

METODI DI DATA MINING ASSOCIATI ALL'APPROCCIO FUZZY: METODOLOGIA INTEGRATA PER L'ANALISI DEGLI INFORTUNI SUL LAVORO

R. LUZZI¹, G. FOIS¹, S. MURÈ³, F. PALAMARA², N. PICCININI²

¹INAIL - Direzione Regionale Piemonte - Consulenza Tecnica Accertamento Rischi e Prevenzione

²SAFeR- Centro Studi su Sicurezza, Affidabilità e Rischi - Dip. Scienza dei Materiali e Ingegneria Chimica - Politecnico di Torino

³ARIA s.r.l. – Analisi dei Rischi Industriali e Ambientali - Torino

RIASSUNTO

L'articolo presenta il risultato delle attività sviluppate presso il Politecnico di Torino nell'ambito del Dottorato in Sicurezza Industriale e Analisi dei Rischi. L'approccio descritto è utile per l'analisi di basi dati ampie come quella mantenuta dall'INAIL sugli infortuni occorsi nell'industria, che classifica e raccoglie per ogni caso di infortunio i fattori che hanno portato all'evento e le relative conseguenze. La prima parte del lavoro descrive lo sviluppo di una procedura di "data mining" che permette di individuare correlazioni significative tra i fattori caratterizzanti gli infortuni, proprio grazie all'elevato numero dei parametri e dei dati esistenti.

Queste tecniche permettono di raggruppare i dati (clustering), classificarli in termini di livello di rischio, e identificare in ogni raggruppamento una specifica sequenza di eventi caratterizzanti gli infortuni, che meritano un'attenzione prioritaria nella definizione degli interventi necessari per ridurre il rischio d'infortunio. Questi strumenti costituiscono il punto iniziale per un'analisi di maggiore dettaglio per una valutazione del rischio infortunistico caratteristico di ogni singola azienda. A questo scopo è stato sviluppato un metodo basato sulla Logica Fuzzy, costruito su quattro fattori: la frequenza di accadimento di un evento (F), il probabile danno da esso derivato (S), il fattore di contatto (E) e una valutazione dell'adeguatezza dei sistemi di protezione esistenti al momento dell'incidente (L).

SUMMARY

The paper introduces the results of work developed by Turin Polytechnic in the within of the Doctorate in Industrial Safety and Risk analysis. The described approach is useful for the analyses of wide databases like that maintained from INAIL on accidents in the industry, being every case of accident classified and collected by factors that have carried at the event and the relative consequences. First part of the job describes the development of a data-mining procedure that allows to characterize meaningful correlations between the characterizing factors, thanks to elevated number of the parameters and the existing data.

These techniques allow to group data (clustering), to classify them in terms of level of risk, and to identify in every grouping a specific sequence of characterizing events the accidents, that they deserve a priority attention in the definition of the participations necessary in order to reduce the risk accident. These instruments constitute the start point for the analysys of greater detail of risk assessment for each company. For this purpose it has been developed a method based on Fuzzy Logic, constructed on four factors: the frequency of occurrence of an event (F), the probable damage from it derived (F), the factor of contact (F) and an appraisal of the adequacy of the existing systems of protection at the moment of the accident (L).

1. INTRODUZIONE

Gli infortuni sul lavoro rappresentano complessivamente una grave perdita che la comunità subisce per la realtà del tessuto produttivo, che risulta presentare, nonostante le regole imposte dalla normativa in vigore siano abbastanza stringenti, lacune organizzative e tecnologiche, che possono

trovare origine da varie cause che spaziano, per citarne alcune, dalle carenze del rapporto contrattuale (precarietà, inidoneità alla mansione, eccessiva mobilità, ecc.), a carenze dei processi produttivi (per inadeguatezza delle soluzioni praticate sotto gli aspetti sia organizzativi che tecnologici), o finanche a carenze nella cultura della sicurezza (formazione parziale, mancata sensibilizzazione, scarso coinvolgimento negli obiettivi di benessere psicofisico). Ma dagli stessi infortuni si possono trarre delle informazioni che alimentano il bagaglio di conoscenze che sta a base dell'azione di prevenzione, trasformandole in insegnamenti che possono orientare la scelta delle misure preventive da adottare nei comparti produttivi. Con l'analisi delle statistiche degli infortuni, originate soprattutto dalla banca dati dell'INAIL, questa raccolta di informazioni per la prevenzione è forse sempre stata fatta, ma ora, con l'evoluzione dei sistemi informatici, ci si trova dinanzi all'opportunità di poter analizzare una enorme quantità di dati organizzati. Infatti, la banca dati dell'INAIL, con l'introduzione della codifica europea ESAW, contiene per ciascun infortunio registrato molte più informazioni: vi si trovano molti elementi utili a descrivere la dinamica dell'infortunio. Ciò ha permesso di arrivare ai flussi informativi a livello regionale: i dati scambiati tra i vari enti che gestiscono gli infortuni vengono così ulteriormente arricchiti. A fronte di questa enorme mole di dati, nasce quindi l'esigenza di affinare le metodiche di analisi e di elaborazione dei risultati utili a definire le misure preventive per i comparti produttivi. Da alcune pregresse esperienze si è riscontrata una difficoltosa lettura dei risultati anche considerando ambiti di settori produttivi abbastanza ristretti, in cui si è svolta la raccolta dei dati ESAW.

Da quanto sopra espresso, sono risultate particolarmente interessanti le attività sviluppate presso il Politecnico di Torino, nell'ambito del Dottorato in Sicurezza Industriale e Analisi dei Rischi (Luzzi et al., 2007). L'approccio sviluppato è utile per l'analisi di basi dati ampie come quella INAIL degli infortuni, che raccoglie per ogni *record*, in forma codificata, i fattori che hanno portato all'evento e le relative conseguenze.

La ricerca svolta, si è articolata:

- nello sviluppo di una procedura di "data mining" che permette di individuare correlazioni significative tra i fattori caratterizzanti gli infortuni (visto l'elevato numero dei parametri e dei dati esistenti),
- nell'utilizzo di tecniche ed algoritmi che hanno permesso di raggruppare i dati (clustering) e classificarli in termini di livello di rischio e forma di accadimento,
- nella valutazione del rischio infortunistico di maggiore dettaglio, sviluppando un metodo basato sulla *Logica Fuzzy*, considerando la frequenza di accadimento di un evento (F), il probabile danno da esso derivato (S), il fattore di contatto (E) e una valutazione dell'adeguatezza dei sistemi di protezione esistenti al momento dell'incidente (L),
- nella validazione del metodo in aziende (volontariamente offertes).

I comparti produttivi considerati sono stati quelli della lavorazione del legno, della manifattura della gomma e dell'industria siderurgica. Ulteriori attività di sperimentazione del metodo in questo ultimo comparto sono ancora in corso.

Le attività non potrebbero prescindere da un consenso ottenuto dalle parti sociali e datoriali del territorio piemontese, che sono state coinvolte sin dall'inizio e da cui sono ancora arrivati positivi segnali di interesse.

2. ANALISI DEGLI INFORTUNI MEDIANTE METODI DI DATA MINING

I dati, su cui la metodologia è stata sviluppata e testata, sono stati forniti dall'INAIL regionale del Piemonte e riguardano circa 1200 casi di infortunio occorsi nel comparto della lavorazione del legno, in Piemonte, tra il 2002 e il 2004. I casi di infortunio riguardano sia laboratori artigianali che vere e proprie industrie e comprendono tutte le fasi della lavorazione del legno: dalla prima fase di lavorazione del legno ai manufatti finiti come mobili o infissi.

Sono stati presi in considerazione i fattori che descrivono la dinamica di occorrenza di ciascun infortunio: l'attività svolta immediatamente prima dell'infortunio, le anomalie per cui l'attività ha subito una deviazione dal suo normale svolgimento, la modalità di contatto che ha causato la lesione

e gli agenti materiali coinvolti. Queste variabili sono categoriche ossia discrete e nominali e non ordinabili, per cui per essere analizzate tramite algoritmi necessitano di una codifica ad hoc. La codifica scelta è di tipo binario ed è esemplificata nella Tabella 1.

Per ogni variabile sono state scelte 8 possibili voci.; ciascuna delle 8 possibilità è caratterizzata da una successione di sette 0 e un 1 e, per ogni variabile, le 8 possibilità differiscono le une dalle altre dalla differente posizione dell'unità.

Tabella 1. Codifica dell'attività svolta al momento dell'infortunio

Operazioni con macchine	1	0	0	0	0	0	0	0
Lavori con utensili	0	1	0	0	0	0	0	0
Conduzione mezzi	0	0	1	0	0	0	0	0
Manipolazione oggetti	0	0	0	1	0	0	0	0
Apertura/chiusura imballaggi	0	0	0	0	1	0	0	0
Spargere/ versare/ riempire	0	0	0	0	0	1	0	0
Trasporto manuale	0	0	0	0	0	0	1	0
Movimento persona	0	0	0	0	0	0	0	1

Oltre all'attività svolta sono stati analogamente codificati: l'anomalia che ha portato all'infortunio (Incendio, Perdite di materiale Rottura – Caduta materiale, Perdita di controllo di macchina-utensile, Caduta o scivolamento persona, Movimenti del corpo, Sorpresa – Spavento), il contatto che ha portato alla lesione e l'agente materiale coinvolto nell'infortunio.

I 1200 casi di infortunio sono stati pertanto trasformati in 1200 vettori. Ciascun vettore è costituito di 48 elementi che corrispondono alla codifica in 8 termini binari dei 6 parametri che caratterizzano i casi di infortunio.

Le tecniche di analisi che sono state applicate a questi dati implicano misure di distanza; nel caso di vettori binari una metrica euclidea non è opportuna; è necessario ricorrere a quelle che sono definite: distanze di similarità. La distanza scelta in questo lavoro è la distanza di Hamming, definita nelle equazioni (1. a) e (1.b); maggiore è il valore di questa distanza, più i due vettori e quindi i due casi di infortunio posti a confronto differiscono tra loro.

$$H(x, y) = \frac{\sum_{i=1}^n \delta(x_i, y_i)}{n} \tag{1.a}$$

$$\delta(x_i, y_i) = \mathbf{0} \text{ if } x_i = y_i \quad \mathbf{1} \text{ if } x_i \neq y_i \tag{1.b}$$

dove:

x e y sono i due dati a confronto

n è il numero delle variabili che caratterizzano ciascun dato.

I casi di infortunio sono stati analizzati con tecniche di 'Data mining'. Si tratta di un insieme di metodi che, applicati a grandi database, permettono di scoprire relazioni significative tra i dati, al fine di renderli più comprensibili e utilizzabili. In particolare, in questo lavoro, l'obiettivo da raggiungere e il tipo di dati a disposizione hanno suggerito di scegliere tecniche di data mining che classificassero i casi di infortunio in funzione della dinamica che caratterizza la loro occorrenza. Si tratta di una classificazione in cui non si conoscono a priori le classi a cui i casi di infortunio appartengono, per cui si parla di classificazione non supervisionata o 'Cluster analysis'. L'obiettivo di una cluster analysis è quello di raggruppare i dati in classi, chiamate clusters, in modo tale che ci sia un'elevata similarità tra i dati all'interno dello stesso cluster e un'elevata dissimilarità tra i dati di clusters differenti. Le tecniche di cluster analysis, utilizzate in questo lavoro, sono descritte brevemente nelle sezioni seguenti.

1) Self Organizing Maps. La struttura della ‘Self Organizing Maps’ (SOM) utilizzata in questo lavoro consiste in una mappa bidimensionale di 504 unità connesse in un reticolo esagonale. Ciascuna unità è caratterizzata da un valore inizializzato con un processo casuale. L’apprendimento della mappa è effettuato utilizzando un algoritmo chiamato ‘Batch training algorithm’ che calcola le distanze tra ciascun dato (ossia ciascun caso di infortunio codificato in termini binari) e ciascuna unità della mappa. In un processo di tipo iterativo i valori delle unità della mappa sono aggiornati in modo tale da rendere ciascuna unità della mappa rappresentativa di un particolare tipo di caso di infortunio. L’aggiornamento dei valori delle unità della mappa, effettuato ad ogni step, è descritto nell’equazione (2).

$$w_{j(t+1)} = \frac{\sum_{i=1}^n h_{c(j)}(t) \cdot x_i}{\sum_{i=1}^n h_{c(j)}(t)} \quad (2)$$

dove

$w_j(t+1)$ sono i valori delle unità, aggiornati durante l’apprendimento;

n è il numero dei dati di input;

x_i sono i dati di input;

$h_{c(j)}(t)$ è una funzione gaussiana riferita ad ogni unità di output c_j che risulta essere rappresentativa di un dato di input;

t sono gli steps in cui si svolge il processo iterativo di apprendimento.

In questo modo, nel processo di apprendimento, casi simili tra loro, sotto il profilo della dinamica, tendono ad essere connessi alla medesima unità nella mappa e dati dissimili tra loro tendono ad essere connessi a unità distanti tra loro nella mappa.

2) K-means. Consiste in un processo di tipo iterativo, in cui i dati sono suddivisi in un certo numero di gruppi, chiamati ‘Clusters’. Il numero k di clusters è scelto a priori e inizialmente, con un processo di tipo random, sono estratti k dati che dovrebbero fungere da ‘centroidi’ dei k clusters. L’algoritmo calcola quindi le distanze tra ogni dato a disposizione e ciascun centroide, utilizzando la relazione (3) e associa ciascun dato al centroide più vicino.

$$x \leftrightarrow Ci \text{ if } |x - c_i| < |x - c_j| \quad (3)$$

dove

c_i è il centroide del cluster Ci , c_j il centroide del cluster Cj , x è un dato di input.

Una volta formati i k clusters, il centroide di ciascun cluster è ricalcolato come descritto dalla seguente equazione (4).

$$c_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ji} \quad (4)$$

dove

c_i è il centroide del cluster Ci

m_i è il numero degli x_j dati raccolti nel cluster Ci .

Il processo continua con il calcolo delle distanze tra i dati e i nuovi centroidi e con la conseguente riassegnazione di ciascun dato al centroide più vicino.

3) Leader Follower. Questo metodo di clustering richiede la definizione a priori del valore di una ‘distanza soglia’. Inizialmente un primo dato di input è selezionato come centro del primo cluster. Poi un secondo dato è confrontato con il primo e, se la distanza tra i due dati risulta minore della ‘distanza soglia’, questo secondo dato si aggiunge al primo nella formazione del cluster e il valore

del centro del cluster è aggiornato secondo l'equazione (5); altrimenti, il secondo dato diventa lui stesso centro di un nuovo cluster. Questo processo si ripete per tutti i dati.

$$wi(t+1) = wi(t) + \frac{(xi - wi(t))}{n} \quad (5)$$

dove

$wi(t+1)$ è il valore aggiornato del centro del cluster i

$wi(t)$ è il valore del centro del cluster i che deve essere aggiornato

xi è il dato assegnato al cluster i

n è il numero di dati raccolti nel cluster i all'iterazione $t+1$.

4) Hierarchical Clustering. Questo algoritmo calcola le distanze tra ogni coppia di dati di input in modo che le coppie di dati che risultano più vicini formino un unico cluster. Il processo si ripete con la formazione di clusters sempre più grandi, finché il numero di cluster scelto è raggiunto.

Le distanze tra i clusters, in questo lavoro, sono valutate in base ad un criterio di legame, chiamato 'Ward method', definito nell' eq. (6).

$$d(r, s) = \frac{n_r n_s d^2(\bar{x}_r, \bar{x}_s)}{n_r + n_s} \quad (6)$$

n_r, n_s sono il numero di casi raggruppati rispettivamente nel generico cluster r e nel generico cluster s

$d^2(\bar{x}_r, \bar{x}_s)$ è la distanza tra il centroide del cluster r e quello del cluster s

5) Visualizzazione dei possibili clusters di dati. I risultati ottenuti possono essere visualizzati in Figura 1 (visione bidimensionale) e in Figura 2 (visione tridimensionale). In queste figure, è visualizzata la mappa SOM dopo il processo di apprendimento. Ciascun esagono rappresenta una unità, etichettata con un numero. I grandi punti neri indicato le 'unità attivate' ossia le unità a cui sono stati associati dei dati di input. In seguito all'apprendimento, a causa dell'aggiornamento del proprio valore, ciascuna unità è diventata più o meno "vicina" alle altre. Le unità vicine tra di loro hanno valori simili e ad esse sono quindi associati dati di input caratterizzati da successioni di 0 e 1 simili tra loro ossia con dinamiche di infortunio analoghe. La scala delle distanze corrisponde sulle mappe, nelle figure, ad una scala di colori: dal blu al rosso. Al colore blu corrispondono le distanze minori, al colore rosso le distanze maggiori. In Figura 2, le zone montuose rappresentano le aree in cui le unità sono distanti tra loro e le valli quelle in cui le unità sono vicine. L'utilizzo di un algoritmo di apprendimento per l'analisi dei casi di infortunio ha permesso di 'proiettare' e quindi visualizzare questi dati, caratterizzati da un elevato numero di parametri, su una mappa bidimensionale. L'algoritmo ha inoltre permesso di associare a unità della mappa topologicamente vicine tra loro e quindi di raggruppare, casi di infortunio simili tra loro, sotto il profilo della dinamica di accadimento. Osservando la mappa, si notano 14 valli in blu che rappresentano 14 possibili clusters in cui possono essere suddivisi i dati di input.

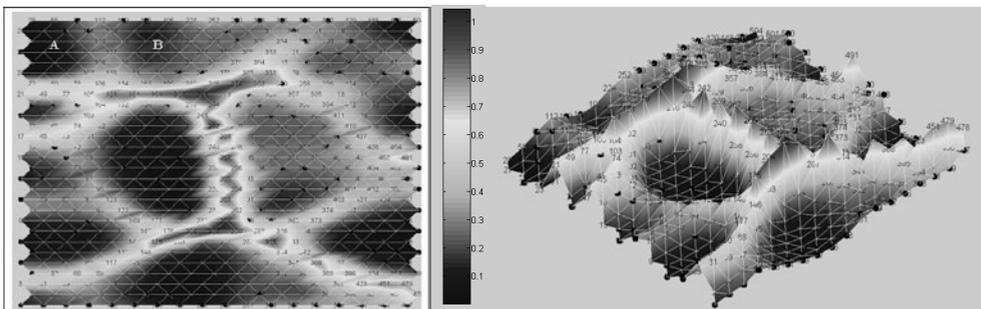


Figura 1. Visualizzazione 2D della SOM Figura 2. Visualizzazione 3D della SOM

Clustering dei casi di infortunio. I risultati descritti sopra hanno suggerito di applicare ai casi di infortunio algoritmi che potessero realizzare un effettivo clustering dei dati.

Pertanto gli algoritmi *K-means* e *Follow the leader* sono stati applicati per suddividere la mappa in aree che raccogliessero casi di infortunio simili. I risultati sono illustrati nelle Figg. 3 e 4 in cui si può osservare come questi algoritmi identifichino proprio le 14 aree che sono state visualizzate con la semplice osservazione della mappa. Le aree delimitate dai tratteggi in rosso sono quelle identificate da K-means, quelle delimitate dai tratteggi in nero sono state identificate dall'algoritmo Follow the leader.

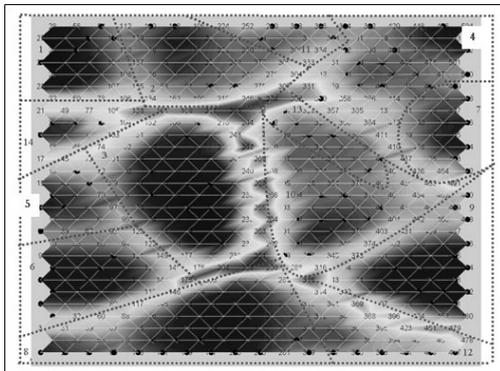


Figura 3. Clusters identificati dal K-means

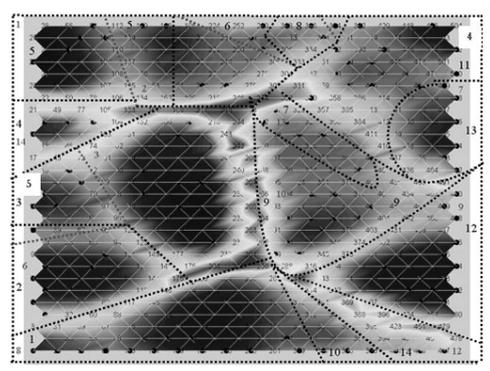


Figura 4. Clusters identificati da Follow the l.

Al fine di valutare l'accuratezza dei risultati sopra descritti è stato applicato, direttamente ai casi di infortunio, un terzo tipo di algoritmo: lo *Hierarchical clustering*. La Figura 5 mostra un dendrogramma che è la sintesi grafica dei risultati ottenuti con questo tipo di algoritmo; sull'asse delle ascisse sono presenti i dati di input e sull'asse delle ordinate le distanze tra i clusters che l'algoritmo ha creato. Questo algoritmo individua i medesimi 14 clusters creati da K-means e Follow the leader.

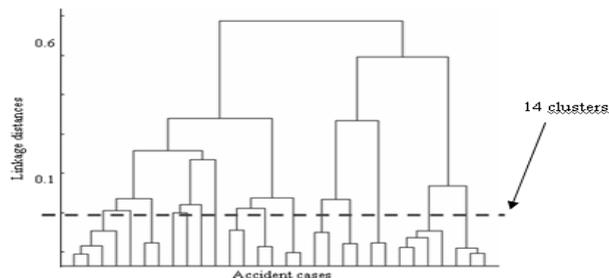


Figura 5. Dendrogram creato dall'algoritmo 'Hierarchical'

Sono stati quindi analizzati i casi di infortunio associati alle unità della mappa che appartengono alla medesima area individuata dagli algoritmi di clustering. Dall'analisi è risultato che i casi di infortunio appartenenti alla medesima area e quindi al medesimo cluster sono caratterizzati dalla medesima dinamica. In questo modo sono state identificate 14 differenti dinamiche che causano gli infortuni nell'industria del legno.

Come illustrato in Tabella 2, i tre metodi di clustering usati sono stati infine confrontati, in termini di accuratezza, mediante l'indice *I* descritto nell'eq. (7).

$$I = \frac{\sum_{v=1}^k a_v}{n} \quad (7)$$

dove

a_v è il numero di casi di infortunio correttamente assegnati al cluster

k è il numero di clusters

n è il numero di casi di infortunio.

Tabella 2. Confronto tra l'accuratezza dei metodi

METODO	I
SOM plus Follow the Leader	94 %
SOM plus K-means	92 %
Hierarchical clustering	88 %

6) Interpretazione dei clusters. Per ogni cluster individuato, è stato calcolato un indice di frequenza, considerando la percentuale dei casi di infortunio raccolti in ciascun cluster rispetto al numero totale di casi di infortunio presenti nella banca dati fornita dall'INAIL. Il numero di casi che hanno implicato perdite anatomiche, presenti in ciascun cluster, è stato invece usato come indice di gravità per ogni cluster. In base ai valori degli indici di frequenza e gravità ad ogni cluster è stato associato un livello di rischio al fine di individuare le dinamiche infortunistiche che prime tra tutte devono essere evitate. Il lavoro con utensili manuali associato ad una perdita di controllo dell'utensile o ad un movimento non corretto dell'operatore è stato individuato come situazione più critica nell'industria del legno, in quanto ha implicato il maggior numero di casi di infortunio e di perdite anatomiche.

3. ANALISI DEGLI INFORTUNI MEDIANTE FUZZY LOGIC

3.1 Descrizione caratteristiche principali FAP (Fuzzy Application Procedure)

Con l'obiettivo di identificare "a priori" le possibili misure preventive e protettive più idonee, atte a ridurre il rischio d'infortunio in un determinato contesto lavorativo è stato messo a punto uno strumento metodologico per una sua valutazione quantitativa o semiquantitativa basato sulla Logica Fuzzy. I vantaggi di un approccio di questo tipo sono stati illustrati mediante l'applicazione ad un campione di dati forniti da un'azienda produttrice di pneumatici.

Nell'ambito di questa ricerca è stata sviluppata un'applicazione software denominata FAP, Fuzzy Application Procedure, che permette di rendere facile all'utente finale l'utilizzo delle enormi potenzialità di uno strumento di calcolo come il fuzzy toolbox di Matlab® e di uno strumento di organizzazione e presentazione dei dati come Excel® (il cui utilizzo è prassi comune). Per sfruttare al meglio queste potenzialità, entrambi i prodotti sono dotati di un'interfaccia COM (Component Object Model) che permette ad altre applicazioni, che sfruttano la metodologia di programmazione ad oggetti (OOP - Object Oriented Programming), di lanciare e pilotare queste applicazioni. Come linguaggio di programmazione è stato scelto Microsoft Visual Basic® (nominato VB in seguito) per la sua facilità, la sua versatilità e perché, mediante il suo utilizzo è possibile creare agevolmente interfacce grafiche molto intuitive all'utente finale. Accanto al linguaggio VB, sono state create delle function Matlab per rendere più semplice e veloce il suo interfacciamento con VB.

Partendo dall'espressione del *rischio tecnologico* $R = f(F, M)$ (Rasmussen, 1975), del *danno probabile derivante da un evento* $M = f(S, E)$ (Patrucco, 1997), e introducendo un giudizio sull'adeguatezza delle misure protettive, si è giunti alla seguente funzione del rischio d'infortunio:

$$R = f(F, S, E, L)$$

dove F è la *frequenza di accadimento dell'evento* [eventi/giorno], S è la *gravità o l'entità del possibile danno* [giorni di assenza dal lavoro], E è il *fattore di contatto* [frazione del turno di lavoro] e L è il *grado di adeguatezza delle misure protettive* [adimensionato].

Con il semplice legame funzionale AND, secondo una prassi consolidata da tempo (Patrucco, 1997) diventa:

$$R=F*S*E*L$$

La costruzione di un sistema fuzzy attraverso il fuzzy toolbox di Matlab si distingue nei seguenti step:

1. Definizione degli insiemi fuzzy delle variabili di input e di output.
2. Definizione delle regole che correlano le variabili di input e di output
3. Aggregazione dei contributi delle regole.
4. Defuzzificazione dei risultati.

Tabella 3. *Insiemi fuzzy, variabili linguistiche e corrispondenti interventi suggeriti.*

Frequenza di accadimento	Gravità	Fattore di contatto	Grado di adeguatezza delle misure protettive	Rischio	Proposte di interventi migliorativi
Remota 0-0.3	Molto lieve 0-0.3	Basso 0-0.3	Ottimo 0-0.3	Basso 0-0.25	Rischio tollerabile: nessuno intervento
Bassa 0.2-0.5	Lieve 0.2-0.5	Medio 0.2-0.5	Buono 0.2-0.5	Medio 0.26-0.45	Rischio moderatamente non tollerabile: interventi di media entità
Media 0.4-0.7	Media 0.4-0.7	Alto 0.4-0.7	Sufficiente 0.4-0.7	Alto 0.46-0.75	Rischio non tollerabile: interventi consistenti
Alta 0.6-1	Grave 0.6-1	Altissimo 0.6-1	Carente 0.6-1	Altissimo 0.76-1	Rischio decisamente intollerabile: interventi urgenti e consistenti

Il numero delle regole è fortemente legato al numero delle variabili di input e al corrispondente numero di insiemi fuzzy.

La FAP è stata applicata su un campione di infortuni avvenuti in un'acciaieria con una forza lavoro pari a 1.000 persone. La maggior parte delle produzioni avviene su turno continuo (ore 6.00 ÷ 14.00, 14.00 ÷ 22.00, 22.00 ÷ 6.00), di seguito sono elencate le aree produttive e l'area di deposito dello Stabilimento:

1. Area di fabbricazione e prima trasformazione: area in cui avviene la prima trasformazione a caldo con produzione di lingotti, tondi, piatti forgiati, ecc.
2. Area di laminazione: area che si compone di un unico reparto "Laminatoio vergelle/barre" e in cui avviene la produzione dei suddetti semilavorati.
3. Area di finitura: in questa area si completa il ciclo di lavorazione con la produzione di vergelle decapate, barre in acciaio inox, barre e blocchi per utensili e barre per valvole di motori endotermici.
4. Area di deposito: il deposito riguarda le materie prime, quali i materiali ferrosi, le ferroleghie, e i refrattari, nel reparto "Deposito rottami" e i prodotti finiti nel reparto "Magazzino".

E' stato possibile individuare i reparti critici (quello del decapaggio, in particolare per la fase dei trattamenti termici, e il laminatoio barre/vergelle) e le mansioni più colpite.

3.2 Applicazione

Il modello di valutazione del rischio d'infortunio sviluppato si basa sulla seguente espressione del rischio:

$$R = F \text{ and } M \text{ and } E \text{ and } L$$

vale a dire sul prodotto logico tra la Frequenza di accadimento, la Magnitudo o gravità del danno, il Fattore di contatto (frazione del turno di lavoro) e il Grado di adeguatezza delle misure protettive. Si osservi come i dati disponibili, tratti dai registri infortuni, non consentano di stimare E e L, pertanto le valutazioni del rischio riportate di seguito si riferiscono ad un'espressione semplificata del rischio, vale a dire $R = F \text{ and } M$.

Con l'obiettivo di quantificare il rischio medio di reparto, si è proceduto con la valutazione del rischio di ciascun infortunio avvenuto in reparto. A tal fine, F è diventata la frequenza di accadimento normalizzata (o relativa), ovvero il rapporto tra la frequenza di accadimento di reparto e la frequenza di accadimento massima di stabilimento, ovvero quella del reparto con il più alto numero di infortuni; M diventa la magnitudo normalizzata (o relativa), ovvero il rapporto tra la prognosi dell'infortunio considerato e la prognosi massima.

Ad esempio, con riferimento all'anno 2002, la frequenza di accadimento relativa del reparto "Acciaieria" è il rapporto tra la frequenza di accadimento del suddetto reparto e quella del reparto "Manutenzione", che registra il più alto numero di infortuni.

Si è ammesso inoltre che l'anno lavorativo abbia una durata di 200 giorni per tutti i reparti e che lo Stabilimento abbia circa 1000 addetti. Si è osservato come più del 90% degli infortuni registrati nel triennio di riferimento abbiano una prognosi minore di 40 giorni (valore limite stabilito dagli Organi di Vigilanza oltre il quale si sottopone l'infortunio ad inchiesta). Al fine di non trascurare i suddetti infortuni e quindi la pericolosità di interi reparti, la magnitudo è stata riferita alla prognosi di 40 giorni e non alla prognosi massima di stabilimento. Gli insiemi fuzzy delle variabili sono riportati nella tabella seguente.

Tabella 4. Insiemi fuzzy basati su frequenza e magnitudo dell'evento

Frequenza di accadimento		Gravità	
Valore normalizzato	Infortuni per giorno per lavoratore (*10 ⁻⁴)	Valore normalizzato	Giorni di prognosi
Remota		Bassissima	
0 - 0,225	0- 0,405	0 - 0,025	0 - 1
Bassa		Bassa	
0,025 - 0,475	0,045 - 0,855	0,225 - 0,475	9 - 19
Media		Media	
0,275 - 0,725	0,495 - 1,306	0,275 - 0,725	11 - 29
Alta		Alta	
0,525 - 0,975	0,945 - 1,755	0,525 - 0,975	21 - 39
Altissima		Altissima	
0,775-1	1,395 - 1,80	0,775-1	31 - 40

Il livello di rischio di reparto è stato calcolato con riferimento agli infortuni con una prognosi inferiore ai 40 giorni; pertanto, al fine di poter stabilire delle priorità di intervento, nei reparti in cui

sono presenti uno o più infortuni con prognosi minore di 40 giorni occorrerà attribuire al reparto un livello di rischio superiore a quello calcolato. Fermo restando la necessità di interventi tecnico/procedurali immediati per gli infortuni più gravi.

La valutazione del rischio medio per reparto è stata fatta considerando esclusivamente gli infortuni con prognosi inferiore a 40 giorni. Pertanto, nei reparti che hanno registrato infortuni con una prognosi maggiore, occorrerà tenerne conto al fine di poter stabilire efficacemente criticità e priorità di intervento.

Si osservi come gli infortuni con prognosi superiore ai 40 giorni, per prassi aziendali siano studiati caso per caso a fronte anche delle inchieste infortuni esterne ed interne, rispettivamente avviate dall'Organismo di vigilanza e dal Servizio di Prevenzione e Protezione aziendale.

Si osservi infine come in riferimento ai reparti più critici ("acciaieria", "colata continua", "fucina", "finitura", "decapaggio", "manutenzione"), alla gravità (prognosi > 40 giorni) e ad una particolare parte lesa, quale è l'occhio, sia stato individuato un campione di infortuni da sottoporre ad approfondimento.

4. CONCLUSIONI

La definizione di correlazioni significative tra i fattori caratterizzanti gli infortuni (visto l'elevato numero dei parametri e dei dati esistenti) ed il raggruppamento dei dati infortunistici (clustering) in termini di livello di rischio e forma di accadimento, sono tra i risultati di maggiore utilità applicativa in ambito prevenzionale. A questi si aggiunge l'utilità del metodo basato sulla *Logica Fuzzy*, che offre la possibilità di effettuare a partire da dati oggettivi la valutazione del rischio infortunistico.

I comparti produttivi su cui si è puntata l'attenzione sono stati quelli della lavorazione del legno, della manifattura della gomma e dell'industria siderurgica e si ribadisce che le attività sono state rese possibili grazie all'interesse condiviso dalle parti sociali e datoriali del territorio piemontese. E' stato raggiunto lo scopo di sviluppare una metodologia di trattamento dei dati utilizzabile sia a larga scala per l'analisi di comparto che in scala ridotta per la valutazione delle singole aziende.

Si osservi come tale metodologia, nonostante sia qui presentata una particolare applicazione, sia facilmente estendibile a qualsiasi azienda con il solo vincolo di disporre di un sufficiente e omogeneo numero di registrazioni infortuni, in modo da istruire correttamente il software e ottimizzare (tuning) i parametri di riferimento.

Il vantaggio di utilizzare una procedura automatizzata permette all'utente meno esperto di poter applicare e utilizzare degli strumenti di calcolo potenti ma complessi e nello stesso tempo di ottenere le informazioni derivanti dalla sua applicazione già organizzati e pronti per essere utilizzati e presentati.

BIBLIOGRAFIA

R. Luzzi, S. Murè, F. Palamara: Metodologie innovative per l'analisi del fenomeno infortunistico basata sui dati INAIL, 2007, 5° Seminario CONTARP, Atti pagg. 175-183.

N. Rasmussen: Nuclear Regulatory Commission Rasmussen Report, "An Assessment of Accidents Risk in US Commercial Nuclear Power Plants", 1975, WASH 1400, Washington.

M. Patrucco: "Sicurezza e Ambiente di Lavoro", ed. Trauben, Torino 1997.