

di Giovanni Pieri

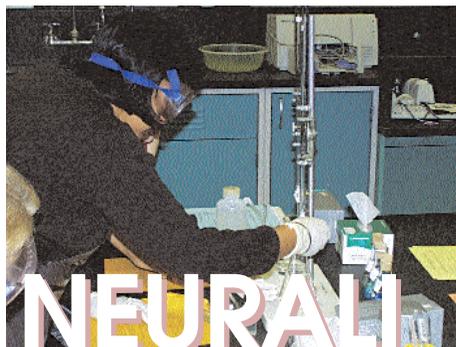
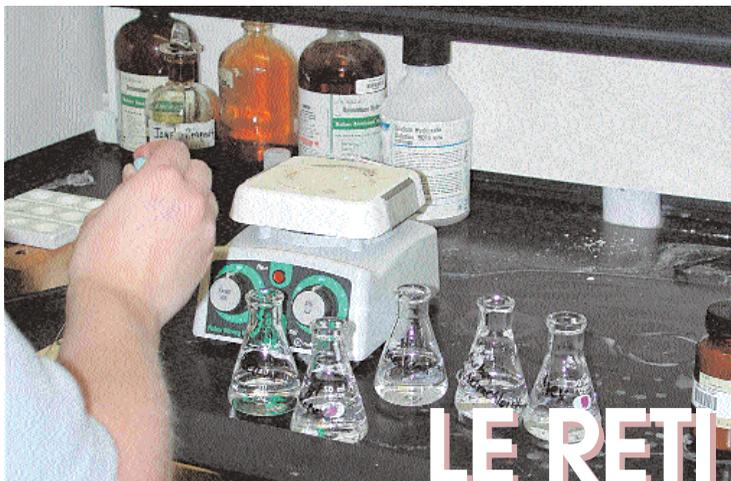
Comitato Ricerca e Innovazione di Federchimica

Massimo Buscema, Vittorio Carlei

Semeion - Trigoria (Roma)

Vittorio Maglia

Direzione Analisi Economiche - Internazionalizzazione di Federchimica



LE RETI NEURALI

Parte 2: Un moderno strumento matematico applicato ai fabbisogni formativi dell'industria chimica

In questa seconda parte dell'articolo viene riportato un sommario della classificazione descritta nella parte prima ed una stima della richiesta totale di laureati da parte dell'industria chimica italiana, eseguita impiegando una rete neurale a ricircolazione (Nrc, New Recirculation).

Nella prima parte dell'articolo è stata riportata una classificazione delle risposte ad un questionario che Federchimica ha proposto ai propri associati per accertare le loro necessità in materia di formazione universitaria. Tale elaborazione è stata eseguita impiegando un particolare tipo di Reti Neurali Artificiali (Rna; in inglese Ann, Artificial Neural Networks), detto Som (Self Organising Map, o mappa autorganizzata). In questa seconda parte una stima della richiesta totale di laureati da parte dell'industria chimica italiana viene eseguita impiegando una rete neurale a ricircolazione (Nrc, New Recirculation), che permette di estrapolare i bisogni globali da un campione comprendente meno del 10% di tutta la popolazione. I dati così ottenuti sono poi impiegati per dare una valutazione orienta-

tiva del bilancio tra domanda e offerta di laureati. Si ritiene che l'esempio interessi i lettori della rivista non solo nel merito, ma anche nel metodo. Se da un lato è centrale che la preparazione dei laureati in materie chimiche si adegui ai cambiamenti subiti dall'industria chimica in Italia, dall'altro il metodo delle reti neurali potrebbe ispirare impieghi tecnici, per esempio nell'elaborazione di dati sperimentali.

Sintesi della classificazione delle risposte al questionario per mezzo di una Som

Nella parte prima è stata impiegata una Som (mappa autorganizzata) per classificare le risposte al questionario che Federchimica ha introdotto tra gli associati per valutare i loro bisogni in materia di formazione universitaria. La mappa ha mostrato che la

classificazione avviene principalmente lungo due assi: dall'alto al basso per dimensione aziendale crescente e da sinistra a destra per propensione crescente all'assunzione di laureati. In questa mappa si possono identificare zone specifiche dove vengono classificate aziende appartenenti a ciascuna delle diverse associazioni federate in Federchimica.

In Figura 1 si riporta una sintesi dei vari diagrammi riportati nella parte prima, ove si cerca di rappresentare le correlazioni tra dimensione, associazione di appartenenza, propensione all'assunzione e tipo di laurea. La denominazione "altre lauree" indica lauree di indirizzo non chimico, principalmente Economia e Agraria. La classificazione ha un valore orientativo sia per lo studente al momento della scelta del corso di laurea, sia per il neolaureato in cerca di un'occu-

pazione, perché permette di determinare con una certa approssimazione la tipologia di azienda, il ramo merceologico e il contenuto professionale del lavoro offerto in funzione del tipo di laurea.

Estrapolazione dell'assunzione di laureati dal campione a tutta la popolazione

Il problema della stima della propensione all'assunzione di chi non ha risposto al questionario non è semplice perché il campione è piccolo rispetto a tutta la popolazione (103 aziende hanno risposto al questionario su 1.500 totali in Italia) e perché le caratteristiche medie del campione differiscono molto da quelle della popolazione (per esempio il numero medio di addetti per le aziende del campione = 280 e il numero medio della popolazione = 80). Per risolvere il problema si è fatto ricorso ad un tipo di

goritmo esamina le relazioni tra loro senza che qualche scelta esterna le abbia previamente associate ad un tipo o all'altro.

Le reti neurali autoassociate sono costituite da nodi e da connessioni. I nodi sono disposti in due strati, connessi in tutti i modi possibili. Ad ogni connessione è associato un peso (la misura della forza della connessione). Perciò se in una rete autoassociata esistono N nodi di input, la matrice dei pesi sarà costituita da N² connessioni. I pesi vengono calcolati per mezzo di una procedura matematica, chiamata addestramento, al termine della quale i pesi rappresentano l'informazione contenuta nei dati. Dei due strati quello di input è chiamato visibile perché accessibile a chi usa la rete e quello di output è chiamato nascosto.

La Nrc elaborata dal Semeion prevede un'architettura a quattro livelli autoassociati. Il primo strato dei nodi (visibile) si connet-

Definiamo il vettore di input a_i^R come input visibile reale (che, come già detto, comprende tutte le variabili). A partire da a_i^R è possibile in un primo tempo calcolare un output nascosto reale a_o^R . In un secondo tempo, retroagendo dall'output all'input, è possibile calcolare, a partire da a_o^R , un input visibile ipotetico, a_i^I . In un terzo tempo, a partire da a_i^I , è possibile calcolare un output nascosto ipotetico a_o^I .

In Figura 2 è rappresentato il sistema di connessioni:

La dinamica di trasferimento del segnale in una Nrc avviene, così, in tre tempi:

- t_1 : input visibile reale (a_i^R)	→	output nascosto reale (a_o^R)
- t_2 : output nascosto reale (a_o^R)	→	input visibile ipotetico (a_i^I)
- t_3 : input visibile ipotetico (a_i^I)	→	output nascosto ipotetico (a_o^I)

L'apprendimento avviene quando l'input visibile ipotetico (a_i^I) riproduce ogni input

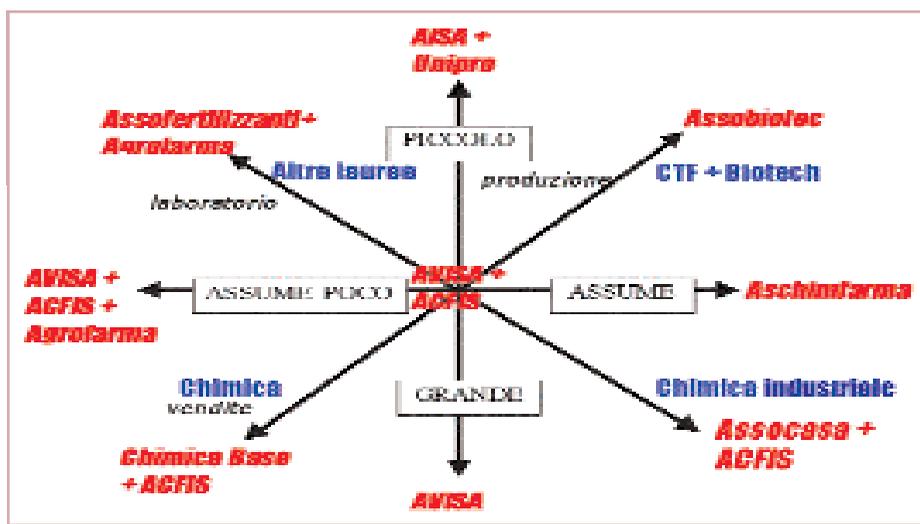


Figura 1 - Sintesi grafica delle relazioni tra variabili principali nella risposta al questionario

rete neurale denominato Nrc (New Recirculation), questo tipo è una variante, inventata e brevettata da Semeion, della rete autoassociata.

La denominazione di autoassociata deriva dal fatto che tra le variabili non viene mantenuta la classica distinzione tra variabili indipendenti (quelle che descrivono le cause e sono degli input per l'algoritmo) e variabili dipendenti (quelle che descrivono gli effetti e per l'algoritmo figurano come output). Le variabili sono tutte in input e l'al-

te con il secondo strato (di output e nascosto). Al secondo è connesso un terzo strato ancora di input (e quindi visibile), ma che rappresenta un input ipotetico, a sua volta lo strato è connesso ad un quarto strato di output, anch'esso nascosto e ipotetico. I pesi delle connessioni tra il primo ed il secondo strato, il secondo e il terzo, il terzo e il quarto, sono gli stessi. Essi vengono aggiornati durante il processo di apprendimento, ma i tre ordini di connessioni cambiano di conserva.

visibile reale (a_i^R); il che significa che i due vettori di unità nascoste, a_o^I e a_o^R , si sono formati una loro rappresentazione interna del vettore di input a_i^R . Uno schema di queste interazioni è rappresentato in Figura 2. Le connessioni incrociate rappresentano i confronti che si fanno per aggiornare i pesi.

L'algoritmo di apprendimento della Nrc prevede due fasi. Ogni volta che il segnale viene filtrato dall'input reale (a_i^R) all'output ipotetico (a_o^I), vengono operate due correzioni sulla stessa matrice dei pesi: la prima considerando la differenza tra l'input reale e quello ipotetico, la seconda computando la differenza tra l'output reale e quello ipotetico. In questo modo la RC converge dopo un numero molto basso di cicli.

La fase di training prevede quindi la presentazione ripetuta dei record di un database affinché la rete modifichi le sue connessioni, individuando la funzione implicita che codifica le relazioni tra le variabili. Una rete di questo tipo si comporta come un oggetto

di plasmabile argilla, sul quale i dati che compongono l'insieme di training lasciano traccia delle loro relazioni.

La fase dell'interrogazione, attraverso il rientro, permette di immettere nella rete input mai visti, inducendola a fornire una risposta compatibile con le informazioni apprese dal database di addestramento. Si potrebbe dire che in questa fase il materiale argilloso si solidifica, permettendo di esplorare le forme che si sono definite durante la fase di addestramento.

L'interrogazione di una rete New Recirculation da parte di un'utente (fase di Recall) corrisponde alla selezione di una o più variabili, lasciando non specificato il valore delle altre variabili. Il processo di ricircolazione completerà le variabili non specificate in coerenza con le variabili specificate.

Nel questionario di Federchimica è stato

guita una validazione del metodo: il data base delle 103 risposte è stato suddiviso con scelta casuale in due parti circa uguali: su una è stato eseguito un addestramento specifico e per l'altra sono state ricostruite le domande a partire dei due soli parametri associazione di appartenenza e numero di addetti. Questa ricostruzione ha permesso di valutare l'accuratezza percentuale della ricostruzione con la seguente formula:

$$\text{accuratezza} = 100 \left(\frac{\text{valore previsto} - \text{valore vero}}{\text{valore vero}} \right)$$

L'accuratezza media è risultata del 99,01% per la ricostruzione delle assunzioni di triennali e 99,11% per la ricostruzione delle lauree quinquennali.

Il totale della assunzioni previste con la Nrc è 750 laureati/anno, pari a 0,5 assunzioni annue su 100 addetti e conferma il valore di 685 assunzioni annue ottenuto con una elaborazione statistica indipendente.

indiretta. Un primo ragionamento può essere impostato confrontando la previsione di nuove assunzioni annue nel prossimo triennio con il turn-over naturale dei laureati di tipo chimico presenti nell'industria chimica. L'industria chimica conta circa 150.000 addetti di cui circa 12.000 laureati chimici. Il tasso di turn-over si può stimare in circa 3% annuo, corrispondente ad una vita lavorativa media di circa 35 anni. Con tale tasso di turn-over i laureati richiesti solo per rimpiazzare gli uscenti, sono circa 350. Le richieste emerse dal questionario sono più o meno il doppio del tasso di turn-over. Tenuto presente che il questionario registra intenti e non decisioni già prese e non è detto che l'eccesso delle previsioni sul turn-over si realizzi, vediamo di procedere ad un bilancio tra domanda e offerta nei limiti di precisione consentiti dai dati a disposizione.

Nel periodo che ci interessa (il prossimo triennio) il numero di laureati immessi sul mercato dalle università italiane si può stimare dalla

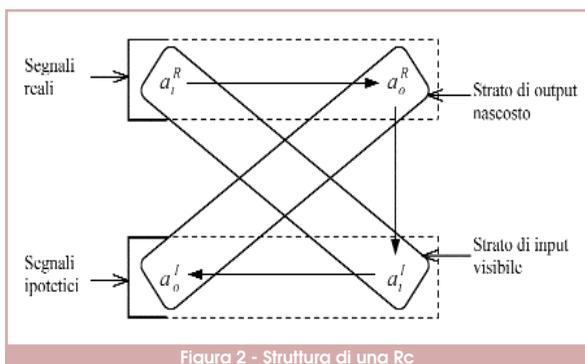


Figura 2 - Struttura di una Rc

quindi sufficiente introdurre l'associazione di appartenenza e il numero di addetti, dati di cui si dispone per ogni associato, per ottenere la sua probabile risposta al questionario, in particolare in merito al numero di assunzioni che l'impresa intende effettuare nel prossimo triennio.

Previsione delle assunzioni annue con la Nrc

La previsione è stata eseguita con la procedura descritta al precedente paragrafo. I risultati cumulati, validi quindi per tutta la popolazione di aziende federate in Federchimica, sono riportati nella Tabella a lato.

L'accuratezza con cui sono stati ottenuti questi risultati è elevata. Infatti, è stata ese-

Bilancio tra le assunzioni e l'offerta da parte dell'università

È diffusa la sensazione, preesistente al questionario, che si possa verificare un deficit nel prossimo triennio, in considerazione del declino degli iscritti alle facoltà di indirizzo chimico, verificatosi negli ultimi anni. La domanda, cui non è facile rispondere, è se i risultati del questionario rafforzano questo punto di vista. Il questionario non è stato disegnato per rispondere direttamente al quesito, però possiamo tentare una stima

media delle iscrizioni annue nel periodo precedente. In questo modo si ottiene che in un anno le facoltà di chimica e chimica industriale immetteranno sul mercato del lavoro circa 800 nuovi laureati, mentre le facoltà di ingegneria chimica faranno altrettanto con 500 nuovi ingegneri per un totale di 1.300 laureati. I 700 circa richiesti dall'industria chimica rappresentano circa il 55% di tale offerta di cui il 27-28% necessario per equilibrare il turn over.

Storicamente è noto che l'industria chimica

Previsione dei laureati totali assunti in un anno

Tipo di laurea	Assunzioni unità/anno	%	Assunzioni unità/anno/100 addetti
Triennale	180	24	0,12
Quinquennale	570	76	0,38
Totale	750	100	0,50

assorbe una quota del 35-40% circa del totale dei laureati di tipo chimico, quota non molto superiore a quella del turn-over. La restante quota maggioritaria (60-65%) non s'impiega nell'industria chimica. In primo luogo, per circa il 30%, s'impiega in altre industrie di tipo non chimico, tra queste l'alimentare, la farmaceutica, la meccanica, le società d'ingegneria. Il restante 30% segue altre strade non industriali: la docenza universitaria, l'insegnamento in scuole secondarie, l'impiego in strutture statali o parastatali come le unità sanitarie locali, le agenzie ambientali e simili. Se queste percentuali si mantenessero anche nel prossimo triennio, resterebbe a disposizione dell'industria chimica un massimo di 520 laureati all'anno, con uno sbilancio negativo di circa 230 unità annue. Quanto è grave questo sbilancio? Non ci sono elementi certi per una valutazione, ma possiamo osservare che tre meccanismi diversi dovrebbero intervenire per mitigarlo:

- l'assunzione di laureati in CTF, ingegneri

gestionali e ingegneri dei materiali in parziale alternativa ai laureati di tipo chimico;

- lo spostamento della preferenza dei laureati in cerca di impiego dalle altre industrie e occupazioni verso l'industria chimica, in altre parole un aumento sopra il 40% della percentuale storica di impiego nell'industria chimica;

- la comprimibilità, senza traumi per le aziende chimiche, della propensione, rilevata con il questionario, ad assumere più del turn-over.

Il primo meccanismo è sicuramente attivo ed è confermato dalla classificazione con la Som in base alle lauree desiderate, che vede un certo tipo di aziende chimiche propense ad assumere laureati in corsi di laurea alternativi. Non vi sono dati per quantificare la richiesta, ma, dati i piccoli numeri dell'industria chimica e i grandi numeri delle

lauree alternative, un eventuale spostamento verso l'industria chimica non metterebbe sicuramente in tensione il mercato del lavoro riguardante quei tipi di laurea.

Impossibile è la quantificazione per gli altri due meccanismi. Specialmente la comprimibilità della domanda manca di una precedente esperienza su cui basare la stima. Peraltro, se anche questi meccanismi potessero contribuire solo per alcune decine d'unità ciascuno, provocherebbero una riduzione consistente dello sbilancio negativo e della tensione sul mercato. Non resta quindi che attendere e monitorare il mercato per cogliere tempestivamente eventuali segnali indicativi di una domanda insoddisfatta.

Conclusioni

Gli algoritmi matematici chiamati reti neurali artificiali o Rna sono stati applicati all'elaborazione di un questionario diffuso tra gli

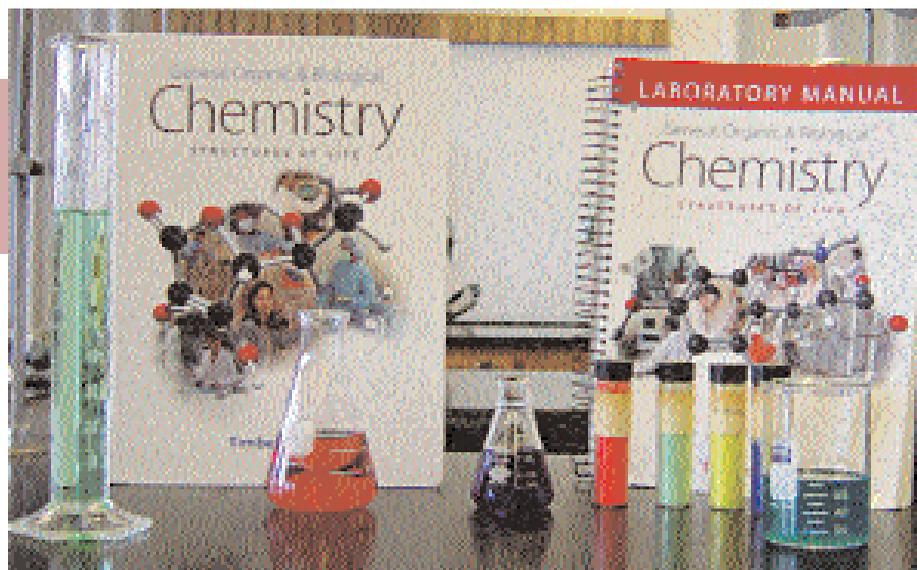
mente il modo di rispondere al questionario;

- l'appartenenza all'una o all'altra delle associazioni di Federchimica fa differenza nel modo di rispondere. Solo tre su 10 delle associazioni sono poco caratterizzate e condividono il modo di rispondere con altre associazioni;

- la propensione alle assunzioni di laureati è maggiore per quelle associazioni che in qualche modo preparano specialità rispetto a quelle che si occupano di chimica di base o materie ad essa affini;

- la propensione ad assumere chimici, chimici industriali ed ingegneri chimici cresce al crescere della dimensione aziendale, mentre a dimensioni più piccole si manifesta una tendenza ad assumere laureati in CTF, biotecnologie ed in facoltà non chimiche;

- il fabbisogno di laureati dell'industria chimica per il prossimo triennio è al massimo 750 unità annue;



associati a Federchimica riguardo alle necessità di formazione universitaria. Le Rna hanno permesso di classificare le risposte al questionario e di ricostruire attendibilmente la propensione ad assumere laureati delle aziende che non hanno risposto al questionario, a partire da due soli parametri: associazione di appartenenza e numero di addetti. I risultati più importanti così ottenuti sono i seguenti:

- la dimensione aziendale influenza forte-

- questa richiesta comporta un possibile deficit rispetto a quanto viene attualmente offerto. Questo deficit può essere almeno in parte colmato, mitigando l'eventuale tensione sul mercato, da assunzioni di laureati in CTF ed altre lauree, da un parziale ritorno verso l'industria chimica di chi in passato preferiva impieghi al di fuori di essa e da una certa comprimibilità della domanda, di cui peraltro manca sufficiente esperienza per tentarne una stima.