

STIMA DELL’AFFIDABILITÀ DEL DATO MACROSISMICO MEDIANTE TECNICHE DI ROUGH SET ANALYSIS

F.T. Gizzi¹, B. Murgante², M.R. Potenza¹, M. Danese², L. Tilio², C. Zotta¹

1 Istituto per i Beni Archeologici e Monumentali (IBAM-CNR)-Tito Scalo (Pz)

2 Laboratorio di Ingegneria dei Sistemi Urbani e Territoriali, Università della Basilicata (Potenza)

Introduzione. È noto come l’informazione macrosismica possa essere utile per selezionare aree potenzialmente soggette a fenomeni di incremento dell’azione sismica. Tuttavia, per non incorrere in analisi errate è necessario avere a disposizione un set di dati sul danneggiamento che sia affidabile. Obiettivo del contributo è, quindi, proporre un criterio per la valutazione, attraverso l’approccio della *Rough Set Theory*, dell’attendibilità dell’informazione macrosismica. La località prescelta come test metodologico è Rapolla (PZ), che subì effetti di danneggiamento pari a VIII MCS a causa del sisma irpino del 23 luglio 1930. Per questa località sono stati analizzati, con il metodo proposto, gli effetti di danneggiamento desunti dalla consultazione delle perizie di accertamento danno.

Metodologia. La *Rough Set Analysis* (RSA) è una metodologia statistica non parametrica (Cohen & Nijkamp, 2002) che consente di identificare nei dati dei pattern ed estrarre delle relazioni, in particolare, individuando legami di causa ed effetto. I pattern individuati generalmente vengono rappresentati attraverso un set di regole decisionali, della forma “if...then”, capaci di spiegare i fenomeni e/o i processi legati ai dati ed interpretabili appunto come una descrizione di una relazione di tipo causa-effetto (Greco et al., 2001). Uno degli aspetti importanti di questo approccio è la possibilità di lavorare su conoscenze imprecise, incomplete ed incerte, ottenendo un raggruppamento dei dati in funzione delle “informazioni simili” che li caratterizzano (Komorowski, 1999).

Il concetto fondamentale alla base della RSA è la relazione di indiscernibilità, generata dall’informazione disponibile. L’esempio (Pawlak, 1998) più utilizzato in letteratura per spiegare facilmente il concetto, preso in prestito dalla medicina, fa riferimento a diversi pazienti che, soggetti agli stessi sintomi, vengono accomunati dalla stessa diagnosi: i pazienti non sono distinguibili dal punto di vista dell’informazione disponibile, rappresentata soltanto dagli stessi sintomi; essi saranno dei *cluster* di oggetti indistinguibili.

Formalmente, si definisce sistema informativo un sistema composto da un insieme di oggetti finito e non vuoto, chiamato universo, ed indicato come U , ed un insieme di attributi, finito e non vuoto, indicato come A , $IS = \{U, A\}$, dove $U = \{x_1, \dots, x_m\}$; $\forall a \in A, \exists f_a: U \rightarrow V_a$ (dove V_a è il dominio dell’attributo a). (Walczak & Massart, 1999).

Ogni $B \subset A$ determina una relazione binaria $Ind(B)$ su U , chiamata relazione di indiscernibilità, e definita come segue: x_i ed x_j , $x_i, x_j \in U$, sono indistinguibili se $b(x_i) = b(x_j)$, $\forall b \in B$; $Ind(B)$ è una relazione di equivalenza, ed ogni elemento di U che appartiene alla $Ind(B)$ può essere indicato come $[x_i]_{Ind(B)}$. La classe di equivalenza di $Ind(B)$ è chiamata elementary set in B , in quanto rappresenta il più piccolo gruppo di oggetti distinguibili.

La relazione di indiscernibilità è particolarmente utile per individuare le relazioni di dipendenza tra gli attributi: un attributo dipendente da un altro non arricchisce il sistema informativo, e la sua eliminazione dunque non inficia la conoscenza su esso, è superfluo: se $Ind(A) = Ind(A - a_i)$, dove a_i è l’attributo di cui si vuole verificare l’indipendenza, allora a_i può essere eliminato, ed il sistema informativo semplificato. I *reducts* sono dei sottoinsiemi che rappresentano la parte essenziale di un sistema informativo, capaci di distinguere tutti gli oggetti distinguibili. La loro intersezione si chiama *core*, sottoinsieme dunque costituito da tutti gli attributi indispensabili.

Se si considera un sottoinsieme $X \subset U$, e si vuol conoscere quali elementi di U sono contenuti in X e quali non sono contenuti, ci si confronta con una condizione di incertezza, per cui si individueranno elementi che senza alcun dubbio appartengono ad X , ed elementi che potrebbero appartenervi o meno. La Fig. 1 schematizza la situazione:

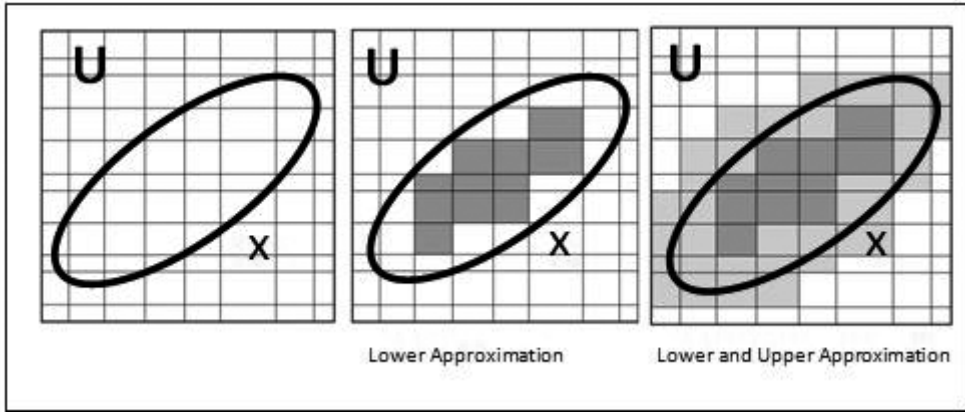


Fig. 1 - Schematizzazione della condizione di incertezza: lower e upper approximation.

La condizione di appartenenza ad un insieme $X \subset U$ con certezza o meno consente di definire due insiemi, rispettivamente la *lower approximation* e la *upper approximation*. Formalmente, i due insiemi vengono definiti come segue:

$$\underline{BX} = \{x_i \in U: [x_i]_{Ind(B)} \subset X\}; \quad \overline{BX} = \{x_i \in U: [x_i]_{Ind(B)} \cap X \neq \emptyset\}.$$

loro differenza costituisce la *Boundary Region*, la cui esistenza rivela che l'insieme X non è definito, è dunque un ROUGH SET.

L'accuratezza dell'approssimazione (Pawlak, 1999) di X è definita dal rapporto tra la cardinalità della *lower approximation* e la cardinalità dell'*upper approximation*:

$$\mu_B(X) = \frac{card(\underline{BX})}{card(\overline{BX})}$$

Un sistema informativo in cui, oltre all'insieme A degli attributi è presente l'insieme D degli attributi cosiddetti "di decisione" si chiama tavola o sistema di decisione (Pawlak, 1997), e si indica come $DS = \{U, A, D\}$. Gli attributi di decisione sono attributi in grado di descrivere l'appartenenza degli oggetti ad una certa classe, attraverso la formulazione di un set di regole di decisione del tipo:

$$a_{k_i} \Rightarrow d_j$$

Quando un sistema informativo contiene sia condizioni sia decisioni, una tavola informativa ridotta, sulla quale è stata calcolata la relazione di indiscernibilità per tutti gli attributi (i.e. è stata costruita una matrice di indiscernibilità) e sono stati definiti i *reducts* ed il *core*, può essere utilizzata per generare delle regole di decisione. Il numero di oggetti che soddisfano le condizioni della regola individuata determina la forza di tale regola, e dunque una misura dell'incertezza nell'assegnazione degli oggetti ad una certa classe di decisione.

Caso di studio. Sono stati analizzati i dati raccolti in riferimento al processo di valutazione danno e ricostruzione che è seguito al terremoto del 1930 nell'area del Vulture. Il ricco database contiene, per ogni edificio, informazioni di carattere amministrativo (proprietario, documenti di riferimento, etc), sulla tipologia dell'edificio (pubblico, religioso etc), sui costi e durata dei lavori, sulla stima del valore degli edifici, sui redditi del proprietario, sui lavori di ricostruzione (variazioni nelle strutture verticali e orizzontali, elementi demoliti, tipologia di interventi adottata etc.), oltre alle informazioni relative al danno, etc. I dati raccolti sono relativi a 728 edifici, tuttavia la tavola informativa usata nell'analisi è stata ridotta, e sono stati considerati soltanto 367 edifici, a causa di lacu-

ne conoscitive. Le analisi sono state condotte anche sull'intero set, provando a far compilare al software di supporto, ROSE, i dati mancanti, ottenendo però un risultato altamente insoddisfacente. Allo stesso modo, è stato deciso di non trattare tutti gli attributi, ma trascurarne alcuni per i quali risultavano molti valori mancanti. Il software ROSE processa un file di tipo ".isf", preparato a partire dal database geografico a disposizione, ed in cui viene indicata la definizione degli attributi di decisione: la variabile decisionale utilizzata è quella relativa alla classe di danno secondo la European Macroseismic Scale (EMS-98), ritenuta fondamentale per la comprensione delle dinamiche legate alla ricostruzione. Il software opera delle analisi e classifica gli oggetti, che nel caso specifico sono rappresentati dagli edifici, e fornisce in output diversi files, relativi alla classificazione, alla induzione delle regole, alla matrice di discernibilità, etc.

La classificazione operata mostra una qualità pari allo 0,7057, dove per qualità della classificazione si intende il rapporto tra la cardinalità della *lower approximation* e la cardinalità dell'intero insieme. L'applicazione della relazione di indiscernibilità ha consentito di individuare 265 insiemi elementari. Per ogni classe di danno, poi, è stata calcolata l'accuratezza, che aiuta a capire la "forza" della classificazione secondo la classe di danno. In particolare, si è osservato che tutte le classi di danno risultano essere dei *rough set*, approssimativamente definiti; in alcuni casi, inoltre, l'accuratezza è molto bassa, sintomo di una alta indefinizione. In un caso, invece, che coincide con la classe di danno d_4 , l'accuratezza è molto alta. Il sistema informativo analizzato risulta caratterizzato dalla presenza di due *reducts*, costituiti da 9 attributi ciascuno, la cui intersezione, il *core*, individua 8 attributi (dai circa 30 di partenza).

Discussione dei risultati. Le regole decisionali inferite sono 97, in particolare 89 regole esatte e 8 regole approssimate. La differenza tra le due tipologie consiste nella forma delle conseguenze di ogni affermazione: nelle regole esatte, si tratta di una univoca conseguenza, supportata soltanto da oggetti appartenenti alla *lower approximation* della classe decisionale corrispondente; nelle regole approssimate, invece, si individuano conseguenze non univoche e supportate da oggetti appartenenti anche alla *boundary region*.

L'interpretazione delle regole mostra che un certo numero (25/88) presenta una chiara inconsistenza nell'attribuzione del livello di danno: si deduce dunque che le dinamiche locali probabilmente hanno influenzato l'attribuzione del danno; tuttavia si rimanda ad un contesto più largo per la verifica e conferma di questa ipotesi, considerando che con un maggior numero di dati (e magari maggiore completezza degli stessi), il fenomeno può diventare più evidente, e possono essere più chiare le dinamiche che lo hanno determinato.

Una possibile interpretazione dell'analisi è quella che le differenze nella distribuzione del danno non siano spazialmente clusterizzate, ma riguardino aree con caratteristiche diverse sia dal punto di vista delle proprietà intrinseche degli edifici sia sotto l'aspetto socio-economico (dimensioni degli edifici, assenza o meno di interventi di manutenzione nel tempo, proprietari benestanti o proprietari in condizioni economiche disagiate, etc). Apparentemente, non sembra evidente una manipolazione volontaria della classificazione del danno (ad esempio attraverso favoritismi richiesti ai tecnici da parte dei proprietari), piuttosto si possono considerare diversi fattori-causa:

1. un rilievo delle condizioni di danneggiamento degli edifici piuttosto superficiale (cosicché l'attribuzione della classe di danno è stata effettuata attraverso un'indagine dall'esterno o limitata soltanto ad alcuni vani);
2. la disomogeneità nelle competenze professionali dei tecnici rilevatori;
3. le modalità di descrizione del danno, che nelle prime fasi post-sisma includeva anche miglioramenti e/o estensione dei lavori non correlati all'evento sismico;
4. l'incompletezza dei dati descrittivi riportati.

Sviluppi futuri. Nel prosieguo della ricerca, l'area di studio sarà ampliata, estendendo l'analisi ai comuni limitrofi soggetti allo stesso terremoto, in modo da portare in conto anche altre variabili riferite all'assetto geologico locale, ed inoltre saranno comparati i risultati derivanti dalla RSA con quelli di altri metodi, usando Visual Analytics.

Bibliografia

- Cohen G., Nijkamp P.;2002: Information and communication technology policy in European cities: a comparative approach, in *Environment and Planning B: Planning and Design*,29, 729-755.
- Greco S., Matarazzo B., Slowinski R.;2001: Rough set theory for multicriteria decision analysis, *European Journal of Operational Research*, 129, 1-47.
- Komorowski J., Pawlak Z., Polkowski L., Skowron A.;1999: Rough Sets: A Tutorial, In: *Rough Fuzzy Hybridization – A New Trend in Decision Making*, Springer, S.K. Pal, A. Skowron, pp. 3-98.
- Pawlak Z.;1997: Rough set approach to knowledge-based decision support. *European Journal of Operational Research*, 99, 48-57.
- Pawlak Z.; 1998: Rough set theory and its applications to data analysis. *Cybernetics and Systems*, 29(7), 661-688.
- Pawlak Z.; 1999: Rough classification. *Int. J. Human-Computer Studies*, 51, 369-383.
- Walczak B. Massart D.L.;1999: Tutorial, Rough sets theory, *Chemometrics and intelligent laboratory systems*, 47, 1-16.