4. LE RETI NEURONALI : UNO STRUMENTO PER L'INTERPRETAZIONE DELLA QUALITÀ DELL'ARIA

Per analizzare l'inquinamento, in relazione alle sue manifestazioni, con riferimento alle cause che lo determinano ed alle caratteristiche dei recettori, è necessario fare un uso combinato delle azioni di *monitoring* e *modeling*. Entrambe le attività, effettuate isolatamente, si risolvono, spesso, in uno spreco di risorse e di informazioni. Utilizzate insieme, invece, forniscono lo strumento adatto per individuare gli interventi più appropriati per garantire la salvaguardia dell'ambiente e della qualità della vita (Tamponi, 1997).

L'avvento di calcolatori sempre più potenti e a costo sempre più basso ha consentito, negli ultimi anni, il decollo nel settore ambientale dei modelli matematici, unico strumento capace di dare trasparenza alle decisioni, perché capace di interpretare i fenomeni e di emularli tutte le volte che se ne presenta la necessità.

Nelle varie attività connesse con il controllo della qualità dell'aria e con la gestione delle reti di rilevamento, sono stati sviluppati ed utilizzati modelli dotati di caratteristiche diverse. Per gestire la qualità dell'aria, infatti, è illusorio ritenere che esistano modelli universali adatti a tutti gli usi, perché lo spettro delle situazioni da fronteggiare è molto ampio, al pari degli obiettivi da perseguire.

Evitando una dettagliata rassegna di modelli matematici, si ritiene opportuno citare i principi informatori di quelli di interesse per lo studio dell'inquinamento atmosferico a scala di bacino. Si citano in particolare i modelli deterministici *a box* e *K* ed i modelli stocastici a regressione multipla. I primi si distinguono per essere i più semplici modelli deterministici in quanto ignorano la struttura spaziale dei fenomeni e, per questo, sono adatti allo studio di situazioni locali nelle quali dominano i meccanismi di reazione chimica.[1]

[1] In questi modelli l'aria viene trattata come un *box* completamente miscelato, all'interno del quale si verificano le immissioni degli inquinanti e le reazioni chimiche, tenendo conto anche del trasporto da e verso l'esterno. Questi modelli permettono di considerare nel dettaglio i meccanismi chimici ma ignorano la

I secondi sono più adatti allo studio di situazioni di bacino, in quanto si basano sulle equazioni della dinamica dei fluidi e tengono conto dei campi di vento tridimensionali. [2]

I modelli di regressione multipla si distinguono per essere stati ampiamente utilizzati ai fini della previsione giornaliera di eventi acuti di inquinamento (Comrie, 1997). Inoltre utilizzano algoritmi semplici che non richiedono necessariamente concettualizzazioni fisiche.

Tutti i modelli fino ad ora studiati, siano essi di ordine deterministico o stocastico, ed applicati all'inquinamento atmosferico, presentano, tuttavia, delle lacune dovute alla loro particolare rigidità, conseguenza degli schematismi delle relazioni di causa-effetto che vengono assunte.

La particolare complessità e non linearità delle relazioni connesse con l'inquinamento atmosferico suggeriscono l'utilizzo di uno strumento più elastico di quelli usuali. In particolare si sono dimostrate (Liguori, 1997) molto efficienti le reti neuronali poiché sono strutture adattive capaci di "apprendere" la soluzione dei problemi a partire da esempi noti. Tali sistemi sono nati ad imitazione del funzionamento del cervello umano e si ispirano al paradigma connessionista (Patarnello, 1991).[3] Le reti neuronali sono, quindi, dotate di proprietà che le rendono completamente diverse da ogni altra classe di modelli. Di particolare rilievo risultano le capacità di adattamento, generalizzazione, apprendimento e le facoltà elaborative altamente parallele (Boznar et Al., 1993).

Le facoltà di elaborazione parallela scaturiscono dalla capacità delle reti neuronali di apprendere e di emulare un sistema reale operando con più informazioni in parallelo, modificando la propria struttura interna fino a minimizzare ogni differenza dal sistema reale.

---

dispersione. Per tali motivi, sono strumenti meno complessi di altri, ma non sempre danno risultati più grossolani (ISTISAN, 1993).

[2] Nei modelli *K* la concentrazione degli inquinanti è calcolata nelle maglie di un reticolo tridimensionale che ricopre il dominio d'interesse (ISTISAN, 1993).

[3] Il paradigma connessionista assume che un sistema, naturale o artificiale, manifesti, al limite, comportamenti intelligenti se possiede le seguenti caratteristiche : (i) ha un numero elevato di componenti elementari (neuroni naturali o artificiali); (ii) ogni componente è interconnesso con un numero elevato di altri componenti; (iii) le connessioni non sono rigide ma plastiche, modificabili con adeguati processi di apprendimento, grazie alla interazione con il mondo esterno o con un opportuno "insegnante". In particolare il processo di apprendimento rinforza selettivamente certe connessioni o crea nuove connessioni, ne indebolisce o elimina altre, in funzione delle prestazioni conseguite ; (iv) non è necessario che un neurone abbia comportamenti complessi, bastando operazioni semplici, come l'esecuzione di somme pesate e decisioni di soglia.

Le reti neuronali sono costituite da unità elementari, chiamate neuroni, interconnesse tra loro ed organizzate in uno o più strati. Ciascun neurone (o nodo) della rete è un processore che riceve degli input e fornisce degli output. L'input ad ogni nodo interno è dato dalla somma pesata dei valori delle connessioni con tutti i neuroni che lo precedono, mentre l'output è ottenuto dall'elaborazione dell'input mediante una funzione di attivazione e trasmesso ai neuroni che lo seguono attraverso una funzione di trasferimento.

In teoria, per apprendere i sistemi più complessi le reti dovrebbero avere più strati, in realtà è dimostrato che una struttura a tre strati (uno di input, uno di neuroni nascosti e uno di output) è sufficiente a definire situazioni molto complesse e può essere, quindi, usata come modello in casi dotati di notevole variabilità nelle caratteristiche degli input e degli output (Marija Boznar et Al., 1993).

4.1. Architettura delle reti neuronali artificiali e loro funzionamento

L'architettura di una rete neuronale è il particolare modo in cui le sue parti elementari (i neuroni) vengono organizzate ed interconnesse. L'architettura *"feed-forward"* è la più semplice e maggiormente utilizzata ed è costituita da più di due strati di neuroni: allo strato di input e a quello di output sono aggiunti uno o più strati intermedi. Ogni neurone di uno strato è connesso a tutte le unità dello strato precedente, ma non ha collegamenti con neuroni dello stesso strato. Il segnale si propaga, unidirezionalmente, dall'input all'output, attraverso la gerarchia degli strati intermedi.

La figura 4.1 dà un esempio di rete neuronale *"feed- forward"*.

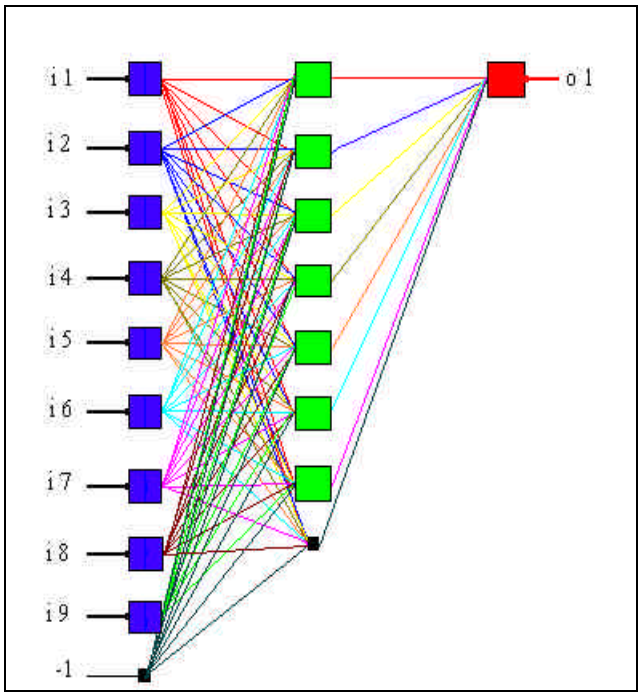

Fig. 4.1. *Esempio di rete neuronale feed-forward.*

Il primo strato contiene le unità d'ingresso, il secondo strato contiene le unità "nascoste" che elaborano l'informazione e la trasmetto alle unità successive di uscita.

Ogni rete neuronale, indipendentemente dalla sua organizzazione strutturale, prevede che ogni neurone sia dotato di alcune proprietà (Cammarata, 1990**)**:

- all'istante t si trovi in uno stato s_i(t);
- abbia una soglia di eccitazione q_i ;
- possa essere stimolato da altri neuroni *j*, ai quali è connesso con intensità w_{ij} (peso sinattico).

Una rete neuronale stabilisce delle relazioni tra un vettore di ingresso ed un vettore di uscita. La dinamica del sistema è rappresentata da due leggi: una legge di attivazione, che aggiorna gli stati dei neuroni; una legge (o algoritmo) di apprendimento, che modifica i pesi delle connessioni.

La legge di attivazione

La legge di attivazione è responsabile dell'aggiornamento dello stato dell'i-esimo neurone al passaggio da un istante temporale al successivo:

$$s_i(t) \rightarrow s_i(t+1)$$

L'input di stimolazione del neurone *i* dello strato σ di una rete *feed-forward* a n strati, è dato dal potenziale P_i^s :

$$P_i^{\sigma} = \sum_j \left[\; w_{ij} \cdot s_j^{\,s-1}(t) - q_i \; \right], \qquad \text{con } \sigma = 2,..., n$$

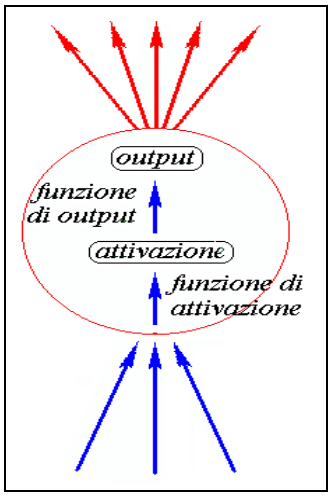

Fig. 4.2. *Funzioni di un neurone.*

La soglia q_i viene talvolta eliminata, aggiungendo un input fittizio s_k=1 e assegnando alla relativa connessione un peso :

$$w_{ik} = - q_i$$

In tal caso il potenziale diventa semplicemente:

$$P_i^s = \sum_j \left[\; W_{ij} \cdot s_j^{\,s-1}(t) \; \right] \qquad \text{con } \sigma = 2,..., n$$

Lo stato successivo del neurone *i*, s_i(t+1), viene calcolato tramite una opportuna funzione del potenziale P_i, denominata funzione di attivazione o di trasferimento :

$$s_i^{\sigma} = f(P_i^{\sigma})$$

La funzione *f* può assumere diverse forme ma quella più comunemente usata è la *funzione logistica,* la quale ha forma sigmoide ed è caratterizzata da valori continui e da saturazione:

$$s_i(t+1) = \frac{1}{1+e^{-P_i^s}}$$

L'addestramento

Nella fase di addestramento la rete neuronale parte da uno stato iniziale, caratterizzato dalla assegnazione di valori arbitrari dei pesi sinattici, ed evolve dinamicamente verso uno stato finale di equilibrio al quale corrisponde l'apprendimento del problema in questione.

Nell'ambito del paradigma connessionista, l'apprendimento assume un'importanza fondamentale: non essendo possibile, in generale, stabilire preventivamente i pesi delle connessioni tra neuroni, in funzione del compito che la rete neuronale deve svolgere. Tali pesi devono essere appresi e la rete neuronale deve comportarsi come un sistema adattivo (Patarnello, 1991).

La fase di apprendimento è generalmente molto lunga e richiede la presenza di un set di valori del vettore di input dotati di output conosciuto (addestramento *supervised*). Questo set di dati costituisce il set di esempi (*training set*) che la rete neuronale dovrà imparare ad emulare.

L'aggiustamento dei pesi delle connessioni avviene attraverso l'uso di procedimenti iterativi. Di questi i più comunemente impiegati sono di tipo retroattivo. Essi consistono nel confronto tra i valori calcolati dalla rete con quelli da emulare ed in una conseguente modifica dei pesi delle connessioni in modo da minimizzare lo scarto di tale confronto. Questo procedimento viene ripetuto su ogni esempio, cioè su ogni vettore del *training set*. Terminata la presentazione degli esempi è terminato un ciclo di addestramento. Affinché la rete neuronale raggiunga l'assetto finale, un solo ciclo di addestramento non basta: sono necessarie molte iterazioni in modo da raggiungere un'approssimazione soddisfacente.

Il procedimento iterativo più conosciuto e maggiormente utilizzato è l'*error back propagation* (EBP) (un caso particolare di algoritmo con supervisore: *supervised-learning*). Si tratta di un algoritmo relativamente poco sofisticato, ma che ha avuto particolare

successo in molte applicazioni, a causa della sua particolare abilità nel conferire alla rete neuronale addestrata una buona capacità di generalizzazione (Logical Designs, 1996).

In una rete *feed-forward*, l'algoritmo *EBP* aggiorna i pesi delle connessioni minimizzando l'errore totale $E_{k,}$ per ogni esempio k presentato in input (Cammarata, 1990):

$$E_k = \sum_j (y_j - \hat{y}_j)^2$$

dove y_j e $\hat{y}_j$ sono rispettivamente l'output calcolato e quello desiderato.

Per effettuare ciò, l'*EBP* prevede due fasi. Nella prima, detta *forward propagation*, si calcolano, per ogni neurone *j* dello strato σ (σ = 2, ..., n) il potenziale di attivazione P_j^{σ} e lo stato di attivazione s_j^{σ}, fino ad arrivare allo strato σ = n.

Nella seconda, detta *backward propagation*, si calcolano gli aggiornamenti dei pesi (da usarsi nella successiva fase *forward*), nell'intento di diminuire l'errore $E = E_k$, procedendo all'indietro, cioè dallo strato σ = n fino allo strato σ = 1. Nell'attuazione di questa fase si procede a molti cicli di presentazione, fino a quando l'errore quadratico medio su tutto il *training set* :

$$E_m = \sqrt{\frac{1}{m} \cdot \sum_{k=1}^{m} E_k}$$

non scende al di sotto di un valore prefissato $\tilde{E}$, idealmente uguale a zero.

L'aggiornamento dei pesi avviene secondo i seguenti calcoli. Per ogni neurone *j* dello strato di output si calcolano :

$$d_j^n = (o_j^n - \hat{o}_j^n) \cdot f'(P_j^n)$$

$$\Delta w_{ji} = -h \cdot d_j^n \cdot s_i^{(n-1)}$$

dove o_j^n e d_j^n sono rispettivamente lo j-esimo output prodotto dalla rete e lo j-esimo output desiderato. Nella seconda formula, che rappresenta l'aggiornamento dei pesi, η è il coefficiente di apprendimento, che consente di regolare la sensibilità dell'algoritmo allo scostamento tra valore prodotto dalla rete e valore desiderato.

Risalendo all'indietro, per ogni strato σ (σ = n-1, n-2, ..., 2) e per ogni neurone *j* del singolo strato, si calcolano:

$$d_j^s = f'(P_j^s) \cdot \sum_r (d_r^{(s+1)} \cdot w_{rj})$$

$$\Delta w_{ji} = -h \cdot d_j^s \cdot s_j^{(s-1)}$$

Gli errori vengono retropropagati attraverso il termine $d_r^{(s+1)}$, partendo dallo strato σ = n, fino ad arrivare a quello d'ingresso (σ = 1).

L'algoritmo di retropropagazione dell'errore trova spazio in numerose applicazioni per la sua capacità di generalizzazione, ma presenta alcuni punti deboli (Cammarata, 1990). Il primo è la lentezza del processo di apprendimento, che richiede spesso un grande numero di cicli prima di conseguire un errore globale sufficientemente piccolo. L'algoritmo *EBP* può essere comunque accelerato con la seguente modifica della legge di aggiornamento dei pesi:

$$w_{ji}(t+1) = w_{ji}(t) - \Delta w_{ji} + b \cdot (w_{ji}(t) - w_{ji}(t-1))$$

dove Δw_{ji} è la variazione dei pesi calcolata con la formula precedente, mentre β è una costante positiva < 1, detta *momentum*. Il termine ora aggiunto, il cui effetto è controllato dal valore assegnato a β, rappresenta una specie di "ricordo" dell'aggiornamento precedente e contribuiscea ridurre le variazioni brusche dei pesi.

Un altro potenziale inconveniente delle reti *feed-forward* ad unità nascoste è la presenza di minimi relativi della funzione d'errore quadratico medio.

La funzione d'errore, infatti, può essere considerata come una funzione dell'energia del sistema: allo stato di minima energia corrisponde lo stato di equilibrio dello stesso. La funzione dell'errore, quando la funzione di trasferimento (o di attivazione) è lineare è rappresentata da una superficie a forma di iperparaboloide dotata di un minimo assoluto. L'apprendimento corrisponde al raggiungimento della configurazione di minima energia ed è garantito, in quanto l'iperparaboloide è dotata di un solo minimo assoluto. Quando la funzione di attivazione è non lineare, ma monotona come la funzione sigmoide, essa è rappresentata da un'iperparaboloide deformata, caratterizzata dalla presenza di minimi relativi. Esiste quindi il pericolo che, durante la fase di apprendimento si rimanga intrappolati in un minimo relativo assestandosi in una configurazione che impedisce il pieno sviluppo delle capacità neuronali.

4.2. Proprietà delle reti neuronali

Le reti neuronali si distinguono particolarmente, in quanto dotate della proprietà di **approssimazione universale**, cioè della capacità di approssimare la legge che descrive un fenomeno. I teoremi, frutto della ricerca di Hornik, Stinchcombe e White (Hornik K., 1991) stabiliscono che reti neuronali *feed-forward* con un solo strato nascosto possono approssimare qualunque funzione continua, indipendentemente dalla funzione di attivazione e dalla dimensione *r* dello spazio degli input.

I teoremi appena citati sono importanti soprattutto dal punto di vista della implementazione; tali teoremi, infatti, ci permettono di codificare reti neuronali relativamente semplici e quindi di ridurre i tempi di stima del modello.

Una ulteriore proprietà delle reti neuronali è quella della **generalizzazione** ovvero la capacità della rete neuronale di effettuare la previsione su dati diversi da quelli utilizzati per il suo addestramento.

Quest'ultima è un'importante caratteristica, in quanto se la rete neuronale si limitasse a fornire risposte corrette Y ai soli input X degli esempi (X, Y) utilizzati nell'addestramento, essa si limiterebbe ad "apprendere a memoria" (Cammarata, 1990). Il vero apprendimento, invece, dovrebbe offrire alla rete capacità previsionali, consentendo quindi risposte Y sostanzialmente corrette anche agli input X non compresi nel *training set* e appartenenti generalmente ad un opportuno *validation set*. La probabilità che una rete neuronale *feed-*