

Roma
13 luglio 2007
Prot. SC/LL/003618

Agli Associati

Loro sedi

Sicurezza anticrimine (pos. 227/204-272) - Strumento di supporto per l'analisi del rischio rapina

In materia di sicurezza anticrimine uno dei principali problemi per il settore bancario è rappresentato dall'evento rapina, sia per la numerosità dei casi, sia per le possibili conseguenze sulle persone (clientela e dipendenti).

Risulta, quindi, evidente la necessità di misurare tale rischio in un'ottica di prevenzione, individuando i fattori che lo producono, per rendere più efficienti gli investimenti nella sicurezza e più efficaci le soluzioni adottate, anche allo scopo di definire parametri omogenei, in conformità dei quali procedere alla valutazione del "rischio rapina" ai sensi del d. lgs. n.626/1994.

Per raggiungere quest'obiettivo, l'ABI ha realizzato – con l'ausilio di uno specifico Gruppo di lavoro composto da esperti di security e safety – lo "Strumento di supporto per la valutazione del rischio-rapina", che ha carattere sperimentale per tutto il 2007.

Lo strumento, che si interfaccia con il data-base di OS.SI.F.:

- archivia in maniera strutturata i dati storici degli sportelli bancari relativamente ai fattori legati alla sicurezza;
- stima dinamicamente il "peso" da associare ai sopra citati fattori, tramite una RNA (Rete Neuronale Artificiale);
- calcola degli indici di rischio (*esogeno, endogeno e combinato*) opportunamente normalizzati, basandosi sui pesi attribuiti ai vari fattori di sicurezza.

Lo strumento, pertanto, in maniera dinamica - per ognuno dei comuni d'Italia ove sono presenti sportelli bancari – calcola il **c.d. rischio esogeno**, determinato sulla base dei seguenti elementi, riferiti agli ultimi anni:

- intensità criminosa su base comunale (fonte Ministero Interno);
- densità sportelli su base comunale (fonte Banca d'Italia);
- densità abitanti su base comunale (fonte ISTAT).

Il rischio esogeno esprime, relativamente alle rapine in banca, la potenziale pericolosità del comune in cui è ubicato lo sportello.

Inoltre, lo strumento valuta le singole misure di sicurezza adottate presso ogni sportello, individuando il **c.d. rischio endogeno**, sulla base dei seguenti parametri:

- peculiarità;
- sistemi di difesa antirapina;
- servizi antirapina;
- rapine subite.

Lo strumento, combinando il rischio endogeno con l'esogeno, determina per ogni singolo sportello il **c.d. rischio combinato**.

Lo strumento consente anche di verificare la conformità delle misure di sicurezza di ogni sportello ai requisiti previsti nel "Protocollo d'intesa ABI-Prefetture per la prevenzione della criminalità".

Inoltre, è stato realizzato un simulatore che permette attraverso l'introduzione fittizia di dati di verificare l'indice endogeno virtuale dello sportello analizzato.

Ai fini di utilizzare correttamente lo strumento e comprenderne i principi fondanti sono stati realizzati uno specifico manuale d'uso (allegato 1) e una scheda tecnica (allegato 2) che illustra, tra l'altro, l'architettura della RNA, la struttura delle variabili considerate e i criteri che hanno orientato la definizione degli algoritmi a supporto dello strumento medesimo.

L'occasione è gradita per porgere i migliori saluti.

Giuseppe Zadra
Direttore Generale



Codice Attività SC6030-LL1190

ALLEGATI

SCHEDA TECNICA DI RIFERIMENTO PER LO STRUMENTO DI ANALISI DEL RISCHIO RAPINA

1. Algoritmi Neuronali vs. Metodi Statistici

Gli algoritmi neuronali nascono nell'ambito della simulazione dei sistemi nervosi biologici come prova oggettiva delle conoscenze che abbiamo della fisiologia di questi, ma sono stati applicati, ampiamente e velocemente, in svariati campi per una infinità di problematiche tecniche e scientifiche, compresa quella dell'analisi statistica dei fenomeni naturali e sociali. In realtà, sono già - anche- considerati "metodi statistici".

Le prestazioni di tali algoritmi, nell'ambito dei "processi previsionali", sono risultate simili a quelle degli algoritmi tradizionali ed, in altri ambiti, persino ("sorprendentemente") migliori - forse perché più intuitivi ed analoghi al nostro modo di percepire la realtà.

La riluttanza (qualche volta vera avversione) di certi matematici e studiosi di statistica nei confronti degli algoritmi neuronali nasce, forse, dal fatto che non tutti questi metodi sono accompagnati da una rigorosa dimostrazione matematica del loro funzionamento e, quindi, della validità dei risultati ottenibili. A parte la back propagation, per la quale solo ultimamente (4-5 anni fa) è stata formulata una dimostrazione rigorosa, gli altri algoritmi neuronali rimangono privi di una spiegazione formale ed esaustiva, nonostante la constatazione dei buoni risultati che essi sono stati in grado di fornire.

In merito alle prestazioni ottenibili, il confronto tra i diversi algoritmi neuronali va imperniato non tanto sulla qualità assoluta di un algoritmo rispetto ad un altro bensì sull'opportunità di impiegare un algoritmo piuttosto che un altro nella particolare problematica presa in considerazione.

Ad esempio, per quanto riguarda il calcolo della media vale il principio (**Chisini**):

"la media è quel numero X tale che, date n grandezze x_i $i=1,2,...,n$ e considerata una loro funzione $f=f(x_1, x_2,... x_n)$ (somma, prodotto ...), X lascia invariato il valore della funzione stessa, vale a dire $f(x_1, x_2,... x_n)=f(X, X,... X)$ "

ed è, quindi, corretto affermare che, ad esempio, la media aritmetica non è più scadente o meno precisa di quella geometrica e che la scelta della media giusta da applicare non dipende dalle qualità generali possedute da ogni singola media, ma dal relativo contesto ovvero dalla particolare funzione che si sta considerando.

"L'apprendimento di caratteristiche invarianti degli stimoli di ingresso, la resistenza alle variazioni casuali e la capacità di generalizzare rendono le reti neurali degli strumenti potenzialmente molto efficaci per le predizioni finanziarie, la valutazione dei rischi in investimenti e prestiti e le analisi del mercato borsistico. Benché sia ben noto che un gran numero di compagnie specializzate in analisi finanziarie fa uso di reti neurali, vi sono relativamente poche pubblicazioni scientifiche a questo riguardo. E' stato comunque mostrato che alcuni modelli neurali ottengono prestazioni migliori delle tecniche tradizionali di regressione specialmente quando il modello viene applicato a nuovi dati. Altri hanno invece evidenziato come le reti neurali ottengano lo stesso

grado di prestazioni di sistemi a regole già esistenti. Comunque sia, la facilità di utilizzo delle reti le rende molto più accessibili di sistemi specialistici molto costosi e complessi” (Floreano – Mattiussi):

Un altro aspetto importante che ha determinato la scelta degli algoritmi neuronali è legato agli obiettivi del progetto, ovvero, stimare la probabilità dell'evento rapina e valutare i pesi di ciascun input (in valore percentuale e assoluto), riunendo il tutto in un'architettura logica a priori con elementi puramente astratti. Così, oltre l'uscita, che rappresenta la probabilità cercata, si sono valutati anche particolari valori interni, detti “pesi”, che sono proporzionali all'importanza di ciascun ingresso.

2. Qualità delle reti neurali

Logica di base semplice ed intuitiva. L'algoritmo del tipo **back propagation**, che risulta essere l'algoritmo di riferimento per il nostro **babapro**, durante la cosiddetta fase d'apprendimento, svolge ripetutamente le seguenti operazioni:

- fornisce una risposta in uscita;
- si corregge internamente (modificando particolari numeri chiamati **pesi**) in funzione dell'errore prodotto dalla risposta fornita.

Il suo comportamento è facilmente immaginabile con l'analogia dell'allievo, la sua capacità di adattamento l'intelligenza, dovuta alla disponibilità di pesi e connessioni interne, lo sbandamento o l'oscillare intorno ad alcuni valori senza una stabile risposta è simile a quello che non capisce per stupidità o perché l'insegnante ha scarse qualità didattiche (alias: i dati non sono uniformi, accurati, corretti...). Di conseguenza, è facile interpretarne il comportamento e, quindi, modificarne l'architettura con grandi vantaggi sulla flessibilità (qui inteso come evoluzione dell'algoritmo stesso).

Ad esempio, nel nostro caso la rete è usata simultaneamente per la predizione dell'indice di rischio e per la valutazione del peso relativo dei vari fattori d'ingresso; inoltre, in **babapro**, con l'incompleta connessione degli strati, l'influenza relativa è strettamente proporzionale alle connessioni imposte e ciò permette la valutazione dei centri di flusso del calcolo, che sono vere e proprie astrazioni, con la certezza della consistenza del peso attribuito. Questa multifunzionalità e facilità d'uso è rara negli algoritmi tradizionali, in genere, piuttosto complessi e pesanti nell'implementazione al calcolatore.

Capacità di isolare i pattern sbagliati. Mentre solitamente un record errato produce valutazioni errate, nelle reti neurali, invece, esso è intenzionalmente inserito per migliorare l'apprendimento della rete. Questa tecnica è usata per “smuovere” la rete da una posizione di minimo relativo e spingerla verso altre direzioni più proficue, con il vantaggio di poter utilizzare dati dal mondo reale anche piuttosto “sporchi”.

Robustezza nella risposta anche con pattern corrotti. Come nel caso precedente, se il record d'ingresso è parzialmente mancante o errato, la rete può continuare a dare risposte coerenti.

Robustezza anche se parzialmente danneggiate. A differenza della maggior parte di algoritmi, se parte delle connessioni sono eliminate, la rete è ancora in grado di dare risposte congrue (alla stessa maniera del cervello umano).

Flessibilità. Un modello neurale può essere impiegato per un grande numero di finalità diverse:

non ha bisogno di conoscere le proprietà del dominio specifico di applicazione perché le apprende in base all'esperienza.

Generalizzazione. Addestrata da un limitato numero di esempi, è in grado di produrre risposte adeguate a nuovi esempi non presentati nella fase di training. In altre parole, è in grado di “capire” delle problematiche particolari e quindi estrarre le logiche interne per riprodurle nel vario mondo reale...

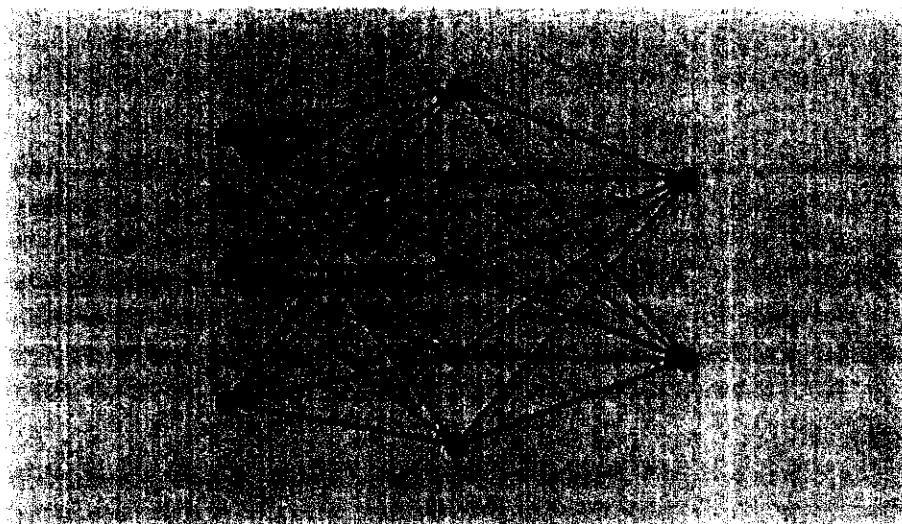
Leggerezza di calcolo. Non considerando il tempo e le risorse di calcolo della fase d'addestramento, una chiamata (recall) alla rete comporta poche e semplici operazioni, programmabili in un telefonino o in una calcolatrice tascabile di vecchia generazione. La nostra rete intera è gestita dal programma di base con sole quattro semplici funzioni.

tipo **back propagation**; durante la cosiddetta fase di apprendimento, tale rete svolge ripetutamente le seguenti operazioni:

- fornire una risposta in uscita;
- correggersi internamente (modificando particolari numeri chiamati **pesi**) in funzione dell'errore prodotto dalla risposta fornita.

3. Architettura di Ba.Ba.Pro

BaBaPro (acronimo di BAYesian BACk PROpagation) è stato pensato perché potesse contemporaneamente calcolare la probabilità di un evento sulla base di una disposizione teorica progettata arbitrariamente e il peso relativo di ogni neurone sul risultato finale.



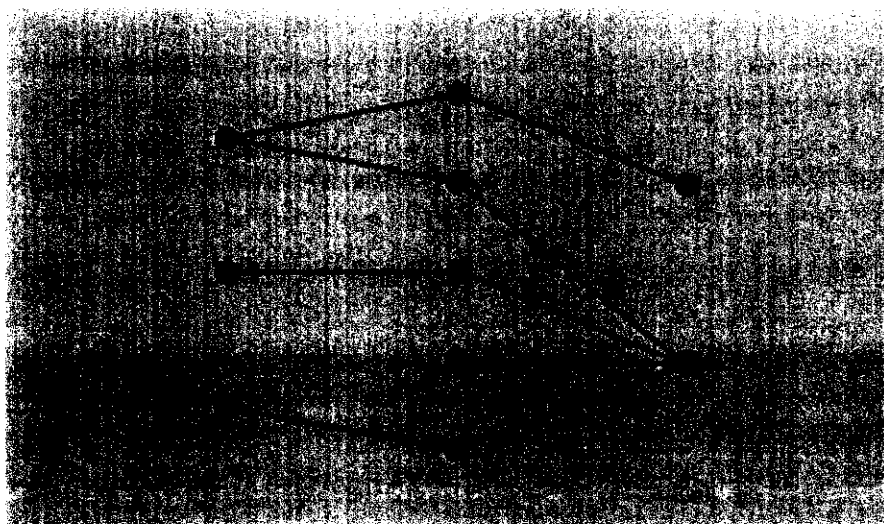
La figura mostra una back propagation interamente connessa. Il calcolo procede da destra a sinistra secondo la direzione delle frecce; i tre neuroni (tondi rossi) posti alla sinistra rappresentano gli input e quelli alla destra le uscite (in questo caso 2), lo strato interno è costituito da 5 elementi.

Esclusi gli input, che vale il valore immesso, per ciascun neurone viene calcolato lo stato di attivazione come sommatoria dei prodotti dei neuroni dello strato precedente e il vettore dei pesi, questo valore verrà successivamente modificato con una funzione di trasferimento, in genere la funzione sigmoide:

$$y = 1 / (1 + e^{-(x_1 + x_2 + \dots + x_n)})$$

E' evidente che ciascun neurone dello strato precedente conterà sulla sommatoria in funzione del valore relativo del proprio peso rispetto agli altri concorrenti. Tra uno strato e l'altro si può apprezzare l'"importanza" di ognuno, invece tra il primo e l'ultimo intercorre uno strato e ciascun input influenzerà un'uscita attraverso 5 sommatorie (strato 1 e 2) + altre 5 (ogni elemento interno e l'uscita) con relative uscite della sigmoide. Si perde così, la linearità che ci permetteva di proporzionare ciascun neurone a qualsiasi altro da esso collegato.

In babapro le connessioni sono limitate alle influenze logiche che si vogliono analizzare, con quest'architettura:



Come si può vedere, le connessioni convergono verso un'uscita senza dividersi in diversi rami per poi confluire in un unico neurone successivo. In questa maniera, le varie proporzioni si mantengono confrontabili. La ridotta interazione rende questo algoritmo piuttosto lento nella convergenza e non riesce a risolvere problemi più complessi.

Nel nostro caso, però, la scarsa dotazione d'intelligenza è bastata.

Nella tradizionale back propagation si ragiona a strati; lo stato di attivazione del successivo è calcolato parallelamente mediante un prodotto di matrici (input e pesi). In babapro, invece, ogni neurone, ricevuti i segnali dai propri ingressi, ne emette uno (nei rari casi di multipla emissione la rete va considerata come diverse reti indipendenti) verso le proprie uscite; non c'è alcuna struttura che supervisiona ed ogni elemento è completamente isolato, un po' come il cervello umano secondo i connettivisti.

La correzione degli errori si propaga all'indietro secondo la regola delta classica.

Va ricordato che la presenza del bayesiano nel nome dato al programma, non rappresenta affatto il calcolo delle probabilità condizionate:

L'espressione del coefficiente appare un po' boriosa e non utile alla comprensione:

$$P(A|B) = P(B|A) P(A) / P(B),$$

ma vuole solo evocare la dipendenza dei vari elementi, alla maniera delle reti ad albero che lo usano. La rete, nella fase di addestramento, riceve dall'esterno solo gli ingressi e le uscite corrette e

alla fine del training, i pesi acquistano valori prossimi ai coefficienti condizionati ma solo grazie alla regola del gradiente ed alle costrizioni nelle connessioni.

4. Descrizione dei vari indici

L'addestramento della rete è avvenuto con un campione di 15.000 agenzie, costituito da agenzie rapinate e non; tra le possibili ipotesi di lavoro futuro vi è quella di addestrare al rete solo con le agenzie rapinate.

Ecco una descrizione analitica dei vari indici:

RISCHIO ESOGENO

Rischio (probabilità) di un evento rapina legato alle seguenti caratteristiche ambientali:

- intensità criminosa su base comunale (dati Ministero Interno);
- densità popolazione su base comunale (dati Istat)
- densità sportelli su base comunale (dati Banca d'Italia).

RISCHIO ENDOGENO

Rischio (probabilità) di un evento rapina legato alle caratteristiche dell'agenzia, sulla base dei seguenti parametri:

- peculiarità;
- sistemi di difesa antirapina;
- servizi antirapina;
- rapine subite.

RISCHIO GLOBALE

Indice di rischio globale associato all'agenzia in esame calcolato mediante una funzione combinatoria applicata ai due precedenti indici (esogeno ed endogeno).

5. Prestazioni della rete ed interpretazione dei risultati

In genere, è considerato il 10% (valore medio d'errore) il valore limite sotto il quale una rete può dirsi convergente, mentre la nostra non riesce a scendere al di sotto del 17%. Questo sembrerebbe un risultato scadente e poco attendibile, ma non è così; in questo paragrafo si cercherà di spiegare il perché di questa affermazione.

Considerando la risposta della rete in relazione a vari fenomeni, possiamo individuare i seguenti casi:

1. **Fenomeno perfettamente modellabile:** se si addestra un algoritmo neuronale a correlare misure di angoli con relativi seno o coseno, la discesa dell'errore è decisa e la sua rapidità o accuratezza dipende solo dal numero di neuroni o dal tempo di addestramento.
2. **Fenomeno con assenza di correlazione tra gli ingressi e le uscite:** se si pretende che la rete correli il colore dei propri calzini con l'andamento del mercato azionario, non ci sarà nessuna riduzione di errore e la risposta della rete oscillerà intorno al 0.5 – che è il punto di mezzo tra 0 e 1, cioè massima indecisione - con fluttuazioni periodiche in funzione del learning rate impostato.
3. **Fenomeno con dati esigui o scadenti:** se il set di dati contiene alcuni errori o sono di scarsa accuratezza (es. si fermano alla prima cifra decimale quando è determinante la seconda)

oppure non se ne dispone di una quantità sufficiente, la rete scenderà sotto il 15% – 30%, ma non di più, neppure modificando la quantità di neuroni o prolungando il periodo di addestramento. Questo è il nostro caso.

La quantità di rapine e il numero d'anni d'osservazione sono insufficienti per una modellazione più accurata, ma la - seppure modesta - discesa d'errore ci assicura che le correlazioni trovate sono valide.

L'uscita della rete, cioè la frequenza di subire eventi in un anno, varia in modo continuo da 0 a 4, e viene normalizzata tra 0 e 1, ma non linearmente. La media di questo valore è del 20%, cioè un quinto di evento per anno, o meglio, un evento ogni 5 anni, la soglia minima, cioè 1 evento in tutto il periodo d'osservazione, è 0.1666... Per esaltare anche un solo caso di rapina e attenuare i record con troppe rapine, si è normalizzato i dati con la seguente espressione:

$$y = (\alpha + \log(\beta + x)) / \gamma$$

in maniera che per $x=0$ (valore minimo):

$$(\alpha + \log(\beta)) / \gamma = 0$$

per un opportuno x_1 (valore di accensione: 0.1666...):

$$(\alpha + \log(\beta + x_1)) / \gamma = 1/2$$

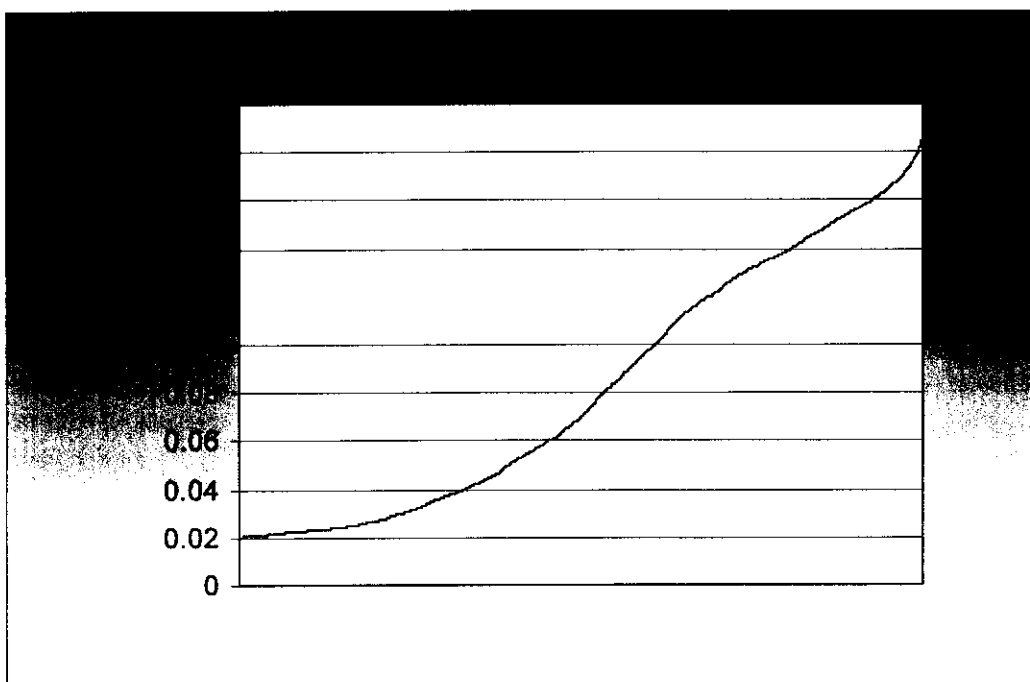
per $x=1$ (valore massimo):

$$(\alpha + \log(\beta + 1)) / \gamma = 1$$

In questo modo, il valore minimo, che con una normalizzazione lineare sarebbe stato: $(1/6)/4=0.041666...$ (4 è il valore massimo), vale poco meno di 0.5, e quello medio (0.2) è esattamente 0.5.

Sempre a causa del doppio significato dell'indice endogeno – probabilità e frequenza – si è voluto verificare quanto la rete sbaglia sui record di agenzie senza rapine e quanto su quelli con rapine. Simultaneamente, si è deciso di verificare come variava la risposta con l'aumentare del tempo di apprendimento. Così, abbiamo addestrato 3 reti per rispettivamente 5,30 e 90 minuti e ne abbiamo confrontato i risultati.

Inoltre, per valutare lo stato di eccitazione del neurone d'uscita, è stato verificato l'andamento per tutti i record ed è stata scelta una soglia, per esempio nel caso della rete da 5 minuti, il valore è stato di 0.1:



Con questi risultati:

	rete da 5'	%	rete da 30'	%	rete da 90'	%
stato eccitazione corretto su eventi rapina	3361	0.54	2796	0.45	1716	0.28
stato eccitazione errato su eventi non rapina	8598	0.24	6348	0.18	3612	0.10

Come si può vedere, la rete meno addestrata è più eccitabile e indovina più delle altre gli eventi di rapina, ma ne sbaglia parecchio nell'altro caso.

Considerando che una quota di rapine accadute può non essere dipeso affatto dalle strutture difensive o dalle altre caratteristiche dell'agenzia (e quindi per niente prevedibile), un risultato del 28% di evento previsto correttamente (rete più esperta e primo rigo) e di appena 0.1 di errore sulla valutazione errata è eccellente.

