\rvert$:Number of objects contribute in concept ϕ $\lvert U\rvert$:Size of data set
Absolute Support	$AS(\phi \to \psi) = \frac{\lvert m(\phi \wedge \psi)\rvert}{\lvert m(\phi)\rvert}$	$\lvert m(\phi \wedge \psi)\rvert$: Number of objects constructing both granules of concepts ϕ and ψ
Coverage	$CV(\phi \to \psi) = \frac{\lvert m(\phi \wedge \psi)\rvert}{\lvert m(\phi)\rvert}$	
Entropy	$H(\psi \mid \phi) = \sum_{i=1}^{n} p(\psi_i \mid \phi) \log(p(\psi_i \mid \phi))$	$p(\psi_i \mid \phi) = \frac{\lvert m(\phi \wedge \psi_i)\rvert}{\lvert m(\phi)\rvert}$

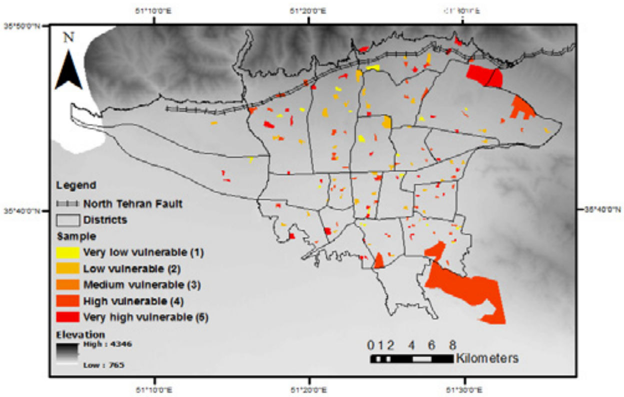


Figure 1: Position of the samples in the study area with different colors indicating earthquake vulnerability.

The two classifiers were trained with 70% of the data set, whereas the remainder of the data was used for testing the models. Accuracy measurements of the two classifiers are expressed in Table 2 and Figure 2. R^2 and $RMSE$ values show that both models provide acceptable results. Moreover, the acquired *p*-values show the similarity of the results to their desired values given by the experts. According to Figure 2, the difference of GrC and NB in their vulnerability class calculated was mostly one or zero, except in some cases. The trained classifiers were applied to the study area and the vulnerability maps are achieved (Figure 3). A correlation coefficient of 0.88 showed a significant similarity between the GrC and NB outputs.

Table 2: Accuracy assessment

Algorithm	NB			GrC		
measurements	R2(coefficient of determination)	RMSE	p-value	R2	RMSE	p-value
Train	0.84	0.49	0.0002	0.87	0.45	0.0000
Test	0.82	0.58	0.0002	0.83	0.49	0.0001

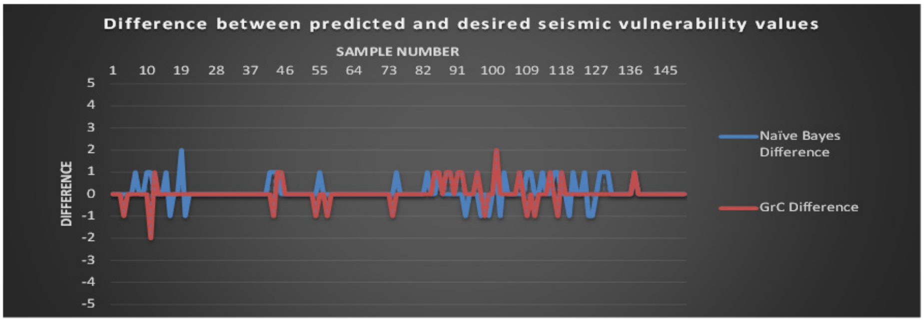
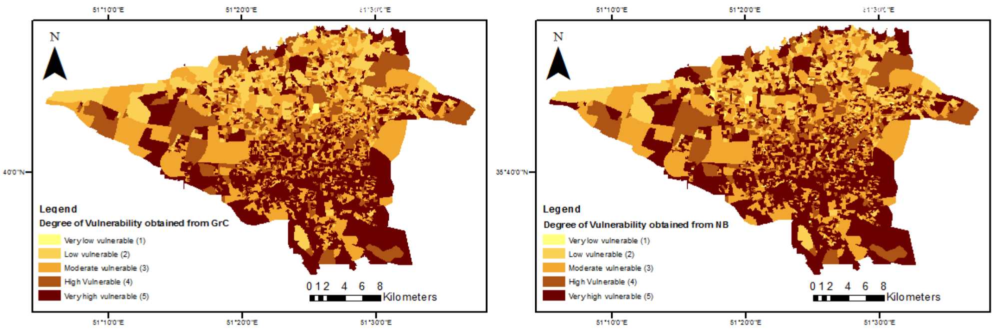


Figure 2: Distribution of errors



a b

Figure 3: Seismic physical vulnerability maps: (a) GrC, (b) NB

The experiment verified that both classifiers provided acceptable accuracy in determining seismic vulnerability of Tehran. Although a significant similarity between GrC and NB results was observed, GrC performed better than NB according to the accuracy measurements presented in Table 1.The classification showed a severe seismic risk in Tehran, as was confirmed with the need for effective disaster reduction plans.

References

1. Lewis, D.D., Naive (Bayes) at forty: The independence assumption in information retrieval,*Machine learning: ECML-98*. 1998, Springer. p. 4-15.
2. Peng, F., D. Schuurmans, and S. Wang, Augmenting naive bayes classifiers with statistical language models.*Information Retrieval*, 2004. **7**(3-4): p. 317-345.
3. Pedrycz, W., S. Bassis, and D. Malchiodi, *The Puzzle of Granular Computing*. Vol. 138. 2008: Springer.
4. Zadeh, L.A., Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic.*Fuzzy sets and systems*, 1997. **90**(2): p. 111-127.
5. Yao, Y., A Unified Framework of Granular Computing, *Handbook of Granular Computing*, W. Pedrycz, A. Skowron, and V. Kreinovich, Editors. 2008, John Wiley & Sons. p. 401-410.