



**UNIVERSITÀ DEGLI STUDI DI NAPOLI
F E D E R I C O I I**

FACOLTÀ DI INGEGNERIA

***DIPARTIMENTO DI INGEGNERIA DEI TRASPORTI
“L.TOCCHETTI ”***

**DOTTORATO DI RICERCA IN INGEGNERIA DEI SISTEMI
IDRAULICI, DI TRASPORTO E TERRITORIALI**

-CICLO XX-

**UNA NUOVA METODOLOGIA D’USO DELLE RETI
NEURALI ARTIFICIALI PER L’AGGIORNAMENTO
DELLA DOMANDA DI TRASPORTO**

DI

VALERIO DE MARTINIS

**IL TUTOR
PROF. ING. LUIGI BIGGIERO**

**IL COORDINATORE DEL DOTTORATO
PROF. ING. BRUNO MONTELLA**

ANNO ACCADEMICO 2007-2008

1	INTRODUZIONE.....	3
2	LE RETI NEURALI ARTIFICIALI (ANN) APPLICATI AI SISTEMI DI TRASPORTO.....	8
2.1	Cenni sulle reti neurali artificiali	12
2.2	Le tipologie di reti neurali artificiali utilizzate	16
2.3	Ultime applicazioni nel campo della previsione dei flussi veicolari.....	30
2.4	La logica fuzzy nelle reti neurali artificiali.....	36
3	LA METODOLOGIA APPLICATA	40
3.1	La matrice Origine Destinazione.....	43
3.2	Il metodo Montecarlo	53
3.3	Test d’ipotesi T-Student.....	57
3.4	Creazione del training set	62
4	APPLICAZIONI A DATI REALI	71
4.1	Le reti neurali artificiali utilizzate	79
4.1.1	Multi Layer Perceptron.....	80
4.1.2	Radial Basis Function Network	82
4.1.3	Generalized Regression Network	84
4.2	Confronto con la metodologia tradizionale	85
4.3	Confronto delle prestazioni di ANN su diversi test site.....	91
5	CONCLUSIONI.....	98
6	BIBLIOGRAFIA	101

“...il compito degli uomini di cultura è più che mai oggi quello di seminare dei dubbi, non già di raccogliere certezze. Di certezze rivestite della fastosità del mito o edificate con la pietra dura del dogma sono piene, rigurgitanti le cronache della pseudo cultura degli improvvisatori, dei dilettanti, dei propagandisti interessati. Cultura significa misura, ponderatezza, circospezione, significa valutare tutti gli argomenti prima di pronunciarsi, controllare tutte le testimonianze prima di decidere e non pronunciarsi e non decidere mai a guisa di oracolo dal quale dipenda in modo irrevocabile una scelta perentoria e definitiva.”

(Norberto Bobbio)

1 Introduzione

Dalla metà del secolo scorso si è assistito allo sviluppo sempre crescente delle aree urbane e al conseguente aumento della domanda di mobilità, dovuto in gran parte alla motorizzazione e a nuovi e sempre più efficienti sistemi di trasporto, e della congestione. Questo trend in decisa crescita ha spinto l’ingegneria dei trasporti a sviluppare nuove politiche di controllo e nuovi modelli applicativi, per un controllo sempre più attivo del traffico urbano. La rete stradale delle città, usualmente considerata congestionata, è quindi oggetto di molteplici esperienze di ricerca, in larga parte finalizzate ad un’ottimizzazione sempre più precisa e flessibile delle possibili politiche di controllo e gestione, in modo da “personalizzare” le scelte nello spazio e nel tempo. Pertanto, si è posta, e tutt’oggi si pone, molta attenzione su modelli e metodologie sempre più affidabili e modulabili rivolte al controllo in tempo reale dei flussi veicolari,

in particolare mediante la creazione di modelli dinamici previsionali che, generalmente, rispondono al sistema sulla base di informazioni, conoscenze e considerazioni, acquisite o note a priori. Queste informazioni sono essenzialmente la domanda di mobilità nota, rappresentata dalla matrice Origine Destinazione, e il flusso veicolare sugli archi stradali, ricavato dai conteggi di flusso di traffico.

In letteratura esistono diversi metodi per determinare la matrice Origine-Destinazione a partire dai conteggi di flusso, e tra questi, i modelli che adottano le reti neurali artificiali che risultano essere interessanti e all’avanguardia, anche se ancora poco sviluppati, soprattutto per la loro rapidità di risposta, oltre che dotati una buona approssimazione, più che sufficiente per le applicazioni odierne.

I modelli di rete neurale non sono di facile applicazione nel campo dei trasporti in quanto necessitano di una numerosità di dati consistente che spesso non è possibile ottenere se non a costo di un notevole impiego di risorse economiche e di tempo.

Per definire una rete neurale sono, infatti, necessari un notevole numero di osservazioni di flusso e domanda. Per quanto riguarda i primi è possibile ottenere una buona base dati attraverso l’installazione di rilevatori automatici di flusso per un periodo di tempo sufficientemente lungo. La domanda, invece, è di difficile stima ed è praticamente impossibile rilevarla con continuità. Per questo motivo, le applicazioni delle reti neurali artificiali hanno riguardato quei casi di controllo nei quali la mancanza di domanda interna all’area di studio permette di prescindere dalla valutazione della domanda che

coincide, nei fatti, con il flusso rilevato. Tipici sono i casi di controllo delle intersezioni, roatorie, brevi tronchi stradali, etc.

Il presente elaborato si inserisce nel filone di ricerca sull’utilizzo delle reti neurali per il controllo dinamico (in tempo reale) del traffico urbano. Il contributo innovativo risiede nella possibilità di applicare la procedura su reti di grandi dimensioni congruentemente con i modelli di simulazione della domanda di trasporto.

Oltre ai già citati limiti di reperimento dei dati, si fa presente anche che i moderni processori e i programmi di gestione attuali non sono ancora pronti per “accogliere” la struttura di calcolo parallelo delle reti neurali. L’architettura seriale dei moderni processori oggi in commercio, infatti, permette solo di simulare una rete neurale, e va facilmente in crisi nei processi computazionali che richiedono una rete neurale di grandi dimensioni. Inoltre, per dati fortemente eterogenei, è spesso conveniente integrare la logica delle reti neurali con la moderna logica fuzzy, così da perfezionarne le capacità predittive, e definire in modo più “ingegneristico” i limiti dei valori dell’insieme di apprendimento.

La necessità di sistemi computazionali ad architettura parallela sarà soddisfatta in tempi non molto lunghi, visto che questo è il trend che le case costruttrici seguono, ed inoltre è comunque possibile realizzare una rete di elaboratori per la simulazione di detta architettura. Le applicazioni attuali, però, devono necessariamente dimensionarsi sugli elaboratori a processori seriali che costituiscono, di fatto, la quasi totalità del mercato.

Sotto questo aspetto, la procedura proposta semplifica molto l’applicazione ed il progetto stesso della rete neurale, rendendola compatibile con le applicazioni richieste su reti di grandi dimensioni.

Le esperienze riportate in letteratura fanno rilevare che l’associazione di input di uno o più ordini di grandezza inferiore all’output, come nel caso specifico il vettore flussi che di solito è nell’ordine di qualche decina di elementi rispetto al vettore matrice che facilmente raggiunge il migliaio di elementi, non è di facile soluzione e necessita di un numero di dati dedicati all’apprendimento dell’ordine delle decine di migliaia di unità. Inoltre il lavoro necessario per approntare una rete che possa rispondere a queste esigenze necessita di molto tempo e risorse, e non risponde velocemente come le si richiede; si parla difatti di problema “mal posto”. La scelta, introdotta nel presente lavoro, di associare al vettore dei flussi un vettore di parametri caratteristici della domanda velocizza di molto la velocità di risposta da parte della rete, senza la quale verrebbe meno uno dei presupposti di utilizzo di tale sistema, ossia la risposta in tempo reale.

Un altro contributo innovativo di questo lavoro di tesi riguarda la procedura di apprendimento della rete neurale. Questa è un sistema di calcolo parallelo che simula il processo di apprendimento del cervello umano. Per fare in modo che la rete risponda ad un dato problema, è necessario presentarle dapprima una serie di dati di apprendimento (training set), in modo tale che la rete “apprenda” l’associazione che in seguito dovrà

fare tra i dati di input e quelli di output senza l’esplicitazione di formulazioni matematiche.

Nella letteratura corrente, infatti, si ricavano una serie di matrici OD modificando percentualmente i valori di una matrice nota a priori, e ricavando i flussi sulla rete di trasporto mediante l’assegnazione di tali matrici creando, nei fatti, una collinearità dei dati che potrebbe far rispondere la rete in modo non corretto.

La metodologia presentata in questo lavoro, si propone di creare il training set, e quindi le matrici, a partire dai parametri dei modelli di domanda della matrice nota a priori e di estrarre in maniera casuale tali parametri secondo la loro distribuzione statistica.

I dati così creati non hanno alcun legame tra loro se non il fatto che discendono da una statistica che fa riferimento ai probabili valori medi di accadimento, evitando quindi la forte correlazione insita nella generazione di dati tutti dipendenti in maniera lineare tra di loro. Inoltre la creazione di dati a partire dai modelli di generazione e distribuzione permette di evitare alcune incongruenze, oggetto di critiche, con la metodologia tradizionale che modifica i singoli valori della matrice OD senza alcun controllo.

Controlli effettuati off-line impongono modifiche che possono inficiare la procedura. È tipico, ad esempio, avere il numero di spostamenti generati da un’origine O tra le 7 e le 8 di mattina superiore al numero di abitanti di detta zona.

Le elaborazioni effettuate tengono ad evidenziare due aspetti: la risposta della rete neurale con diverse tipologie di rete neurale sia nella metodologia tradizionale che nella

metodologia proposta, e la risposta della rete neurale inserita in una ipotetica procedura di previsione dei flussi in senso spaziale su diverse reti urbane usate come test site. Il primo aspetto rivolge l’attenzione esclusivamente alla risposta della rete neurale in termini di vettore dei parametri di domanda, il secondo riguarda la bontà di previsione dei flussi e, ancora, la bontà di risposta della rete neurale artificiale in termini di vettore dei parametri di domanda.

2 Le reti neurali artificiali (ANN) applicati ai sistemi di trasporto

Negli ultimi decenni, sono stati proposti e sviluppati molti modelli di tipo deterministico o stocastico, atti a risolvere problemi complessi nel campo dell’ingegneria dei trasporti. La tradizionale metodologia seguita presuppone l’utilizzo di modelli matematici, formule e complessi sistemi di equazioni, per risolvere questi problemi. I modelli matematici, in generale, sono però espressione di quella che in letteratura viene chiamata spesso “conoscenza oggettiva”; in altre parole, sono rappresentativi di una parte della realtà spiegabile, in maniera più o meno facile, attraverso osservazioni del fenomeno, ipotesi, formule, sistemi di equazioni, verifica delle ipotesi. Tuttavia nel caso di problemi d’ingegneria particolarmente complessi, vi sono una serie di informazioni appartenenti ad una realtà difficilmente spiegabile o inspiegabile, dette “conoscenza soggettiva”, di difficile riproduzione modellistica o di difficile quantificazione. Tali informazioni risultano quindi essere per lo più incognite, nel senso che non si ha

percezione di quanto possano influire sulla bontà di un modello matematico che non tiene conto delle stesse. Il prezzo che si paga per l’utilizzo di precisi modelli matematici per la risoluzione di complicati problemi ingegneristici, è la limitazione nel loro campo d’applicazione, attraverso ipotesi restrittive di utilizzo.

Là dove la modellizzazione della realtà non può, per ragioni fisiche, essere più precisa entra in gioco l’evoluzione tecnologica di sistemi computazionali sempre più potenti.

Nel campo dell’ingegneria dei trasporti, il problema di modellizzare sistemi capaci di prevedere ed elaborare accurate previsioni a breve termine dei parametri di traffico, come flusso, velocità di viaggio e occupazione, è uno dei temi di ricerca più interessanti nell’ambito di sistemi ITS (Intelligent Transportation System). Ad esempio, le tecnologie ITS, come ATIS (Advanced Traveler Information System) e ATMS (Advanced Traffic Management System), ricercano soluzioni ai problemi di congestione del traffico e di tempi di percorrenza, sincronizzando al meglio gli strumenti di gestione del traffico e assistendo i viaggiatori nella scelta del percorso di viaggio, il tutto basandosi su accurate informazioni, ottenute in tempo reale, sulle condizioni di traffico.

In particolare, quello che riguarda la modellizzazione e la predizione a breve termine dei parametri di traffico ha provocato un significativo interesse scientifico che incrementa man mano nel tempo, dovuta al continuo aumentare della velocità computazionale,

conseguenza di una evoluzione esponenziale della tecnologia, e allo sviluppo di nuove e sempre più flessibili metodologie.

Fin dagli anni '80 i ricercatori del settore hanno usato una grande varietà di specificazioni delle caratteristiche dei modelli di traffico per la previsione a breve termine, con una grande produzione di pubblicazioni su predizioni da modelli per intervalli, da ARIMA a modelli lineari generalizzati in campo dinamico fino ad arrivare, specialmente negli ultimi anni, alle Reti Neurali Artificiali (o ANN).

La letteratura corrente ha mostrato che le reti neurali artificiali sono una delle migliori alternative nella modellizzazione e predizione dei parametri di traffico in quanto hanno la capacità di approssimare qualsiasi tipo di funzione, senza tenere conto del suo grado di non linearità e senza una conoscenza a priori della sua forma funzionale. C'è da dire che, per loro stessa natura, le reti neurali artificiali presentano un elevato grado di incertezza nello scegliere la struttura della rete ottimale e nell'implementarla.

Recentemente, in un articolo di *Cantarella, de Luca*¹, relativo all'utilizzo delle reti neurali nella previsione di scelta del modo di trasporto, si distinguevano tre modalità di approccio alla modellizzazione, distinti in base al modo in cui vengono specificate le formule nelle funzioni di modellizzazione e di implementazione:

¹ Giulio Erberto Cantarella, Stefano de Luca “*Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility model*” Transportation Research partC (2005)

White box: le formule derivano da considerazioni teoriche che cercano di spiegare il fenomeno stesso. Questo è il caso in cui i modelli derivano dalla diretta applicazione di leggi fondamentali, quali ad esempio quelli della fisica. In questo caso i parametri del modello hanno, il più delle volte, una chiara interpretazione.

Black box: Le funzioni derivano da considerazioni di tipo empirico. Questo è il caso in cui non esiste alcun paradigma teoretico soddisfacente. I parametri hanno una difficile interpretazione anche se possono essere prese in considerazione una quantità notevole di funzioni; in altre parole non si ha precisa cognizione di ciò che sta avvenendo nel momento in cui vengono applicate queste funzioni in un paradigma di calcolo. Questo è il caso delle Reti Neurali Artificiali.

Grey box: è un qualsiasi tipo di mix dei due tipi sopra menzionati e usualmente utilizzati quando è disponibile solo una parziale definizione teoretica delle funzioni. Questo è generalmente il caso in cui entra in gioco il processo di decision-making dell’ essere umano, rappresentato ad esempio attraverso i modelli di utilità aleatoria (RUM – Random Utility Model)

Su questo tema, le prime esperienze con le reti neurali artificiali hanno avuto come ambito applicativo le intersezioni stradali, così come avvenne per lo studio della domanda mediante i conteggi di flusso, per poi estendersi all’intera rete. Solitamente gli studi fatti propongono un addestramento basato su delle misure di flusso da usare come

input, e su delle informazioni a priori della domanda usate come esempi di output corretti. E’ importante aggiungere che, come molti autori hanno fatto notare nei loro lavori, non sempre un metodo applicabile su un’intersezione lo si può estendere all’intera rete, ma è sicuramente da considerare come un buon punto di partenza per focalizzare i problemi e sviluppare le estensioni modellistiche.

2.1 Cenni sulle reti neurali artificiali

In breve, il concetto alla base di una rete neurale artificiale è quello di riprodurre quello che è il comportamento del cervello umano agli stimoli esterni, il suo processo di apprendimento, la sua capacità di associare ad input anche parziali gli effettivi output desiderati; quindi una rete neurale artificiale, al pari del cervello umano, apprende i legami che esistono tra input e output attraverso degli esempi che le vengono presentati, ed impara a riconoscere nuovi input, mai presentati prima, che rispondono a quei legami. La capacità che una rete neurale ha di rispondere in maniera accurata ad un input è legata al numero di “neuroni” (unità computative virtuali) di cui è fatta ed alle modalità con cui si attivano, dalle connessioni sinaptiche tra neuroni e neuroni e dalla tipologia di esempi presentati. La bontà di risposta di una rete può quindi venire meno a causa della ridondanza degli elementi computazionali elementari costituenti la rete (i neuroni e le connessioni) e quindi parliamo di *sovraparametrizzazione* o *overfitting*, oppure quando si impiega un insieme di esempi ridotto o poco significativo nel corso di lunghe fasi di apprendimento e quindi avremo un *sovrapprendimento* o *overtraining*.

Nell’ultimo decennio, nonostante la consolidata esperienza basata su studi a carattere statistico riguardo la natura spaziale e temporale dei flussi veicolari, c’è stata una grossa produzione di lavori riguardanti l’utilizzo delle reti neurali artificiali nella previsione a breve termine del traffico. Il motivo è dato dalla diversa impostazione nell’implementazione dei modelli rispetto ai convenzionali modelli statistici di previsione, che permette alle reti neurali di auto adattarsi e di catturare le relazioni “nascoste” senza il bisogno di ipotesi a priori, riguardo i problemi esaminati. Inoltre la loro natura non lineare e la loro abilità nel generalizzare, li rende uno strumento molto interessante quando si lavora con database dotati di rumore statistico.

Le principali caratteristiche delle reti neurali artificiali sono

- ***Resistenza al rumore:*** ossia è in grado di continuare a dare una risposta corretta anche se alcune delle sue connessioni vengono eliminate o se viene aggiunto del “rumore” al segnale di ingresso. Quindi anche in presenza di input affetti da disturbo statistico, la rete è in grado di dare una risposta coerente con l’apprendimento che ha ricevuto.

Man mano che il rumore aumenta, le prestazioni della reti neurali artificiali subiscono un decadimento graduale, con un comportamento molto variabile da rete a rete, ma generalmente la percentuale di errore aumenta in modo lento e pressappoco uniforme su tutto il campo di risposta, oppure può perdere la capacità di rispondere correttamente a certi stimoli e mantenere una risposta inalterata per

tutti gli altri. In ogni caso le reti neurali artificiali, come nel caso di sistemi biologici, anche “lesionate” possono essere riaddestrate e riacquistare le capacità perse. La differenza con la modalità di funzionamento dei consueti modelli computazionali basati su sistemi seriali è enorme, se si pensa che la perdita anche di un solo anello della catena di elaborazione, comporta per questi ultimi una catastrofica caduta delle prestazioni del sistema.

- *flessibilità d’uso*: i modelli basati su reti neurali artificiali possono essere impiegati per un gran numero di finalità diverse; infatti una delle caratteristiche più importanti delle reti neurali artificiali, ed uno dei motivi per cui hanno molto appeal nel settore della ricerca, è che non ha bisogno di conoscere le proprietà del dominio specifico di applicazione, perché queste vengono dedotte dall’esperienza accumulata durante l’apprendimento. Questo, da un lato, dà la possibilità di affrontare questioni particolarmente complicate di cui non si conoscono le soluzioni analitiche, dall’altro non dà l’opportunità all’utente di comprendere a fondo la natura del problema, e quindi di “non aumentare la sua conoscenza” (Dewdney, 1993). Inoltre per poter essere utilizzato un modello su reti neurali artificiali, l’utente deve possedere un certo numero di dati “certi”, da usare come input e output di addestramento, nonché avere chiare le finalità del lavoro che si vuole far svolgere alla rete, il tipo di compito ed una serie di vincoli, volti a definire la struttura della rete.

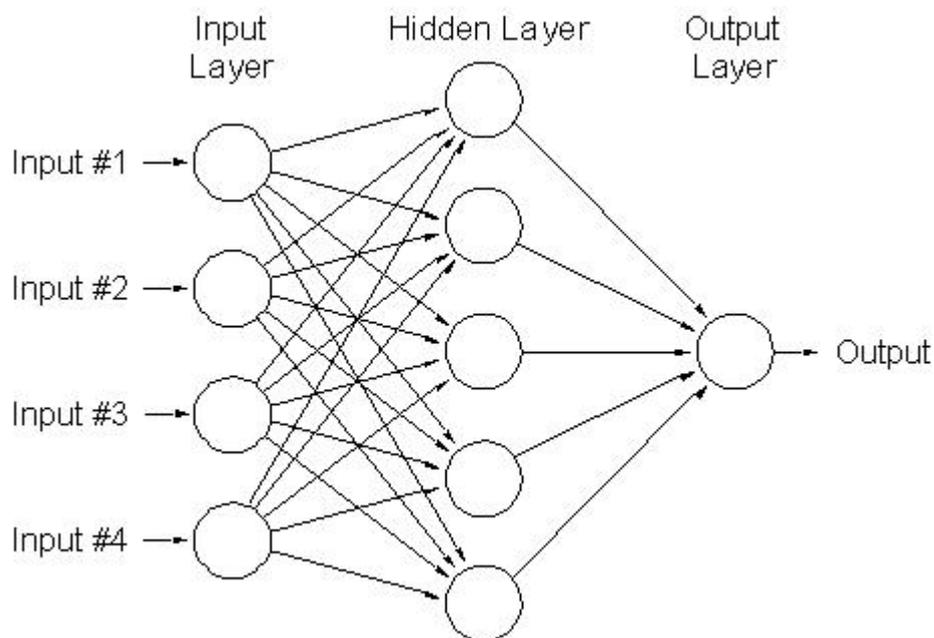
- **generalizzazione:** una rete neurale artificiale è capace di dare una risposta corretta anche in presenza di informazioni parziali di input, o di input mai visti in precedenza; come nei sistemi nervosi biologici, avviene una specie di riconoscimento per somiglianza, e tutto questo a prezzo di un modesto sforzo elaborativi, e quindi con un elevata velocità di risposta. Questa caratteristica deriva dal fatto che al loro interno vengono rappresentate, di solito, un numero di associazioni stimolo-risposta più grande del numero di sinapsi disponibili; nel far questo, la rete neurale tende ad estrarre le caratteristiche invarianti degli input piuttosto che memorizzare ogni singolo input. La misura in cui una rete neurale è in grado di generalizzare in presenza di un nuovo input dipende dal grado in cui queste caratteristiche invarianti possono essere ritrovate nell’input. Al contrario, i modelli computazionali di tipo seriale, anche quelli più recenti e potenti, per poter ottenere le stesse prestazioni in termini di generalità, pagano un enorme sforzo elaborativo.

- **recupero di memoria in base al contenuto:** partendo da dati di input incompleti, simili o corrotti dal rumore è possibile comunque dare una risposta completa. Come nei sistemi biologici, è sufficiente un “indizio” per dirigere l’attivazione del sistema nella direzione giusta, completando o recuperando l’intera memoria. I calcolatori seriali, invece, recuperano i dati impiegando un numero, che rappresenta l’indirizzo

di memoria corrispondente; se questo viene alterato o viene perso non è più possibile recuperare l’intero dato.

2.2 Le tipologie di reti neurali artificiali utilizzate

La tipologia di una rete neurale è quindi il particolare modo di come le sue parti elementari vengono organizzate ed interconnesse. La tipologia **Feed Forward** è la più semplice e maggiormente utilizzata ed è costituita da più di due strati di neuroni; in altre parole allo strato di input e di output se ne aggiunge uno o più strati nascosti (hidden layer). In questo tipo di rete neurale ogni neurone è connesso con tutti i neuroni dello strato precedente ma non ha connessioni con i neuroni del suo stesso strato e il segnale si propaga in maniera unidirezionale dall’input all’output attraverso gli hidden layer.



Schema di una rete Feed Forward generica

Il modo in cui questo segnale viene processato varia a seconda della specifica architettura implementata. In generale, seguendo la compilazione originale del

Perceptron, una delle prime reti neurali implementate, e considerando come esempio una rete con un solo hidden layer, la relazione tra un generico vettore input x e una variabile di output y si può esplicitare nella seguente espressione:

$$y = g \left[\beta_0 + \sum_{m=1}^M \beta_m \phi \left(\alpha_{0m} + \sum_{i=1}^d \alpha_{im} x_i \right) \right] + \varepsilon$$

Dove:

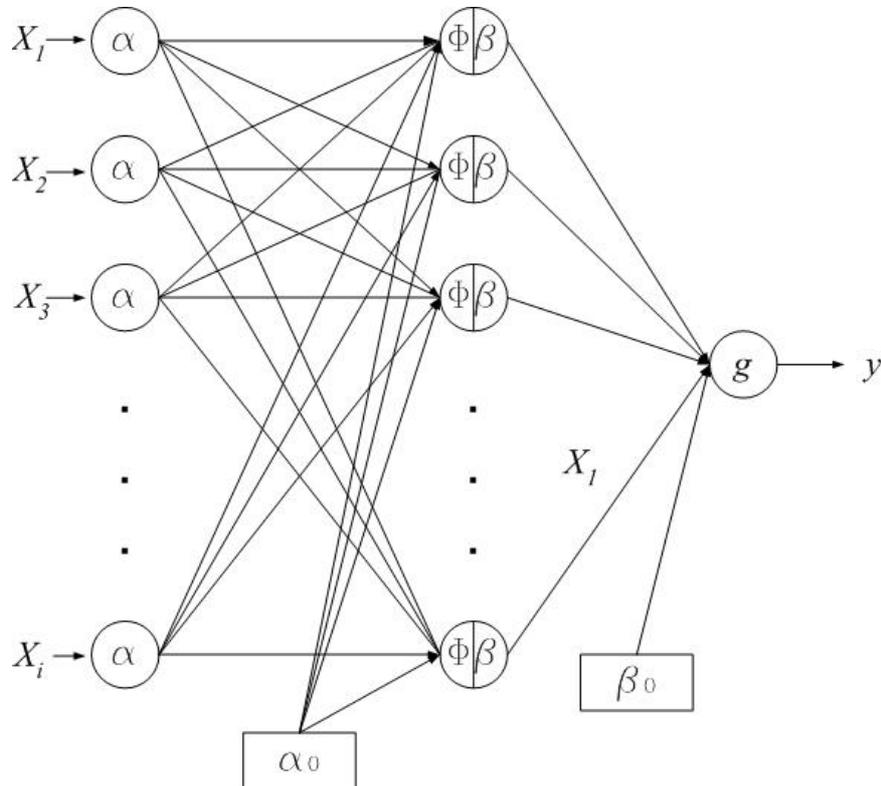
- α_{im} = è il peso della connessione tra i -esima variabile di input e l’ m -esima unità nascosta
- α_{0m} = è il bias dell’ m -esima unità nascosta
- $\Phi(*)$ = è la funzione di trasferimento che lega l’input alle unità nascoste
- β_m = è il peso del collegamento tra l’ m -esima unità nascosta e il vettore y di output
- β_0 = è il bias dell’output
- $g(*)$ = è la funzione di trasferimento che lega le unità nascoste all’output
- ε = è una variabile casuale

La quantità $z_m = \Phi(\alpha_{0m} + \sum_{i=1}^d \alpha_{im} x_i)$ rappresenta il valore dell’ m -esima unità nascosta. Come funzione $g(*)$ viene spesso utilizzata la funzione identità, mentre come funzione $\Phi(*)$ si usa una funzione sigmoideale, ossia a forma di S allungata, come ad esempio la funzione logistica :

$$\phi(z) = \frac{1}{1 + \exp(-z)} \quad \text{che ha condominio nell’intervallo } (0, 1)$$

Ovvero la funzione tangente iperbolica:

$$\phi(z) = \tanh(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \text{ che ha condominio nell'intervallo } (-1, 1)$$



Schema di una rete MLP ad un solo layer

Le reti **MLP, Multi Layer Perceptron**, che rispondono a questo tipo di relazione tra input e output, sono reti ad apprendimento supervisionato, in quanto la rete è addestrata (calibrata) con degli esempi (osservazioni) che contengono un input e l’output ad esso associato; esistono reti ad apprendimento non supervisionato, dove la rete deve essere in grado di estrarre delle informazioni di similitudine tra i dati forniti in input (senza associazioni con output desiderato) al fine di classificarli in categorie.

Per la stima dei parametri, si considera una funzione d’errore da minimizzare; un algoritmo iterativo molto utilizzato e molto studiato è quello di backpropagation in cui l’equazione di aggiornamento dei pesi al ciclo (i+1)-esimo è:

$$\underline{\hat{\theta}}_{i+1} = \underline{\hat{\theta}}_i - \eta \left. \frac{\partial E(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\underline{\hat{\theta}}_i} + \mu \Delta \underline{\hat{\theta}}_i$$

In cui in cui, $\underline{\hat{\theta}}_i$ è il vettore stima dei pesi (α_{im}, β_m) , mentre i parametri η e μ sono parametri positivi detti rispettivamente tasso di apprendimento e momento.

Le reti MLP godono della proprietà di essere ”approssimatori universali”, ossia, per M sufficientemente elevato, riescono ad approssimare qualunque funzione continua definita su un insieme compatto $D \subset R^d$.

Da un punto di vista applicativo però, è indispensabile che M non risulti eccessivamente grande, in modo che la rete abbia una buona capacità di generalizzazione. Per raggiungere questo obiettivo (ossia buona capacità di generalizzazione della rete), spesso si aggiunge un termine di penalizzazione alla funzione di errore; il termine di penalizzazione più usato per le reti neurali è chiamato ”weight decay” ed è:

$$\Omega = \sum_{m=0}^M \left[\beta_m^2 + \sum_{i=0}^d \alpha_{im}^2 \right]$$

Per applicare tale tecnica è essenziale che le variabili di input siano standardizzate, in modo tale che l’unità di misura delle variabili non influisca sulla grandezza dei pesi. Un altro approccio è quello chiamato Early Stopping; in altre parole viene suddiviso (casualmente) il campione in 2 sottoinsiemi; un sottoinsieme, definito campione di

addestramento (training set), viene utilizzato per stimare i parametri del modello ed un altro sottoinsieme, detto campione di verifica o test (test set), consente di misurare la capacità di generalizzazione del modello.

In pratica, ad ogni passo dell’algoritmo di ottimizzazione, si misura l’errore del modello, ossia $E(\hat{\theta})$, in entrambi i sottoinsiemi. L’errore calcolato sul training set tenderà a decrescere e sarà uguale a zero, nel caso di perfetto adattamento ai dati osservati (il che però implica una scarsa capacità di generalizzazione), mentre l’errore calcolato sul test set seguirà un andamento parabolico, ossia all’inizio tenderà a decrescere e poi, raggiunto un minimo, inizierà ad aumentare; ciò significa che il modello inizia a seguire eccessivamente la dinamica dei dati osservati, per cui si considera come stima dei parametri del modello, il valore assunto dai parametri nel passo in cui l’errore sul test set aveva raggiunto il minimo.

Per ridurre il numero di parametri si utilizzano anche tecniche di pruning (potatura): una volta addestrata la rete, si procede ad eliminare ad ogni passo quella connessione la cui rimozione produce il maggior peggioramento di adattamento e si procede a rimuovere pesi e/o neuroni fin quando non si raggiunge un adattamento prefissato accettabile. Per migliorare le performance di una rete potrebbe essere importante trasformare le variabili esplicative: ad esempio standardizzandole o normalizzandole, oppure ricorrendo a tecniche di riduzione e selezione delle variabili, come ad esempio l’analisi delle

componenti principali (PCA). Esistono tanti altri metodi di ottimizzazione, come ad esempio il metodo di Newton, il cui algoritmo è:

$$\underline{\hat{\theta}}_{i+1} = \underline{\hat{\theta}}_i - H^{-1} \left. \frac{\partial E(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\underline{\hat{\theta}}_i}$$

dove H^{-1} è l’inversa della matrice Hessiana della funzione di errore $E(\cdot)$, calcolata in $\underline{\hat{\theta}}_i$.

Sovente, per evitare problemi numerici, si considerano i metodi quasi-Newton, in cui si considera ad ogni iterazione un’approssimazione G_i della matrice H^{-1} , ossia:

$$\underline{\hat{\theta}}_{i+1} = \underline{\hat{\theta}}_i - \delta_i G_i \left. \frac{\partial E(\underline{\theta})}{\partial \underline{\theta}} \right|_{\underline{\theta}=\underline{\hat{\theta}}_i}$$

dove δ_i è un parametro che viene calcolato in ogni ciclo, mediante opportune minimizzazioni. Considerando una rete MLP, con d unità di input, M unità nascoste ed una sola unità di output, ci sono da stimare $M(d+2)+1$ parametri.

La rete neurale tipo **Radial Basis Function (RBF)** è una particolare architettura di reti feed forward con un singolo strato nascosto, in cui la funzione di attivazione $g(\cdot)$ per l’output è la funzione identità; quindi la parte deterministica del modello che lega y ad x è una combinazione lineare di M funzioni base non lineari.

Mentre nelle reti MLP, la funzione $\Phi(\cdot)$ dipendeva dal prodotto interno fra i vettori α_m ed x , nelle reti RBF essa dipende dalla distanza fra il vettore x ed un vettore prototipo μ_m , definito centro dell’ m -esima funzione radiale. La distanza $d(x, \mu_m)$ usata più spesso è quella euclidea, per cui si pone:

$$d(x, \mu_m) = \|x - \mu_m\| = \sqrt{\sum_{i=1}^d (x_i - \mu_{im})^2}$$

Quindi una generica rete RBF è definita dalla seguente equazione:

$$y = \sum_{m=0}^M \beta_m \phi(\|x - \mu_m\|) + \varepsilon$$

$$\text{con } \phi(\|x - \mu_0\|) = 1$$

Le reti RBF differiscono dalle reti MLP oltre che per l’argomento di $\Phi(*)$, che risulta

$\|x - \mu_m\|$ anzicchè $\sum_{i=1}^d (\alpha_{im} x_i)$, anche per la funzione stessa che, in questo caso, è una

funzione radiale (solitamente gaussiana):

$$\phi(\|x - \mu_m\|) = \exp\left\{-\frac{1}{2} \frac{\|x - \mu_m\|^2}{\sigma_m^2}\right\} \quad \text{con } \mu_m \in R^d, \sigma_m \in R^+ \quad \forall m = 1, \dots, M$$

Il parametro σ_m è detto ampiezza della m-esima funzione base radiale. Quindi in definitiva si ha che la risposta è data dalla seguente formulazione:

$$y = \sum_{m=0}^M \beta_m \exp\left\{-\frac{1}{2} \frac{\|x - \mu_m\|^2}{\sigma_m^2}\right\} + \varepsilon$$

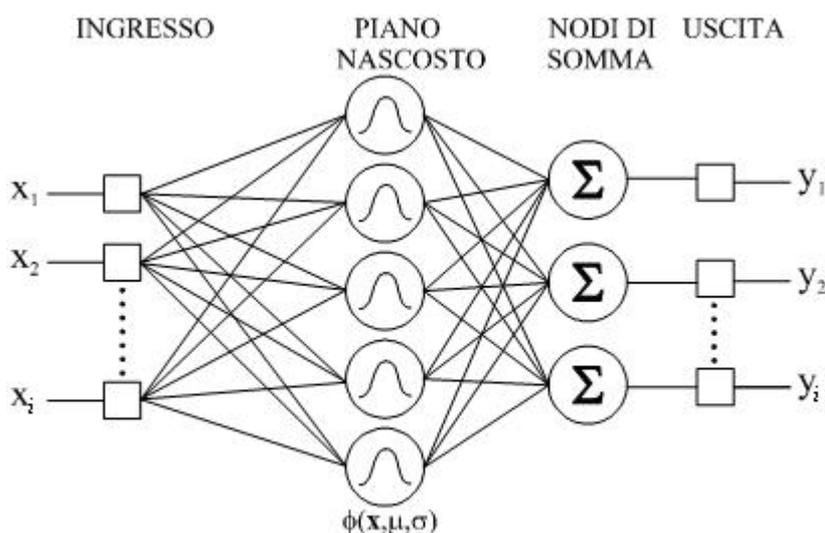
In tale modello i parametri da stimare sono $M(d+2)+1$.

E’ possibile generalizzare la funzione $\Phi(*)$ considerando una matrice \sum_m al posto di σ_m e scrivere:

$$\phi \left(\left\| \underline{x} - \underline{\mu}_m \right\| \right) = \exp \left\{ -\frac{1}{2} \left(\underline{x} - \underline{\mu}_m \right)^T \Sigma_m^{-1} \left(\underline{x} - \underline{\mu}_m \right) \right\}$$

Ma anche in questo caso cresce il numero di parametri da considerare.

Anche le reti RBF sono ”approssimatori universali”, per M sufficientemente elevato come per le MLP.



Schema di una rete RBF

Per la stima dei parametri si possono seguire due strategie alternative: considerare un processo di apprendimento supervisionato, oppure effettuare una procedura di stima a 2 stadi. Con la prima scelta, si considera una funzione d’errore (con la possibilità di aggiungere un termine di penalizzazione) e si determinano i parametri che simultaneamente la minimizzano. Tale scelta non è la più praticata; infatti spesso si preferisce considerare un algoritmo di stima a 2 stadi: nel primo stadio si stimano i centri delle funzioni base, mediante un processo di apprendimento non supervisionato e

successivamente si stimano i pesi β_m con un procedimento di apprendimento supervisionato, utilizzando le stime dei parametri delle funzioni base ottenute al primo passo. Tale scelta è molto più veloce da applicare della prima. I metodi di apprendimento non supervisionato, usati per stimare i parametri che definiscono le funzioni base, sono diversi.

Un primo approccio è quello di considerare i valori di $\hat{\underline{\mu}}_1, \hat{\underline{\mu}}_2, \dots, \hat{\underline{\mu}}_m$ pari ad un particolare sottoinsieme dei dati di input osservati; più in dettaglio si potrebbe seguire il seguente algoritmo: si inizia col considerare una sola funzione $x_{t_1} = 1, 2, \dots, N$, si pone $\hat{\underline{\mu}}_1 = x_{t_1}$ e si stimano i vari pesi $\hat{\beta}_1^{(t)}, \hat{\beta}_2^{(t)}, \dots, \hat{\beta}_N^{(t)}$ attraverso l’insieme completo di N osservazioni (sia di input sia di output). Finito il ciclo di osservazioni N il valore $\hat{\underline{\mu}}_1$ sarà pari al valore x_{t_1} che rende minima la funzione d’errore $E(\cdot)$. Successivamente si considera una rete con due funzioni base e si pone $\hat{\underline{\mu}}_2$ pari a x_{t_2} con $t_2 \in (\{1, \dots, N\} - \{t_1\})$ che rende minimo l’errore $E(\cdot)$ ed il procedimento continua fino alla determinazione di $\hat{\underline{\mu}}_m$.

Tale approccio è computazionalmente oneroso, in quanto in ogni ciclo si devono invertire diverse matrici. Una soluzione simile alla precedente, ma computazionalmente meno onerosa (se la funzione d’errore è quella dei minimi quadrati), è quella definita “*orthogonal least squares*”, che aggiunge ad ogni passo una funzione base, con i parametri stimati come descritto in precedenza, ma per scegliere quale dato porre uguale

al centro della funzione base, si considerano insiemi di funzioni base ortogonali, applicando il processo di ortogonalizzazione di Gram-Schmidt alle funzioni base $\{\phi(\|\underline{x} - \hat{\underline{\mu}}_m\|)\}$, quindi ad ogni ciclo si invertono matrici diagonali, che non richiedono elevati tempi di elaborazione. Le funzioni base ortogonali, vengono considerate solo per stimare i parametri $\hat{\underline{\mu}}_m$, una volta stimati si considerano tutte le funzioni base stimate e si stimano i $\hat{\beta}_1^{(t)}, \hat{\beta}_2^{(t)}, \dots, \hat{\beta}_N^{(t)}$ con algoritmi supervisionati.

Un altro approccio per stimare i parametri delle funzioni base è quello di considerare l’algoritmo di clustering delle K-medie; esso, a partire dalle osservazioni $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$, tenta di trovare quella partizione dei dati in K sottoinsiemi disgiunti (S_1, S_2, \dots, S_k) , ognuno contenente N_j osservazioni, tale da minimizzare la seguente funzione obiettivo:

$$J = \sum_{j=1}^K \sum_{t \in S_j} \|\underline{x}_t - \hat{\underline{\mu}}_j\|^2$$

Dove $\hat{\underline{\mu}}_j$ è la media delle osservazioni del j-esimo gruppo. Ossia:

$$\hat{\underline{\mu}}_j = \frac{1}{N_j} \sum_{t \in S_j} \underline{x}_t$$

L’algoritmo procede nel seguente modo: scelte casualmente K osservazioni, si assegna un’osservazione ad ognuno dei sottoinsiemi S_j e si pone la media del gruppo pari al

valore osservato assegnato al cluster considerato; successivamente ogni osservazione x_t

è assegnata a quel set S_j per la quale è minima la distanza $d(\hat{x}_t, \hat{\mu}_j)$,

assegnate le N osservazioni ai vari gruppi, si ricalcolano le medie $\hat{\mu}_j$ quindi si reitera il

procedimento, fin quando in due iterazioni consecutive, tutte le unità vengono assegnate

agli stessi gruppi. I valori finali di $\hat{\mu}_j$ sono le stime dei centri delle K funzioni base.

$$y = \beta_0 + \sum_{m=1}^M \beta_m \phi \left(\left\| \underline{x} - \hat{\mu}_m \right\| \right) + \varepsilon$$

Tali pesi si possono stimare con il metodo dei minimi quadrati ordinari, per cui:

$$\hat{\underline{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_M \end{pmatrix} = (\phi^T \phi)^{-1} \phi^T \underline{Y}$$

Dove:

$$\phi = \begin{pmatrix} 1 & \phi_{11} & \phi_{12} & \dots & \phi_{1M} \\ 1 & \phi_{21} & \phi_{22} & \dots & \phi_{2M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \phi_{N1} & \phi_{N2} & \dots & \phi_{NM} \end{pmatrix}$$

Con :

$$\phi_{tm} = \phi \left(\left\| \underline{x}_t - \hat{\mu}_m \right\| \right) \text{ e } \underline{Y} = \begin{pmatrix} y_1 & y_2 & \dots & y_N \end{pmatrix}^T.$$

Se invece si considera la funzione d’errore:

$$E = \frac{1}{N} \sum_{t=1}^N \left(y_t - \beta_0 - \sum_{m=1}^M \beta_m \phi \left(\left\| \underline{x}_t - \hat{\mu}_m \right\| \right) \right)^2 + \lambda \sum_{m=0}^M \beta_m^2$$

Allora si avrà:

$$\underline{\hat{\beta}} = (\phi^T \phi + \lambda I)^{-1} \phi^T \underline{Y}$$

Dove I è la matrice identità di ordine $M+1$.

La procedura di stima in 2 stadi, dove si utilizzano anche algoritmi di apprendimento non supervisionato, è nettamente meno complessa e per questo più veloce da implementare rispetto ad un processo di apprendimento supervisionato; però nel caso in cui ci sono predittori molto variabili, ma che hanno una relazione irrilevante con l’output, questi non vengono ignorati dall’algoritmo, per cui, quando i centri delle funzioni radiali sono scelti utilizzando solo dati di input, non c’è la possibilità di distinguere gli input rilevanti da quelli irrilevanti.

Le Reti neurali di regressione generalizzate (GRNN) sono una particolare espressione delle RBF e sono usate spesso per l’approssimazione di funzione. La differenza principale consiste nell’ hidden layer nella quale c’è un neurone dedicato ad ogni singolo caso presentato per l’addestramento. È stato dimostrato che, dato un certo numero di neuroni nascosti, una GRNN è in grado di approssimare una funzione continua con un certo grado di precisione. Come è noto, la regressione di una variabile dipendente x rispetto ad una variabile indipendente y consiste nel calcolo del valore di y più probabile dato x , sulla base di una serie di coppie (x,y) affette da errore sperimentale, dove x e y possono essere anche dei vettori; nel campo dell’identificazione di sistemi, la variabile y corrisponde all’uscita del sistema corrispondente ad un valore x fornito come input. Per realizzare un modello, è

necessario ipotizzare una qualche forma funzionale che assumerà la relazione tra le variabili, scegliendo un insieme di parametri da determinare in base ai dati sperimentali raccolti. I valori di questi parametri incogniti vengono determinati in modo da fornire il miglior adattamento possibile tra la funzione ipotizzata ed i valori registrati nella realtà. Il modello di rete neurale conosciuto con il nome di GRNN supera la necessità di tale ipotesi funzionale, permettendo di ricavarla direttamente senza la necessità di alcuna assunzione preliminare. In particolare, facendo affidamento solo sul risultato della funzione che definisce y , possiamo dire che y è connessa con x dalla media della sua funzione densità di probabilità (pdf). Se la pdf è nota possiamo scrivere:

$$\bar{y} = \frac{\int_{-\infty}^{+\infty} y \phi(x, y) dy}{\int_{-\infty}^{+\infty} \phi(x, y) dy}$$

Dove $\phi(x, y)$ è la funzione di probabilità condizionata che può essere stimata attraverso il metodo nonparametrico di Parzen:

$$\phi(x, y) = \frac{1}{(2\pi)^{3/2} h^3 n} \sum_{i=1}^n \exp\left(-\frac{D_i^2}{2h^2}\right) \exp\left(-\frac{(y_i - \bar{y})^2}{2h^2}\right)$$

con $D_i^2 = (x_i - \bar{x})^2 + (y_i - \bar{y})^2$

Sostituendo la formulazione di Parzen, si ottiene:

$$\bar{y} = \frac{\sum_{i=1}^n y_i \exp\left(-\frac{D_i^2}{2h^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2h^2}\right)} \quad (*)$$

O anche:

$$\bar{y} = \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i}$$

$$w_i = \exp\left(-D_i^2 / 2h^2\right)$$

La formula (*) è meglio conosciuta come il Regression Kernel Nadaraya-Watson Estimator, mentre la formulazione di w_i , può essere vista come la formulazione dei pesi dell’i-esimo punto, funzione della distanza dal centro. L’unico parametro libero è h e rappresenta la “larghezza di banda” della funzione di probabilità utilizzata.

Le principali caratteristiche delle reti GRNN sono le seguenti:

- La rete GRNN impara in un solo passo (one pass learning algorithm) ed è immediatamente in grado di generalizzare nel momento in cui vengono inseriti in memoria i dati da analizzare
- La superficie di regressione ottenuta converge a quella ottimale tanto più quanto più numerosi sono i dati disponibili
- La stima effettuata è limitata dai valori massimi e minimi presenti nei dati utilizzati
- Non sussistono problemi di convergenza a soluzioni corrispondenti a minimi locali della superficie di errore, come invece può accadere nel caso della back-propagation

- Una volta determinati i cosiddetti centri di cluster, ossia un sottoinsieme dei dati di partenza che conservi inalterate le informazioni sul comportamento del sistema, l’applicazione della tecnica di clustering dei dati permette di ridurre i tempi di calcolo e di realizzare modelli che si adattano in tempo reale ad eventuali variazioni del sistema che rappresentano.

Le reti GRNN hanno dunque molti vantaggi, ma presentano un inconveniente: sono molto lente ad operare all’aumentare della numerosità del vettore di input, perché utilizzano più calcoli rispetto ad altri tipi di rete per effettuare l’approssimazione di funzione o il compito di classificazione (per queste ultime applicazioni la rete neurale viene detta Probabilistica o PNN).

2.3 *Ultime applicazioni nel campo della previsione dei flussi veicolari*

Il crescente interesse verso l’utilizzo delle reti neurali artificiali nel campo dell’ingegneria dei trasporti ha portato allo sviluppo di moltissimi modelli progettati ad hoc per rispondere alle varie esigenze di ricerca. Un dato su tutti, a simboleggiare l’interesse per questa architettura computazionale, è dato da *Mark Dougherty*² che segnala 40 articoli pubblicati su rivista internazionale negli anni dal 1990 al 1995, a fronte di soli 5 articoli pubblicati su rivista internazionale tra il 1980 e il 1990. Numerosi infatti sono i tentativi fatti per riuscire ad individuare l’architettura di ANN ideale per poter risolvere problemi predittivi di un gran numero di parametri con un

² Mark Dougherty “*A review of neural networks applied to transport*” Transportation Research part C (1995)

basso numero di informazioni come input, come la previsione dei flussi di traffico con il conteggio dei flussi veicolari su alcuni rami ma non esiste ancora una risposta precisa a questo problema.

Una delle ultime esperienze significative in questo campo è stata fatta da *Chen e Grant-Muller*³ hanno sviluppato un’architettura molto avanzata di rete neurale, testandola sulla rete autostradale inglese, in particolare sulla zona che riguarda l’autostrada M25, di cui studi precedenti, fatti dallo stesso Chen nel 1997, hanno fornito un gran numero di dati per l’addestramento e i confronti.

L’iniziale studio di Chen fatto per l’ente autostrade inglese, prevedeva un classico schema multi layer con un algoritmo di apprendimento di tipo back propagation. La rete era addestrata su una parte dei dati, validata su la restante parte dei dati e poi usata per prevedere futuri valori. Dopo l’iniziale selezione di dati e input, i pesi della rete erano fissati senza che la rete fosse riaddestrata. Questo suggerì agli autori la ricerca di un algoritmo di apprendimento “dinamico”, che automaticamente riaggiusta il numero delle unità per simulare la natura mutevole e la complessità dei dati di traffico non appena vengono fuori. La rete neurale “dinamica” usata si rifà ad una rete di allocazione di risorse RAN (resource allocative network), sviluppata con l’intento di superare il problema di incompletezza non-deterministica polinomiale rilevabile nell’apprendimento con reti a valori fissi. La prima novità sta nel fatto che le unità

³ Haibo Chen, Susan Grant-Muller “*Use of sequential learning for short-term traffic flow forecasting*” Transportation Research part C (2001)

nascoste rispondono solo in un certo campo dello spazio dei valori degli input, attraverso le loro funzioni gaussiane a forma di campana, in tal modo le unità appena riallocate non interferiscono con le unità allocate precedentemente. In altre parole, non è richiesto alcun riaddestramento se arrivano altri dati nuovi. La RAN in oggetto è una rete ad un unico layer nascosto e adotta, inoltre, un apprendimento chiamato “sequenziale” o “dinamico” (sequential learning) che presenta alla rete i punti dei dati in maniera sequenziale. Ogni punto dei dati può essere rimosso una volta che è stato usato e non è necessario avere in memoria set di dati completi. Le prestazioni della rete possono essere testate aggiungendo dinamicamente un’unità nascosta che inizialmente apprende dal set di dati di input corrente; il numero di unità nascoste rappresenta così la complessità della funzione “nascosta”. La rete, così costruita, può essere usata in qualunque momento nel processo di apprendimento e può essere addestrata in qualunque momento durante il suo processo lavorativo. I risultati ottenuti sono stati significativi, con un grosso potere di generalizzazione da parte della rete e un PAE (percentage average error) pari al 9,5% contro oltre l’ 11% delle reti “statiche”, e aprono le porte ad un’ applicazione pratica utilizzando “on-line” i dati di traffico del centro di controllo dell’autostrada.

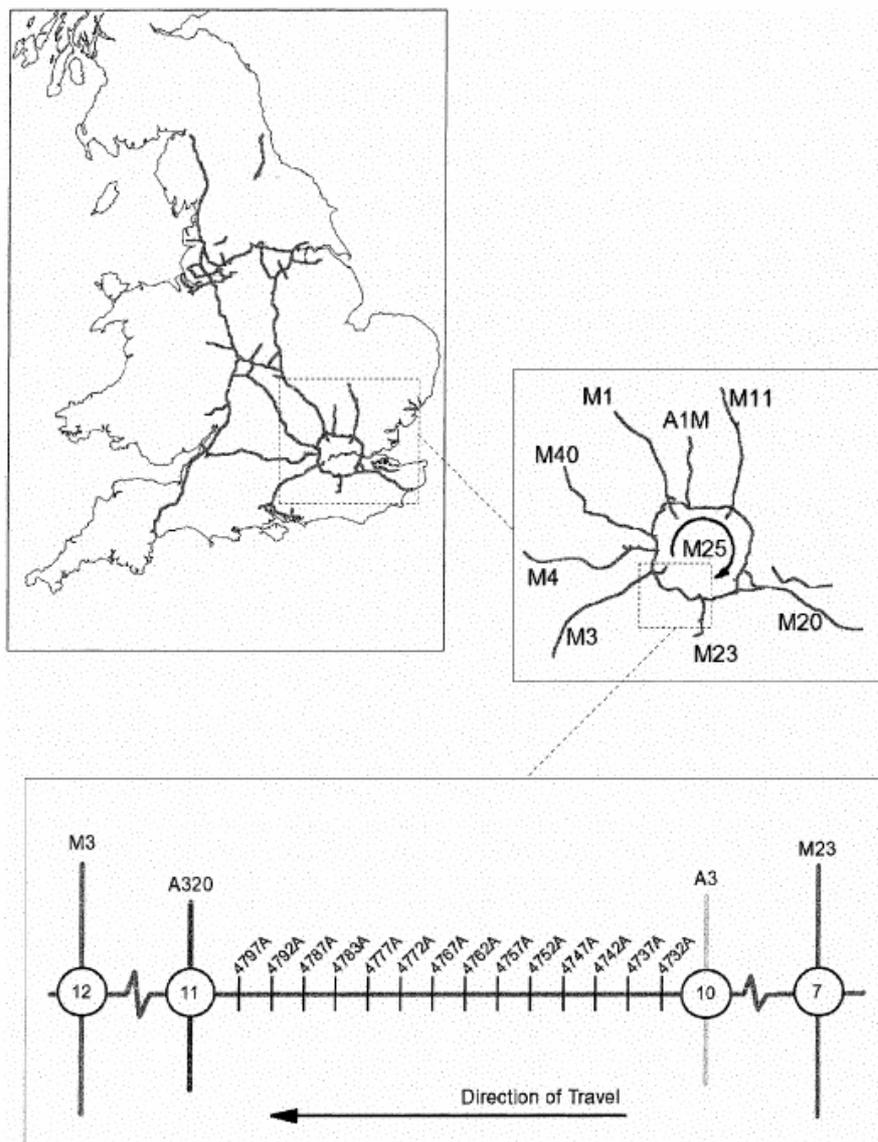


Figura 1 - Ambito di applicazione del lavoro di Chen e Grant.Muller (2001)

Un'altra esperienza significativa, è stata fatta nell'ultimo anno da *Vlahogianni, Karlaftis e Golias*⁴, che ha focalizzato l'attenzione sullo sviluppo di modelli di reti neurali ottimizzati per la previsione dei flussi veicolari su arterie fortemente congestionate. Il

⁴ Vlahogianni, Karlaftis, Golias “Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach” Transportation research part C (2005)

lavoro in oggetto, frutto di elaborazioni di lavori passati, cerca di definire una strategia di ottimizzazione strutturale dell’ANN, basata su un algoritmo di tipo genetico, capace di dare una risposta sia nella rappresentazione dei dati dei flussi veicolari con caratteristiche spaziali e temporali sia nella scelta della migliore struttura di rete neurale compatibile con il set di dati disponibile per l’addestramento. In particolare viene usato questo algoritmo di tipo genetico (GA), al posto del più comune procedimento finora usato di trial and error, nella fase di ottimizzazione di una rete di tipo Multy Layer Perceptron (MLP), ricercando così il numero di unità nascoste h , la funzione momentum α e il learning rate η . Gli algoritmi genetici o GA sono algoritmi di ricerca stocastica che differiscono dai più classici algoritmi di ricerca di una soluzione all’interno di una popolazione, sostanzialmente per 4 motivi:

- Lavorano con una codifica dei parametri e non con i parametri stessi
- Cercano su una popolazione di possibili soluzioni, non la singola soluzione
- Utilizzano solo informazioni di tipo oggettivo
- Utilizzano regole di tipo probabilistico, non deterministico

I GA basano il loro meccanismo di funzionamento su 3 principi della genetica: la selezione, il crossover o incrocio tra due segnali, la mutazione. Inizialmente quindi c’è una popolazione A di possibili soluzioni, codificate come cromosomi, e ogni cromosoma consiste di geni. Durante la fase di selezione, viene scelto un cromosoma dalla popolazione corrente per essere messo nella prossima popolazione. La selezione

non crea nuovi cromosomi, ma li ripropone in nuove popolazioni, secondo la regola di incrementare i cromosomi che meglio si adattano e diminuire il numero di cromosomi che si adattano peggio. Prima di entrare nella nuova generazione di popolazione i cromosomi scelti subiscono un processo di crossover, ossia due cromosomi generano un nuovo cromosoma che prenderà le migliori caratteristiche dei due che lo hanno generato e che sarà migliore di entrambe i due cromosomi generatori. La mutazione riguarda invece il cambiamento dei geni all’interno di un cromosoma e permette di avere una soluzione migliore di quanto sia precedentemente possibile. Queste due fasi hanno un’importanza fondamentale nella ricerca di un’ottimizzazione della risposta: la mutazione serve per evitare un’eccessiva permanenza dei geni in un cromosoma, il crossover serve a evitare ripetizioni di risposte uguali tra loro.

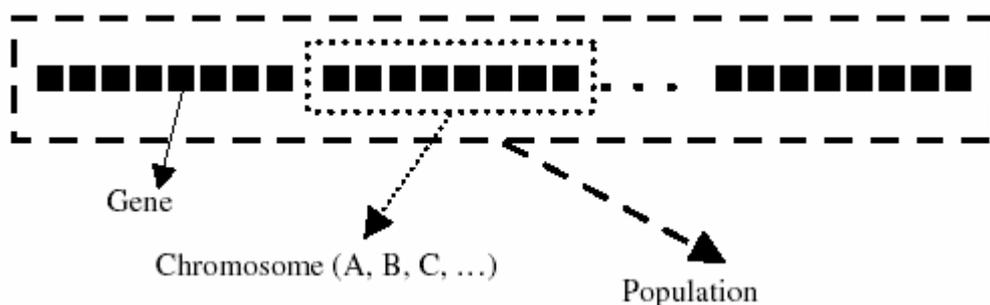


Figura 2 Rappresentazione di una popolazione (Vlahogianni et al. - 2005)

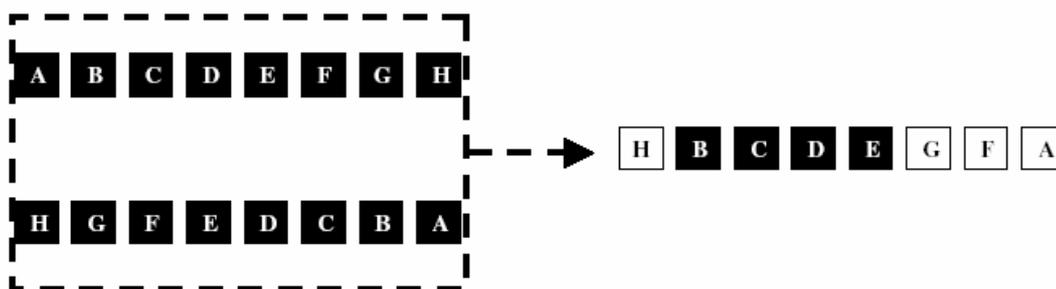


Figura 3 Processo di crossover (Vlahogianni et al. - 2005)

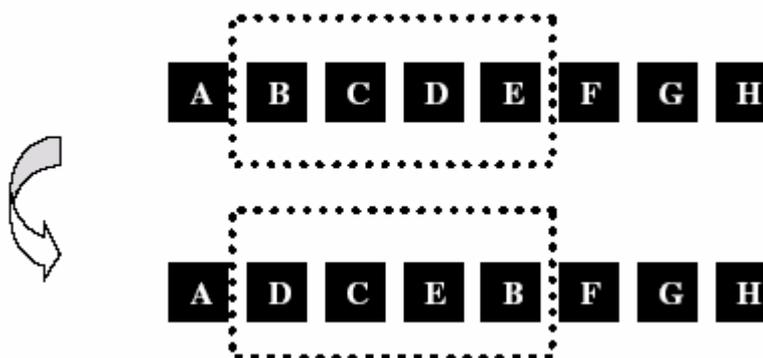


Figura 4 Processo di mutazione (Vlahogianni et al. - 2005)

2.4 La logica fuzzy nelle reti neurali artificiali.

L’integrazione di questi modelli con architettura a rete neurale con la logica fuzzy risulta essere l’ultimo indirizzo di ricerca dell’argomento. Tale applicazione deriva dal fatto che, talvolta, i dati di input risultano essere eccessivamente rumorosi rispetto al set di dati di addestramento, e questo implica, per la maggior parte dei casi, risultati scadenti o addirittura completamente sbagliati. In campo trasportistico, basti pensare alla notevole intensità di fluttuazione dei flussi di traffico e della stessa domanda di spostamento, anche in intervalli molto brevi, che si accentuano esponenzialmente su reti

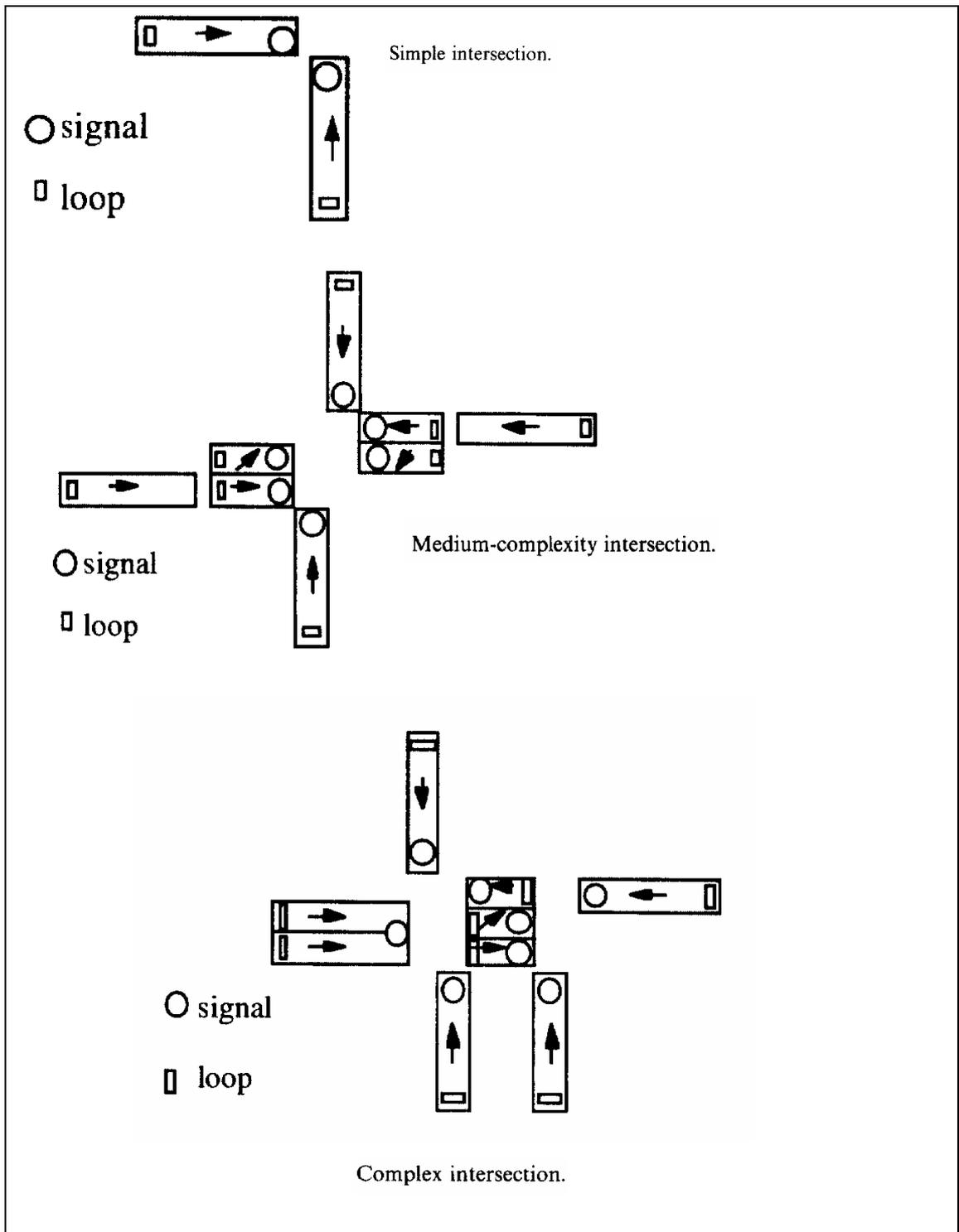
di grandi dimensioni. Per ovviare a questo inconveniente, si è pensato all’integrazione del ragionamento “sfumato” o fuzzy, che pone a base del suo sviluppo il concetto che un’ipotesi non è mai completamente vera o falsa, ma ha un suo “grado di verità”, che inciderà sulla forza con cui verrà seguita una regola e applicata la sua conseguenza. Il grado di verità di un’ipotesi è da interpretare, nei campi applicativi ingegneristici ed economici, come il grado di appartenenza o “credibilità” ad un determinato range e ciò è dovuto al fatto che questi range non sono separati nettamente, ma si sovrappongono, con funzioni di appartenenza variabili dal tipo sinusoidale a quello triangolare; in definitiva una variabile di input potrebbe appartenere a due o più range, con altrettanti gradi di credibilità. Dal punto di vista computazionale, tale tipo di ragionamento, appesantisce molto l’architettura seriale dei processori attualmente in uso, e l’utilizzo combinato della rete neurale con la logica fuzzy risulta, quindi, essere estremamente oneroso; l’avvento di un’architettura parallela nei nostri computer, tra qualche anno, dovrebbe dare l’impulso definitivo all’applicazione di questa metodologia. In campo trasportistico, nella ricerca della matrice OD, un modo diverso per affrontare il problema della capacità di generalizzazione della rete neurale senza ricorrere alla logica fuzzy consiste nell’utilizzo di una matrice a priori sufficientemente affidabile e perturbarla su intervalli di confidenza prefissati dall’analista. Questo set di dati perturbato sarà anche quello che utilizzeremo per l’addestramento della rete e, di conseguenza, quanto maggiore sarà l’affidabilità della matrice iniziale, tanto migliori

saranno le stime della domanda restituite dalla rete neurale in risposta ai flussi reali misurati sul sistema di trasporto. La possibilità di evitare la fuzzyficazione dei dati per ottenere maggior potere di generalizzazione non ha però limitato le esperienze in questo campo, che anzi hanno messo in luce la “veridicità” dei risultati ottenuti ma anche la limitata applicabilità, per ora, ai soli modelli semplici. *Teodorovic*⁵, uno dei maggiori sostenitori della logica fuzzy, ha messo in luce quelle che sono le principali differenze tra la teoria delle probabilità e le nuove teorie alternative di cui la fuzzy theory è la più rappresentativa. Nonostante ci siano molti studiosi scettici sull’utilizzo di tale teoria, egli ritiene che la logica fuzzy sia un ulteriore strumento che dà la possibilità di spiegare in modo più “ingegneristico” alcuni fenomeni e che comunque non è uno strumento sostitutivo della teoria delle probabilità, piuttosto uno strumento da affiancare per risolvere i problemi trasportistici più complessi.

*Henry et al.*⁶ hanno testato le potenzialità di una rete neuro-fuzzy ad alcune intersezioni, selezionate secondo la loro complessità (simple, medium-complexity, complex), per definire un metodo di controllo in tempo reale dell’intersezione. I risultati ottenuti hanno messo in luce proprio questa limitata applicabilità; sono stati soddisfacenti per le intersezioni semplici e mediamente complesse, ma piuttosto scadenti per quelle complesse. Nel contempo è però emerso l’enorme potenziale che questo strumento può avere negli anni a venire.

⁵ Teodorovic “*Fuzzy logic system for transportation engineering: the state of the art*” Transportation Research Part A (1999)

⁶ J.J.Henry, J.L.Farges, J.L.Gallego “Neuro-Fuzzy techniques for traffic control” Control Engineering Practise vol.6 (1998)



3 La metodologia applicata

I modelli di rete neurale non sono di facile applicazione nel campo dei trasporti, soprattutto in ambito urbano, in quanto necessitano di una numerosità di dati consistente, da utilizzare nella fase di training o addestramento della rete, che spesso non è possibile ottenere se non a costo di un notevole impiego di risorse economiche e di tempo.

In particolare, per definire una rete neurale sono necessari, nell’ambito di tale problema legato alla teoria dei sistemi di trasporto, un notevole numero di osservazioni di flusso e di domanda. Per quanto riguarda i primi, la moderna tecnologia permette di ottenere una buona base dati attraverso l’installazione di rilevatori di flusso. La domanda, invece, è di difficile stima ed è praticamente impossibile rilevarla con continuità.

Per questo motivo, le applicazioni delle reti neurali artificiali hanno riguardato quei casi di controllo su aree ristrette nei quali la mancanza di domanda interna all’area di studio permette di prescindere dalla valutazione della domanda che coincide, nei fatti, con il flusso rilevato in limitato numero di sezioni (tipici sono i casi di controllo delle intersezioni, rotatorie, etc...). Nella letteratura corrente, per ovviare a tale inconveniente, vengono ricavate una serie di matrici OD modificando percentualmente i valori di una matrice nota a priori, e ricavando i flussi sulla rete di trasporto mediante l’assegnazione di tali matrici. Questo, se da un lato riduce ad uno i parametri, dall’altro

crea altresì una collinearità dei dati che potrebbe far rispondere la rete neurale in modo non corretto.

Un ulteriore aspetto legato all’utilizzo delle reti neurali artificiali riguarda la potenza degli elaboratori su cui si implementano tali architetture neurali. Essendo, per stessa natura delle ANN, un processo computativo di tipo parallelo, esso mal si adatta a processori seriali che costituiscono, di fatto, la quasi totalità del mercato. Le esperienze riportate, infatti, in letteratura fanno rilevare che l’associazione di input di uno o più ordini di grandezza inferiore all’output, ad esempio il vettore flussi (qualche centinaio di elementi) rispetto al vettore matrice (che facilmente supera il migliaio di elementi), non è di facile soluzione e necessita di un numero di dati dedicati all’apprendimento dell’ordine delle decine di migliaia di unità, di molto tempo e risorse, e la rete non risponde velocemente come le si richiede.

Nell’ambito della stima e aggiornamento della matrice Origine Destinazione, ci si propone, in questo lavoro, di sviluppare una metodologia per creare il training set, e quindi le matrici di domanda, a partire dai parametri dei modelli di domanda della matrice nota a priori e di estrarre in maniera casuale tali parametri secondo la loro distribuzione statistica. I dati così creati non hanno alcun legame tra loro se non il fatto che discendono da una statistica che fa riferimento ai probabili valori medi di accadimento, evitando quindi la forte correlazione insita nella generazione di dati tutti dipendenti in maniera lineare tra di loro. Inoltre la creazione di dati a partire dai modelli

di generazione e distribuzione permette di evitare alcune incongruenze, oggetto di critiche, con la metodologia tradizionale che modifica i singoli valori della matrice OD senza alcun controllo.

Inoltre, data la netta inferiorità del numero di parametri che definiscono la domanda rispetto alla numerosità della stessa matrice origine destinazione, la metodologia proposta semplifica molto l’applicazione ed il progetto stesso della rete neurale rendendola compatibile con le applicazioni richieste su reti di grandi dimensioni. La scelta, effettuata in questa fase, di associare ad un sottoinsieme dei flussi dell’area, simulando in tal modo il rilevamento dei flussi su alcune arterie urbane, un vettore composto dai pesi degli attributi per la domanda di mobilità interna e dal flusso misurato in entrata e in uscita dall’area di studio per sezione di controllo per quanto riguarda la domanda di scambio-attraversamento, velocizza di molto la velocità di risposta da parte della rete, senza la quale verrebbe meno uno dei presupposti di utilizzo di tale sistema, ossia la risposta in tempo reale; in tal modo infatti sia input che output risultano dello stesso ordine di grandezza, di qualche decina di elementi, e tale dimensione dei vettori riduce di conseguenza anche il numero di esempi da presentare alla rete per l’addestramento. Verranno di seguito riportati alcuni cenni relativi alla domanda di mobilità e agli strumenti statistico-matematici che interverranno nella definizione della metodologia proposta. Nel paragrafo seguente si richiameranno i

concetti principali relativi alla definizione e stima della domanda di trasporto, necessari per introdurre nel prosieguo la proposta metodologica oggetto della tesi.

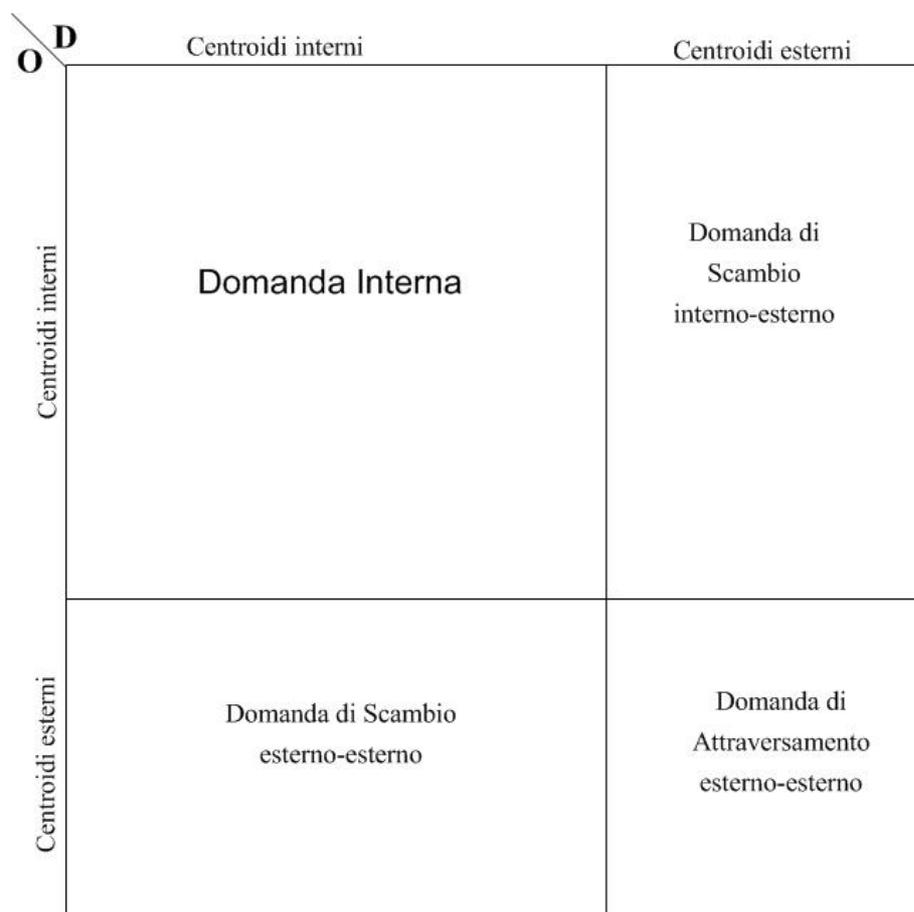
3.1 La matrice Origine Destinazione

La domanda di trasporto passeggeri risulta essere *l’insieme degli utenti o dei potenziali tali (con i loro comportamenti e le loro caratteristiche) che, in un intervallo di tempo di riferimento, chiedono di usufruire dei servizi e delle infrastrutture di trasporto.* Giacchè la mobilità può essere vista, nell’ambito del più generale sistema complessivo delle attività territoriali, come un’esigenza derivata da esigenze primarie legate allo svolgimento di attività economiche e sociali sul territorio, la domanda di trasporto è da considerarsi come una domanda indotta.

Ancorché la dislocazione delle attività sul territorio influenzi sempre la domanda di mobilità, tale influenza può caratterizzarsi in maniera più o meno diretta e più o meno vincolante a seconda dei motivi che caratterizzano gli spostamenti e della modalità con cui avvengono gli stessi.

In generale, dunque, la domanda di trasporto è definita secondo una caratteristica temporale e una spaziale; la caratterizzazione spaziale è intrinseca al fenomeno della mobilità, per questo motivo gli spostamenti devono essere suddivisi secondo la loro *Origine e Destinazione*, e la rappresentazione degli spostamenti è solitamente la **matrice Origine-Destinazione**, o **matrice OD**. Inoltre poiché la domanda ha anche una

dimensione temporale, le matrici OD fanno riferimento ad un precisato intervallo di tempo.



Esempio di Matrice OD

Tutti gli approcci relativi alla stima delle matrici OD sono classificabili in due famiglie principali: i metodi di *stima diretta* e i metodi di *stima da modello*.

La *stima diretta* della matrice OD si riferisce ad indagini campionarie sull’utenza del sistema. Tali tipologie di indagini, note come indagini Origine–Destinazione, sono da definire in funzione delle caratteristiche e della qualità delle informazioni che si intende ottenere. Le indagini più comunemente utilizzate sono le indagini durante il viaggio o “a bordo”, le indagini a domicilio, le indagini a destinazione e le indagini postali.

Le indagini a bordo sono indagini che possono essere effettuate a bordo strada quindi indagini di tipo “puntuale”, o all’interno dei mezzi di trasporto pubblico, sono indagini che tengono conto del numero di persone in transito e di una parte di essi che viene intervistato, utilizzando quindi un campionamento di tipo casuale semplice. Le indagini a domicilio, sicuramente le più costose, sono però le più complete; si effettua un campionamento, semplice o, più spesso, stratificato, delle persone o famiglie residenti nell’area di studio a cui viene sottoposto un questionario più o meno articolato. Le indagini a destinazione vengono effettuate sui luoghi di lavoro, scuole e, generalmente, in tutti quei luoghi che possono essere di destinazione per uno spostamento. Le indagini postali consistono in un questionario inviato per posta e successivamente ritirato.

In tutte queste indagini campionarie ha grande importanza la definizione della popolazione e le sue caratteristiche per individuare il miglior tipo di campionamento relativo all’ambito analizzato. Le strategie di campionamento utilizzate, e qui sopra accennate, sono tutte di tipo probabilistico, ovvero fanno riferimento a metodi di estrazione del campione per i quali è possibile definire i possibili risultati dell’estrazione e del campione, assegnare una probabilità a ciascun risultato ed estrarre a caso gli elementi del campione con quella probabilità. Si distinguono quindi:

- Campionamento casuale semplice, dove ogni elemento della popolazione campionaria ha uguale probabilità di essere estratto

- Campionamento casuale stratificato, ovvero la popolazione è suddivisa in strati secondo una certa caratteristica, in maniera tale da non far sovrapporre i vari strati e da esaurire la popolazione. Tutti gli elementi appartenenti ad uno strato hanno uguale probabilità di essere estratti, mentre due elementi di strati diversi possono avere probabilità diverse.
- Campionamento a grappolo, dove le unità di riferimento sono gruppi di elementi i quali hanno medesima possibilità di essere estratti, oppure sono suddivisi anch’essi in strati (campionamento a grappolo stratificato). Questo tipo di campionamento permette anche un campionamento a due stadi, ovvero operando prima un’ estrazione sui grappoli e successivamente sugli elementi del grappolo.

Alla fase di campionamento segue la fase di stima dei risultati, attraverso uno *stimatore statistico* che, a partire dalle informazioni ottenute attraverso campionamento, permette di ottenere una stima della quantità incognita cercata. La scelta dello stimatore da adottare risulta critica in quanto dipende fortemente dal tipo di grandezze di cui si intende ottenere una stima e della strategia di campionamento adottata; difatti uno stimatore statisticamente “efficiente” per una data strategia di campionamento, può non esserlo affatto per un’altra.

Ad esempio, nel caso di campionamento casuale semplice, si considera un campionamento di n elementi su un universo di numerosità N , e si assume che il numero totale di spostamenti effettuati dal campione n_{OD} sia somma degli spostamenti dell’i-

esimo elemento che si sposta da O a D per uno scopo ed in un orario determinati,

ovvero $n_{OD} = \sum_{i=1..n} n_{OD}^i$. Pertanto una stima corretta della domanda viene data

dall’espressione:

$$\hat{d}_{OD} = (N/n)n_{OD} = (1/\alpha)n_{OD} = N\bar{n}_{OD}$$

Dove α è detto “tasso di campionamento” e \bar{n}_{OD} è il numero medio di spostamenti sulla determinata coppia OD per singolo elemento campionato. La varianza è data dalla seguente espressione:

$$Var[\hat{d}_{OD}] = \frac{N^2 \hat{s}^2 (1-\alpha)}{n} \quad \text{con} \quad \hat{s}^2 = \frac{\sum_{i=1..n} (n_{OD}^i - \bar{n}_{OD})^2}{(n-1)}$$

La caratteristica \hat{s}^2 altro non è che la stima campionaria della varianza della v.a. n_{OD}^i .

Accade sovente che durante le indagini di domanda (soprattutto nelle indagini al cordone) un elemento del campione è in possesso o meno di una caratteristica richiesta (ad esempio effettua lo spostamento per un certo motivo). In tal caso n_{OD}^i assume i valori 0 ovvero 1 e pertanto la \bar{n}_{OD} diventa una stima campionaria della percentuale di utenti che hanno effettuato lo spostamento secondo quella o quelle determinate caratteristiche, e viene quindi distinta dalla precedente indicandola con:

$$\hat{P}_{OD} = \sum_{i=1..n} \frac{n_{OD}^i}{n}$$

In questo caso, la stima campionaria della varianza di n_{OD}^i può essere espressa in maniera proporzionale alla varianza di una variabile bernoulliana.

$$\hat{s}^2 \cong \hat{P}_{OD}(1 - \hat{P}_{OD})$$

Con il segno di \cong dovuto all’approssimazione di aver considerato uguali n e $(n-1)$, solitamente accettata visto i numeri di solito in gioco. Pertanto la varianza di \hat{P}_{OD} può essere calcolata come:

$$Var[\hat{P}_{OD}] = \frac{\hat{P}_{OD}(1 - \hat{P}_{OD})(1 - \alpha)}{n}$$

Nel caso di un campionamento casuale stratificato, largamente utilizzato nella pratica delle indagini campionarie, le espressioni di stima della domanda e della sua varianza sono essenzialmente simili a quelle utilizzate nel campionamento semplice con la particolarizzazione delle v.a. rispetto allo strato e con l’aggiunta di un peso relativo allo strato. Supponendo di aver suddiviso quindi l’universo in K gruppi di utenti, si avranno n^k elementi estratti da N^k elementi appartenenti al k -esimo strato:

$$\hat{d}_{OD} = N \sum_{k=1..K} w_k \sum_{i=1..n_k} \frac{n_{OD}^{ik}}{n_k} = N \sum_{k=1..K} w_k \bar{n}_{OD}^k$$

$$Var[\hat{d}_{OD}] \approx N^2 \sum_{k=1..K} \frac{w_k^2 \hat{s}_k^2 (1 - \alpha_k)}{n_k}$$

Dove \hat{s}_k^2 è la stima campionario della v.a. n_{OD}^{ik} espressa come:

$$\hat{s}_k^2 = \frac{\sum_{i=1..n_k} (n_{OD}^{ik} - \bar{n}_{OD}^k)^2}{(n - 1)}$$

E α_k è il tasso di campionamento dello strato k -esimo.

Le varianze delle stime campionarie sono utili per calcolare i limiti dell’intervallo di confidenza della stima della domanda, ovvero l’intervallo entro il quale si ha una probabilità $(1-\gamma)$ di trovare il valore vero della stessa. Difatti se il numero di elementi campionati è sufficientemente numeroso, per il teorema centrale del limite, si può assumere che lo stimatore della domanda segua una legge Normale della probabilità.

Pertanto i limiti suddetti si ricavano dalla seguente formulazione:

$$L_{\gamma}^S(d_{OD}) = \hat{d}_{OD} + z_{1-\gamma/2} \text{Var}[\hat{d}_{OD}]^{1/2}$$

$$L_{\gamma}^I(d_{OD}) = \hat{d}_{OD} + z_{\gamma/2} \text{Var}[\hat{d}_{OD}]^{1/2}$$

Dove $z_{1-\gamma/2}$ e $z_{\gamma/2}$ sono i percentili della variabile Normale Standard che hanno una probabilità $1-\gamma/2$ e $\gamma/2$ di non essere superati. Per $\gamma = 0.05$ tali variabili sono rispettivamente 1.96 e -1.96 e i limiti di confidenza rappresentano gli estremi dell’intervallo che con il 95% di probabilità contiene il valore vero.

La matrice OD, nella **stima da modello**, non è altro che il risultato di una simulazione della domanda di trasporto mediante dei modelli di utilità aleatoria. Tali modelli debbono essere necessariamente specificati, calibrati e validati; in altre parole è necessario definire la forma funzionale e le variabili che compaiono nel modello, stimare i coefficienti delle variabili e verificare la “qualità statistica” del modello. La forma funzionale del modello, come ad esempio un Logit Multinomiale o un Logit Gerarchizzato, dipende dalla trattabilità computazionale, le esperienze fatte in casi

analoghi, le aspettative a priori sull’indipendenza dei residui aleatori, mentre la scelta della variabili esplicative dipende dal tipo di modello scelto e da alcune considerazioni su di esse in modo da evitare qualsiasi genere di col linearità. Tali modelli matematici possono essere visti come delle relazioni matematiche che forniscono la probabilità $p_i[j](X^i, \beta, \theta)$ che l’individuo i scelga l’alternativa j in dipendenza del vettore attributi X^i di tutte le alternative, del vettore β dei parametri relativi alle utilità sistematiche e del vettore θ della probabilità congiunta dei residui aleatori.

Questa dipendenza, ricordiamo, si esplicita nella dipendenza di $p[j]$ dall’utilità sistematica V_j dell’alternativa, che di solito viene espressa come combinazione lineare degli attributi delle j alternative X_z^i per i parametri β_z :

$$V_j (X_j^i) = \sum_z \beta_z \cdot X_{zj}^i$$

La dipendenza dal parametro θ avviene invece attraverso la forma funzionale dello specifico modello utilizzato che a sua volta dipende dalla funzione di distribuzione di probabilità congiunta dei residui aleatori ε . Assunti noti gli attributi delle alternative, possiamo dire che gli elementi caratterizzanti la matrice OD sono i parametri β e θ , e che dalla loro stima dipende l’accuratezza nel riprodurre la distribuzione dei comportamenti osservati da un gruppo di decisori considerato.

In fase di calibrazione, il metodo di stima dei parametri comunemente usato è quello della Massima Verosimiglianza (o ML che sta per Maximum Likelihood), che fornisce i

valori dei parametri incogniti che massimizzano la probabilità di osservare le scelte effettuate dagli utenti, ed si può esprimere nella forma generale come:

$$[\beta, \theta]^{ML} = \arg \max L(\beta, \theta) \quad L(\beta, \theta) = \prod_{i=1 \dots n} p^i[j(i)](X^i, \beta, \theta)$$

O anche in forma di logaritmo naturale:

$$[\beta, \theta]^{ML} = \arg \max \ln L(\beta, \theta) \quad \ln L(\beta, \theta) = \sum_{i=1 \dots n} \ln p^i[j(i)](X^i, \beta, \theta)$$

La stima mediante il metodo della Massima Verosimiglianza non ci consente di stimare direttamente i parametri β ma i rapporti β/θ , per cui il vettore θ dei parametri della matrice di dispersione viene stimato “congiuntamente” alla stima dei parametri β di reciproca sostituzione. Il problema di massimizzazione della verosimiglianza, esposto formalmente nella formula sopra, può essere risolto con un algoritmo del gradiente, la cui specificazione dipenderà dal modello di utilità aleatoria utilizzato. E’ inoltre opportuno ricordare che è possibile ottenere una stima delle varianze e delle covarianze delle stime di β^{ML} ottenute con il metodo della Massima Verosimiglianza in quanto la sua matrice di dispersione Σ tende asintoticamente, all’aumentare del campione utilizzato, ad essere pari all’inverso della matrice Hessiana della funzione log-likelihood, calcolata nel punto $(\beta, \theta)_{ML}$, cambiata di segno:

$$\Sigma_{\beta, \theta} = - \left[\frac{\partial^2 \ln L(\beta, \theta)}{\partial \beta \partial (\beta, \theta)^T} \right]_{(\beta, \theta)_{ML}}^{-1}$$

Per un campione di utenti sufficientemente numeroso, può essere utilizzata questa importante proprietà per calcolare le varianze e i limiti di confidenza dei coefficienti del modello considerato.

La bontà di stima dei coefficienti può essere appurata mediante alcuni test formali (fase di validazione) che verificano alcune ipotesi come, ad esempio, il test t-Student che verifica l’ipotesi H_0 che un coefficiente β_k sia pari a zero e che invece la sua stima β_k^{ML} sia diversa da zero, in conseguenza di alcuni errori campionari:

$$t = \frac{\beta_k^{ML}}{Var[\beta_k^{ML}]^{1/2}}$$

Oppure, equivalentemente, che due coefficienti β_k e β_j siano uguali:

$$t = \frac{\beta_k^{ML} - \beta_j^{ML}}{[Var(\beta_k^{ML}) + Var(\beta_j^{ML}) - 2cov(\beta_j^{ML}, \beta_k^{ML})]^{1/2}}$$

Un altro tipo di validazione del modello può essere effettuato attraverso un test sulla bontà dell’accostamento (goodness of fit) del modello, in altre parole sulla capacità da

parte del modello di riprodurre le scelte realmente effettuate dall’utente. Tale test solitamente viene fatto con un campione di utenti diverso da quello utilizzato per la calibrazione (hold out sample) utilizzando la statistica rho quadro (ρ^2):

$$\rho^2 = 1 - \frac{\ln L(\beta_{ML})}{\ln L(0)}$$

Tale statistica, normalizzata nell’intervallo [0, 1] restituisce un indice sulla capacità del modello di riprodurre le scelte osservate. Difatti vale 0 se $L(\beta_{ML})$ è uguale a $L(0)$, inteso quindi come se il modello non avesse nessuna capacità esplicativa del fenomeno, vale 1 se il modello fornisce una probabilità pari a 1 di osservare le scelte fatte dall’utente, inteso come se il modello avesse una perfetta capacità riproduttiva dell’evento.

3.2 Il metodo Montecarlo

Il metodo Montecarlo è ritenuto uno dei metodi di simulazione numerica più efficaci ed importanti tutt’oggi usati. In generale il Montecarlo è un metodo per risolvere determinati problemi di calcolo matematico mediante la costruzione di un processo stocastico i cui parametri convergono stocasticamente alle quantità che si vogliono calcolare. L’effettiva valutazione di tali quantità viene realizzata eseguendo delle osservazioni del processo stocastico e stimandone i parametri secondo le procedure fornite dalla statistica. Così, come nel nostro caso, per valutare un parametro θ_0 , di cui si ha una definizione matematica, si può costruire un schema stocastico cioè una sequenza,

teoricamente infinita, di variabili casuali X^n_θ tale che, al crescere di n , la X^n_θ converga stocasticamente verso θ_0 , ossia:

$$\lim_{n \rightarrow \infty} \Pr \left\{ X^{(n)} - \theta_0 \mid < \varepsilon \right\} = 1$$

Da un punto di vista generale il funzionamento del metodo Montecarlo prevede sempre due fasi fondamentali: la prima consiste nella creazione di un modello che presenti un equivalenza numerica con il sistema che si vuole studiare, la seconda è rappresentata dalla rilevazione di alcune esperienze casuali i cui risultati vengono poi analizzati secondo le procedure di natura statistica.

L’applicazione di tale metodo si presta a due diversi tipi di lettura, secondo quindi due diversi approcci. Da un lato esso permette lo studio di sistemi, ed in particolare la valutazione di certe grandezze, mediante una serie di esperienze casuali basate su di un modello probabilistico che non ha nessuna somiglianza relazionale con il sistema di studio ma presenta soltanto un equivalenza di natura numerica. Dall’altro il metodo affronta lo studio di sistemi aventi un comportamento probabilistico mediante la creazione di un modello che presenta delle evidenti analogie relazionali con il sistema di studio, per cui l’equivalenza relazionale non è altro che la necessaria conseguenza.

Ciò che è importante notare è che, qualsiasi sia l’approccio che venga utilizzato, esso ritrova sempre il suo fondamento logico ed analitico nella convergenza stocastica.

La generazione di numeri casuali è uno dei problemi più importanti in termini di simulazione numerica, e l’applicazione del metodo Montecarlo alla risoluzione di tale problema risulta essere, ad oggi, una delle migliori soluzioni. Attorno alle sue linee guida sono stati creati tanti algoritmi per una soddisfacente generazione di numeri casuali, e non è questa la sede per una loro disquisizione, ma basti sapere che esistono due indirizzi di studio per la realizzazione di tali algoritmi. Il primo indirizzo parte dalla constatazione dell’inefficienza dei primi metodi usati per la generazione di numeri casuali, detti anche metodi “congruenziali” in quanto ricavati da formule matematiche definite a priori e, a partire da un certo momento, ripetitivi; il problema diviene quindi quello di determinare i parametri che controllano la frequenza, la ripetizione e la distanza (il cosiddetto *periodo di ritorno*) tra due risultati uguali in modo da allungare il periodo di ritorno e, possibilmente, massimizzarlo. Il secondo indirizzo è invece orientato ad ottenere delle sequenze di numeri più soddisfacenti nei confronti dei test di casualità, semplicemente utilizzando e modificando di volta in volta una combinazione di vari metodi congruenziali.

Nella generazione di numeri casuali a distribuzione di probabilità prefissata ci sono delle relazioni fondamentali che legano i numeri casuali S_i , aventi una certa distribuzione di probabilità, ed i numeri R_i , con una distribuzione uniforme nell’intervallo (0,1). La possibilità di ottenere tali relazioni risiede nel seguente teorema:

“se una variabile casuale ξ ha una funzione densità di probabilità $f(x)$, allora la distribuzione della quantità:

$$\eta = \int_{-\infty}^{\xi} f(x)dx$$

è uniforme nell’intervallo $(0,1)$ ”.

Sulla base di questo teorema possiamo stabilire la seguente regola: *per ottenere un numero che appartenga all’insieme di numeri casuali S_i aventi una funzione densità di probabilità $f(x)$, si deve risolvere rispetto a S_i l’equazione:*

$$R_i = \int_{-\infty}^{S_i} f(x)dx$$

Se ammettiamo di disporre di un generatore di numeri aventi una distribuzione uniforme, la relazione sopra riportata in taluni casi può essere utilizzata direttamente per generare dei numeri casuali con distribuzione di probabilità prefissata. I casi sono quelli in cui tale equazione è risolvibile in modo esatto. Per tutti gli altri casi, come ad esempio quando si ha una distribuzione Gaussiana, si può fare in modo da ricadere nelle condizioni per cui sono validi alcuni teoremi limite del calcolo della probabilità

Nel nostro caso, una distribuzione normale di media μ e varianza σ , possiamo sfruttare il teorema del limite centrale per cui, sotto alcune condizioni molto generali, le somme di un gran numero di variabili casuali sono asintoticamente distribuite secondo una Gaussiana. È possibile impiegare l’addizione di variabili rettangolari per simulare approssimativamente dei numeri casuali distribuiti in modo normale. Poiché in questo

caso le componenti hanno una distribuzione uniforme nell’intervallo (0,1), si sa dal teorema limite per variabili casuali identicamente distribuite che se le variabili casuali indipendenti $\xi_1, \xi_2, \dots, \xi_n$ hanno tutte la stessa distribuzione di probabilità e se ogni ξ_i ha media μ^* e varianza σ^* allora la somma delle variabili ξ_i :

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n$$

È asintoticamente normale con media pari a :

$$\mu = \mu^* \cdot n$$

e scarto quadratico medio pari a :

$$\sigma^2 = \sigma^* \cdot \sqrt{n}$$

Si può dimostrare che la somma ξ ha una distribuzione molto vicina a quella normale anche per valori di n relativamente piccoli.

3.3 Test d’ipotesi T-Student

La validità statistica dei numeri estratti in modo casuale e distribuiti secondo una distribuzione normale viene verificata secondo uno dei test parametrici più comuni in ambito trasportistico: il test *t-Student*, così chiamato perché utilizza la statistica *t*.

I test parametrici non fanno altro che verificare un’ipotesi formulata in precedenza tramite un esperimento; per ipotesi statistica intendiamo un’affermazione relativa ad una popolazione statistica, riguardante di norma una o più caratteristiche o qualità di quest’ultima. Se i risultati ottenuti dall’esperimento risultano essere probabili sotto detta ipotesi, questa non viene respinta, non essendoci un’evidenza statistica per farlo.

Viceversa, se i risultati costituiscono eventi poco probabili, l’ipotesi è respinta. È importante notare che respingere o meno un’ipotesi a seguito di un test statistico non vuol dire smentirla o meno con certezza, visto che si tratta comunque di un’induzione e non di una deduzione.

In generale, nelle indagini sperimentali, viene spesso formulata un’ipotesi di lavoro H_1 che può, ad esempio, derivare direttamente dalla teoria che si sta esaminando. Essa viene di solito espressa in termini di differenza di parametri, ad esempio $H_1=(m \neq n)$, oppure viene ipotizzata anche la direzionalità di tale differenza, ad esempio $H_1=(m > n)$, il che induce ad un affinamento dell’ipotesi H_1 , in quanto, restringendosi il campo dei valori accettati, c’è una maggiore probabilità di accettarla qualora essa sia effettivamente vera. Viene solitamente utilizzato l’approccio di Neyman e Pearson secondo i quali, per verificare l’ipotesi H_1 si deve innanzitutto definire l’ipotesi alternativa H_0 , detta nulla in quanto indica l’assenza di ciò che prevede l’ipotesi H_1 , ad esempio può essere, in contrapposizione agli esempi fatti, $H_0=(m = n)$.

All’ipotesi H_0 è legata la *significatività* di un test; infatti questo risulta tanto più significativo (nel rigettare l’ipotesi H_0) quanto minore è la probabilità α , detta *rischio di errore di I specie*, di rigettare l’ipotesi H_0 , qualora questa fosse vera. Per cui la probabilità $1-\alpha$ è chiamata *livello di significatività* del test.

Successivamente, ma in verità anche prima di raccogliere i dati, si deve individuare l’insieme dei possibili risultati estremi, ossia le zone di rigetto, nelle quali il risultato ha

una probabilità molto bassa di verificarsi se l’ipotesi H_0 fosse vera. Questa probabilità deve essere così bassa che, se il risultato dell’indagine sperimentale cadrà in detta regione, tra le seguenti alternative:

- Ritenere che si sia verificato un evento molto raro ma comunque compatibile con H_0
- Ritenere che ciò che si è verificato è verosimilmente una conseguenza di una realtà non conforme a H_0 bensì ad H_1

non si dovrebbero avere dubbi nello scegliere la seconda.

Possiamo dire quindi che quanto più è bassa α , tanto minore sarà la probabilità di commettere un errore di I specie (essendo proprio pari ad α), nel caso di rigetto di H_0 .

Nel nostro caso, come già detto, la bontà di stima dei coefficienti può essere appurata mediante il test t-Student che verifica l’ipotesi H_0 che un coefficiente β_k sia pari a zero e che invece la sua stima β_k^{ML} sia diversa da zero, in conseguenza di alcuni errori campionari:

$$t = \frac{\beta_k^{ML}}{Var[\beta_k^{ML}]^{1/2}}$$

Oppure, equivalentemente, che due coefficienti β_k e β_j siano uguali:

$$t = \frac{\beta_k^{ML} - \beta_j^{ML}}{[Var(\beta_k^{ML}) + Var(\beta_j^{ML}) - 2 \text{cov}(\beta_j^{ML} \beta_k^{ML})]^{1/2}}$$

La statistica t-Student è distribuita secondo una variabile t-Student con un numero di gradi di libertà pari alla dimensione del campione meno il numero dei coefficienti stimati. Spesso però, se la numerosità del campione lo consente, si assume una distribuzione secondo una normale Standard $N(0,1)$, che è la distribuzione a cui tende la variabile t-Student al crescere del numero dei gradi di libertà; i valori delle varianze e covarianze campionarie possono essere dedotti dalla proprietà della matrice di dispersione Σ , che tende asintoticamente all’inversa della matrice Hessiana calcolata nel punto $(\beta, \theta)_{ML}$ e cambiata di segno.

Per l’ipotesi H_0 assunta, è stata scelta una probabilità $\alpha = 0,95$ di commettere l’errore di rifiutare un ipotesi vera se i valori della statistica t-Student sono esterni all’intervallo di estremi $(t_{\alpha/2}, t_{1-\alpha/2})$, e questi estremi, nel nostro caso, valgono $\pm 1,96$.

Data l’incertezza dei risultati che sarebbero potuti uscire in termini di adattabilità all’istruzione della rete neurale, è stato fatto un ulteriore test con la stessa ipotesi H_0 ma con un $\alpha = 0,50$ che porta a restringere l’intervallo a $t = \pm 0,68$. In effetti ci si è posti il problema che dati troppo dispersi sarebbero stati non significativi per l’istruzione della rete e, probabilmente, non attinenti alla realtà, per cui con questo nuovo intervallo di confidenza si prendono in considerazione tutte le matrici che hanno parametri molto vicini alla matrice originaria che, ricordiamo, risulta essere, per le assunzioni fatte, media delle osservazioni rilevate. Queste matrici, quindi saranno molto simili alla

matrice originaria, ma non è detto che generino, nella fase di assegnazione, flussi sui rami molto simili ai flussi di origine, essendo l’assegnazione di tipo stocastico.

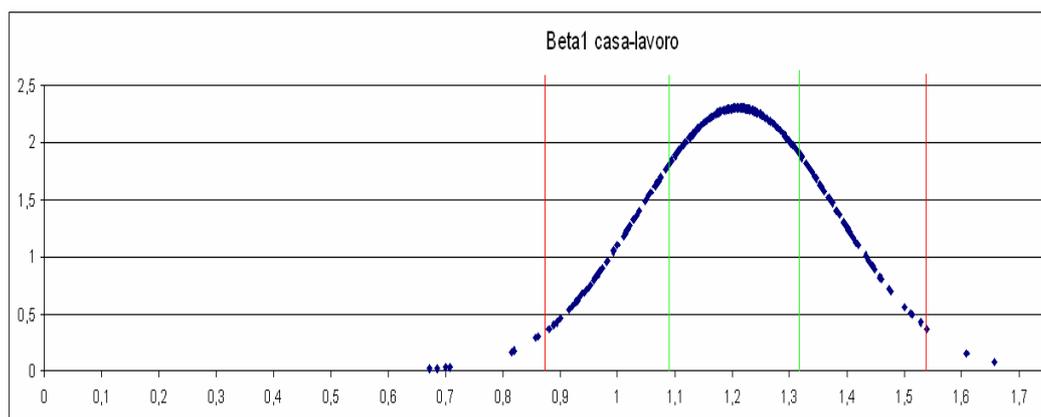
In questo modo, dopo aver fatto le dovute sperimentazioni sulla rete, si potrà capire in che direzione ci si può muovere nella produzione di matrici adatte all’istruzione della rete neurale, ossia, in sostanza, se parametri molto vicini ai parametri originari generano una buona varietà di flussi oppure no.

Più chiaramente, quindi, possiamo scrivere:

$$\beta^{\text{sup}} = \beta_k^{ML} + t_{1-\alpha/2} \cdot \text{Var}[\beta_k^{ML}]^{1/2}$$
$$\beta^{\text{inf}} = \beta_k^{ML} + t_{\alpha/2} \cdot \text{Var}[\beta_k^{ML}]^{1/2}$$

dove $t_{\alpha/2}$ è pari a -1,96 e $t_{1-\alpha/2}$ pari a 1,96 nel caso di $\alpha = 0,95$, ovvero $t_{\alpha/2}$ è pari a -0,68 e $t_{1-\alpha/2}$ pari a 0,68 nel caso di $\alpha = 0,50$.

In questo modo possiamo visualizzare direttamente sulle distribuzioni quelli che sono i limiti entro i quali possiamo ritenere di non aver commesso errori di valutazione. A titolo di esempio si riporta la distribuzione dell’estrazione un parametro analizzato nel presente lavoro con i rispettivi limiti. In rosso c’è la statistica con $t = \pm 1,96$, in verde quella con $t = \pm 0,68$.



Come si può vedere, così facendo abbiamo la possibilità di controllare tre tipi di comportamento: il primo con tutti i parametri, anche quelli esterni alla statistica, il secondo con i parametri che confermano il test d’ipotesi con $\alpha = \pm 1,96$ e il terzo con parametri che confermano il test d’ipotesi con $\alpha = \pm 0,68$. E’ importante comunque ricordare che queste considerazioni così come i risultati ottenuti hanno una loro validità solo nel caso analizzato e vanno ripetuti per ogni nuovo caso.

Nonostante la mole di lavoro necessaria per preparare una prova del genere, questa operazione consente di stimare la bontà della risposta della rete neurale ad una matrice i cui parametri ricadono in una statistica ovvero nell’altra e quindi di scegliere appropriatamente i parametri da utilizzare.

3.4 Creazione del training set

Come più volte ripetuto, l’addestramento delle reti neurali necessita di un set di dati piuttosto numeroso e diversificato, che, in genere, nel caso di aggiornamento della matrice OD mediante conteggi di flusso, non si ha. Si rende quindi indispensabile

creare un set di dati “in laboratorio”, mediante una tecnica che rispecchi le ipotesi dei modelli utilizzati per la definizione delle matrici di viaggio.

Fino ad oggi la tecnica più sofisticata per generare le matrici è stata l’estrazione pseudocasuale degli elementi della matrice OD, implementata dai software più comuni, o, nei casi di rete di studio piccola, si è proceduto alla stima diretta della matrice OD mediante indagini su targa.

La notevole approssimazione nel caso di implementazione pseudocasuale induce l’operatore a porre notevole attenzione ai vettori di domanda generati con tale metodo, dovendone giudicare la bontà e la veridicità di tale dato nonché l’adattabilità alle diverse esigenze che si riscontrano durante la fase di training o di validazione del modello. In effetti non vi è alcun controllo sui dati generati, e si rende necessario quanto meno definire dei controlli di tipo esogeno, quali ad esempi il rispetto del numero di abitanti per zona; capita sovente, infatti, che da questo tipo di estrazione, il numero di persone che si spostano da O a D di mattina sia superiore al numero di abitanti della zona O!

È opportuno notare che la matrice OD rappresenta un’immagine della distribuzione spaziale della domanda corrispondente ai flussi che si rilevano sulla rete. In particolare, tra le infinite matrici cui corrisponde l’insieme dei flussi misurati sulla rete, la matrice stimata secondo le consolidate procedure descritte in precedenza risulta essere la migliore stima. I valori dei coefficienti di calcolo della nuova matrice OD sono,

statisticamente, stima di valori attesi. Se gli attributi sono considerati come valori deterministici, il corrispondente valore di ogni cella della matrice OD rappresenta un valore medio della domanda sulla coppia OD considerata. In quest’ottica, considerando variabili aleatorie i coefficienti degli attributi, esisteranno infinite matrici OD corrispondenti ai valori probabili dei coefficienti in dipendenza delle loro funzioni di probabilità. A tali matrici “probabili” corrisponderanno diverse configurazioni di flussi che, così come le matrici, hanno una loro probabilità di verificarsi.

Partendo da queste ipotesi, si è sviluppato il lavoro di tesi, nell’ottica di creare un insieme di addestramento della rete neurale (training set) quanto più vicino alle reali condizioni i rete, flussi, domanda. Essendo, infatti, le reti neurali un particolare sistema regressivo, legano l’affidabilità della loro risposta alla “bontà” della base dati necessaria per la sua calibrazione. In particolare, per quanto detto precedentemente, si intende procedere mediante un’estrazione casuale di parametri caratteristici della domanda mediante il metodo Montecarlo, all’interno quindi della loro distribuzione statistica.

Dalla Teoria dei Sistemi di Trasporto la matrice OD risulta essere dipendente, dai parametri (X^i, β, θ) . I valori di tali parametri sono frutto di una stima di un campione degli stessi, e pertanto hanno una loro distribuzione statistica e, inoltre, per una condizione di omogeneità, sono riferiti ad una precisa fascia oraria, relativamente alla stessa modalità di spostamento e ad una stessa categoria di persone, per un certo motivo. Il metodo di stima ML (Maximum Likelihood) dei β definisce questi distribuiti

secondo una *t-Student*, con media pari a β^{ML} e varianze facilmente calcolabili tramite la proprietà della matrice Hessiana. Si ricorda che la variabile *t-Student* è la distribuzione dei valori attesi, essendo solo stimata la varianza. Analogamente in una matrice OD derivante da una stima diretta dei flussi, come le sottomatrici di Scambio-Attraversamento, si può ipotizzare che la domanda sulla generica coppia OD sia la media di un certo numero di osservazioni ovvero di indagini campionarie, che rispettano sempre la condizione di omogeneità, ossia riferiti ad un certo intervallo di tempo, secondo una certa modalità di spostamento e una certa categoria di persone, per un certo motivo, e che la varianza di tali osservazioni sia proporzionale alla distanza quadratica di tali osservazioni rispetto alla media, ovvero, secondo quanto detto prima, pari a:

$$\hat{s}^2 = \frac{\sum_{i=1..n} (n_{OD}^i - \bar{n}_{OD})^2}{(n-1)}$$

Viene inoltre fatta l’ipotesi che le domande Interno-Esterno, Esterno-Interno e Esterno-Esterno, siano state stimate in maniera diretta, tramite delle indagini al cordone.

Per la domanda Interno-Interno, che si esaurisce all’interno dell’area di studio, andremo essenzialmente ad agire sui parametri calibrati nei modelli di emissione e di distribuzione.

I primi forniscono la probabilità di ottenere un numero medio di spostamenti $m^i[osh]$ per la categoria i , per il motivo s , nel periodo h , a partire dall’origine degli spostamenti

o . I secondi forniscono la probabilità che un utente della categoria i , per un motivo s , da un’origine o , in un periodo temporale h , si rechi nella zona d .

Il modello di emissione utilizzato è del tipo “*indici per categoria*” e restituisce il numero di spostamenti che la popolazione compie per uscire dalla zona o , per il motivo s , nel periodo h ; tale numero risulta essere la sommatoria estesa a tutte le categorie del prodotto tra il numero di persone appartenenti ad una categoria (operai, professionisti, studenti, famiglie, etc.), per l’indice medio di spostamenti che effettua la singola persona che appartiene a tale categoria.

Il modello di distribuzione utilizzato è un modello di distribuzione del tipo gravitazionale:

$$p[d / osh] = \frac{A_d^{\beta_1} \cdot C_{od}^{-\beta_2}}{\sum_d A_d^{\beta_1} \cdot C_{od}^{-\beta_2}}$$

Dove A_d è la variabile di attrattività e C_{od} è la variabile di costo, mentre β_1 e β_2 sono i parametri calibrati per le diverse categorie di utenti, di motivi e di periodi di tempo considerati.

La tecnica di validazione dei modelli sopra citati ci dà l’opportunità di poter considerare i valori di m^i , di β_1 e di β_2 trattandole come delle medie di una serie di rilevamenti effettuati, distribuite secondo una variabile *t-Student*, la cui media è proprio il valore originario noto a priori; in altre parole i valori dei parametri della matrice originaria possono essere considerati come delle medie di una distribuzione di medie.

L’estrazione casuale dei valori contenuti all’interno di tale distribuzione risulta essere, quindi, un’estrazione casuale di medie probabili con densità di probabilità minore secondo la legge di distribuzione stocastica considerata (*t-Student*). Tuttavia l’estrazione casuale, condizionata al rispetto della lontananza dai limiti di confidenza fissati dalla distribuzione del test *t-Student*, ci permette di estrarre un numero di parametri che, sebbene siano lontani dalla media, risultano essere sicuramente possibili; in altre parole stanno ad indicare che è stato scelto un parametro rappresentativo di uno scenario possibile.

Analogamente, per quanto riguarda le sottomatrici di scambio Esterno-Interno, di scambio Interno-Esterno e di attraversamento Esterno-Esterno, si fa riferimento ad una stima diretta di tali valori, attraverso un’indagine al cordone con un campionamento stratificato sul quarto d’ora, su più giorni equivalenti. L’espressione “giorni equivalenti” è spesso utilizzata nella progettazione delle indagini campionarie su strada e sta ad indicare quei giorni che, per considerazioni fatte sul tipo di domanda considerata e sulle caratteristiche del sistema di trasporto, possono considerarsi equivalenti; in altre parole si assume che i campionamenti svolti, con le stesse modalità, in giorni equivalenti siano osservazioni di uno stesso tipo di domanda. Si può supporre, in questo caso che la distribuzione dei valori osservati sia di tipo Normale con media pari proprio ai valori della matrice nota a priori mentre una stima della varianza di detti valori può essere ottenuta, attraverso i valori delle singole celle relative a detta sezione

e relativi alle singole osservazioni; tale posizione è possibile in quanto si suppone, peraltro correttamente, un’uniformità di svolgimento, ossia di campionamento, delle indagini effettuate al limite dell’area di studio (cordone).

Le ipotesi sui possibili modi su come far variare queste parti della matrice OD possono essere molteplici, così come i possibili motivi che inducono tale tipo di variazione. Si potrebbe andare ad intervenire sulla generazione dei singoli centroidi esterni e, volendo, anche sulla distribuzione dei flussi che dal centroide esterno si dirigono verso gli altri centroidi; ad esempio considerando costante il coefficiente di correlazione C_v relativo alla generica coppia OD:

$$C_{V_{od}} = \frac{\hat{s}_{OD}}{\hat{d}_{OD}} \quad \text{o anche spesso} \quad C_{V_{od}} = \frac{\hat{d}_{OD}}{\sqrt{Var[\hat{d}_{OD}]}}$$

che altro non è che un rapporto tra la media di alcune osservazioni statistiche e la radice della loro varianza. Oppure in un ottica ancor più disaggregata ma con una gran mole di lavoro, forse non necessario per tutte le approssimazioni insite nei modelli, si può andare a far variare il singolo elemento della matrice secondo la sua distribuzione statistica ipotizzata.

Quindi, congruentemente a quanto fatto per la matrice interna, si considera la legge di distribuzione statistica con cui variano i valori dei flussi al cordone in entrata e in uscita; tale scelta è dettata dal fatto che, in ipotesi di applicazioni ITS, si ipotizza l’installazione di semplici contatori di flusso in ingresso e in uscita dalle sezioni. Se, quindi, si considera ogni somma di valori di ciascuna riga o colonna delle sottomatrici

come realizzazioni statistiche delle variabili “flusso in uscita o in entrata dall’area di studio”, è possibile considerare tale osservazione come una variabile normale con media e varianza note. La media, infatti, in questo caso, è proprio pari alla somma dei valori delle celle dell’*i*-esima riga della matrice OD nota a priori, mentre la varianza è ricavabile, secondo la sua definizione matematica, tenendo presente i valori delle somme dell’*i*-esima riga delle singole osservazioni avvenute nei giorni equivalenti. In altre parole:

$$Var[\hat{d}_{i\bullet}] = \frac{\sum_{g=1\dots G} (\hat{d}_{i\bullet}^g - \hat{d}_{i\bullet})^2}{G - 1}$$

Dove *G* è il numero di giorni equivalenti utilizzato, o anche il numero di osservazioni, $\hat{d}_{i\bullet}^g$ è la somma dei flussi entranti o uscenti dalla singola sezione *i* durante l’osservazione *g*, mentre $\hat{d}_{i\bullet}$ è la media della somma dei flussi entranti o uscenti dalla sezione *i* in relazione alle *G* osservazioni, in altre parole il valore che si ritrova nella matrice nota a priori. La ripartizione sulle singole celle dell’*i*-esima riga verranno fatte in riferimento alla proporzionalità tra singola coppia OD e la variabile “flusso in entrata o in uscita dall’area di studio” nota a priori relativa alla singola sezione.

Le statistiche dei parametri scelti così definite, in termini di media e varianza, sono sufficienti ad effettuare un estrazione casuale di valori di detti parametri all’interno della loro distribuzione statistica; in altre parole risulterà che la probabilità di estrazione di un dato valore di un parametro è tanto più elevata quanto più è vicino tale valore alla

media del parametro stesso (metodo Monte Carlo). Il risultato di tale estrazione ha una duplice valenza per la presente applicazione:

- 1 Le estrazioni fatte sui singoli parametri, tutti indipendenti tra loro e dotati ognuno di una propria statistica definita (media e varianza), non generano correlazioni pericolose. Gli esempi, composti da tali parametri, sono effettivamente da considerare come osservazioni indipendenti del fenomeno indagato, senza legami con altri esempi se non per l’appartenenza alla stessa legge di probabilità.
- 2 La scelta di intervenire su parametri rappresentativi della domanda risulta essere conveniente oltre che per un minor numero di parametri da considerare, il che rende l’applicazione su ANN più semplice e veloce, anche per il maggior controllo sul risultato delle estrazioni. Vengono evitate in tale maniera alcune incongruenze che spesso si verificano in particolare durante le estrazioni dei valori delle celle dove, ad esempio, accade che il numero di emessi da una zona è superiore al numero di residenti!

Vengono così costruiti dei vettori, rappresentativi della domanda di mobilità, composti da parametri della domanda interna e della domanda esterna. Le matrici, esplicitazioni dei vettori di parametri di domanda creati, attraverso il processo di assegnazione alla rete generano altrettanti vettori di flussi. In tal modo viene così definito il set di apprendimento della ANN, basato appunto su vettori di parametri rappresentativi della domanda e di relativi flussi su strada da essi generati. Prima della fase di training della

ANN, il set di dati destinato all’apprendimento viene standardizzato; in altre parole il singolo dato viene manipolato sottraendogli una misura di allocazione e dividendolo per una misura di scala. Questa pratica, nel caso di valori estratti da una variabile Gaussiana, corrisponde alla standardizzazione di tali valori secondo quella che è la comune definizione statistica, ovvero sottraendo la media e dividendo per la deviazione standard. Nonostante si aggiunga un processo in più nell’elaborazione, è una pratica largamente utilizzata in letteratura e gli effetti che derivano dalla sua applicazione sono una maggiore velocità di esecuzione e un miglior comportamento di risposta nella fase di training.

4 Applicazioni a dati reali

La procedura di creazione del training set così sviluppata, è stata testata e validata su più tipi di ANN e in diversi contesti urbani, reali e simulati. La metodologia proposta è stata inoltre comparata con la metodologia usualmente utilizzata per la creazione del training set, basata sulla definizione di diverse percentuali e sull’incremento percentuale di tutte le singole celle secondo ogni percentuale definita. Si viene così a definire un training set di dati tutti linearmente dipendenti secondo una percentuale. Questo tipo di procedura ha evidentemente dei limiti sulla risposta nel caso in cui si presentano dati reali in caso di simulazione dove è molto improbabile che la regola di associazione imparata (tutti i coefficienti proporzionalmente dipendenti tra di loro) sia l’effettiva risposta su dati reali, ma non è detto che non possa trovare applicabilità in alcuni casi

semplici o particolari. Volendo simulare tale tipo di metodologia a partire dal tipo di dati che si è scelto di utilizzare, si può affermare che tale tipo di metodologia consiste in effetti nell’estrazione (o anche scelta casuale) della percentuale d’incremento dei soli coefficienti di generazione e che ogni percentuale di incremento/decremento rispetto ai coefficienti di generazione noti a priori dev’essere la stessa per ogni coefficiente.

I siti scelti per i test sono il quartiere di Fuorigrotta, nella città di Napoli, e la cittadina di Acerra nella periferia di Napoli. La scelta è ricaduta su queste due località per la quantità di dati a disposizione, in quanto sono da sempre abbondantemente studiate e prese a modello per numerose esperienze.

Il quartiere di Fuorigrotta, nella città di Napoli, per la sua posizione, per i servizi al suo interno e per i collegamenti con l’esterno risulta essere un’area particolarmente delicata. Si trova, difatti, in una posizione al limite dell’area urbana di Napoli, ed è un punto di accesso/uscita alla parte Nord-Ovest della città, attraverso il vecchio tracciato della Domiziana che lo connette al comune di Pozzuoli e attraverso la tangenziale di Napoli su cui c’è un uscita dedicata. Le due grotte, una per senso di marcia, che attraversano la collina di Posillipo e che la mettono in comunicazione con i quartieri di Chiaia San Ferdinando e Posillipo, costituiscono la principale connessione di questo quartiere all’area urbana centrale di Napoli. Inoltre per la configurazione orografica del sito, risulta essere la principale via d’accesso al quartiere di Bagnoli e all’area di trasformazione urbana Bagnolifutura.

I servizi al suo interno sono molteplici: è presente l’Università degli studi di Napoli Federico II con la Facoltà di Ingegneria, il centro espositivo Mostra d’Oltremare, lo stadio San Paolo, la sede della RAI, oltre a numerose scuole, uffici, sedi distaccate di altre facoltà universitarie.

La cittadina di Acerra si trova nella parte Nord Est della provincia di Napoli, ed è il terzo comune per vastità territoriale, dopo Napoli e Giugliano. È una cittadina dai contorni ben definiti, al contrario di quanto spesso si rileva scorrendo la mappa della provincia di Napoli, dove spesso comuni limitrofi si “compenetrano” formando un unicum indefinito. È una cittadina molto ben connessa con le infrastrutture stradali extraurbane, avendo accesso a Nord all’asse di supporto e a Sud all’asse mediano. Questi due assi stradali la collegano direttamente con le autostrade A16 Napoli-Canosa Bari, A30 Caserta–Salerno e A1 Napoli-Roma.

Acerra è una cittadina che considerevolmente può considerarsi autonoma avendo un gran numero di aziende nei vari settori industriali, basti pensare che nella sola area comunale insistono ben 2000 aziende agricole. Inoltre, stando ai dati del censimento 2001, la popolazione si attesta sulle 45mila unità.

I motivi di scelta di questi due siti risiedono, oltre che nella facile reperibilità di dati, nella necessità di utilizzare diverse tipologie di rete urbana e diverse tipologie di domanda. Difatti la domanda del quartiere di Fuorigrotta, dalle considerazioni precedenti, risulta essere una domanda dove è preponderante la parte di Scambio e

Attraversamento, mentre al contrario la cittadina di Acerra risulta avere una domanda prevalentemente interna, in quanto la concentrazione di beni e servizi e la sua concentrazione urbana permettono di avere un sistema delle attività fortemente concentrato nell’area stessa.

	COEFFICIENTI	MEDIA	DEV. STD.	DESCRIZIONE VARIABILE
INTERNI	A1	0.65	0.108333	generazione Casa lavoro
	A2	0.28	0.108333	generazione Casa Altri motivi
	B1	1.21	0.172857	distribuzione addetti Casa lavoro
	B2	1.83	0.653571	distribuzione costo Casa lavoro
	B3	0.68	0.172166	distribuzione addetti Casa altri motivi
	B4	3.21	0.243182	distribuzione costo Casa altri motivi
ESTERNI	C1	1701	413.514411	flussi in entrata sez. 1042
	C2	3291	704.134650	flussi in entrata sez. 1041
	C3	5997	1247.544423	flussi in entrata sez. 1040
	C4	1749	394.287902	flussi in entrata sez. 1039
	C5	3374	722.257396	flussi in entrata sez. 1038
	D1	4248	877.389661	flussi in uscita sez. 1038
	D2	2723	610.778221	flussi in uscita sez. 1039
	D3	3838	785.699129	flussi in uscita sez. 1040
	D4	2250	452.530369	flussi in uscita sez. 1041
	D5	1016	248.935921	flussi in uscita sez. 1042

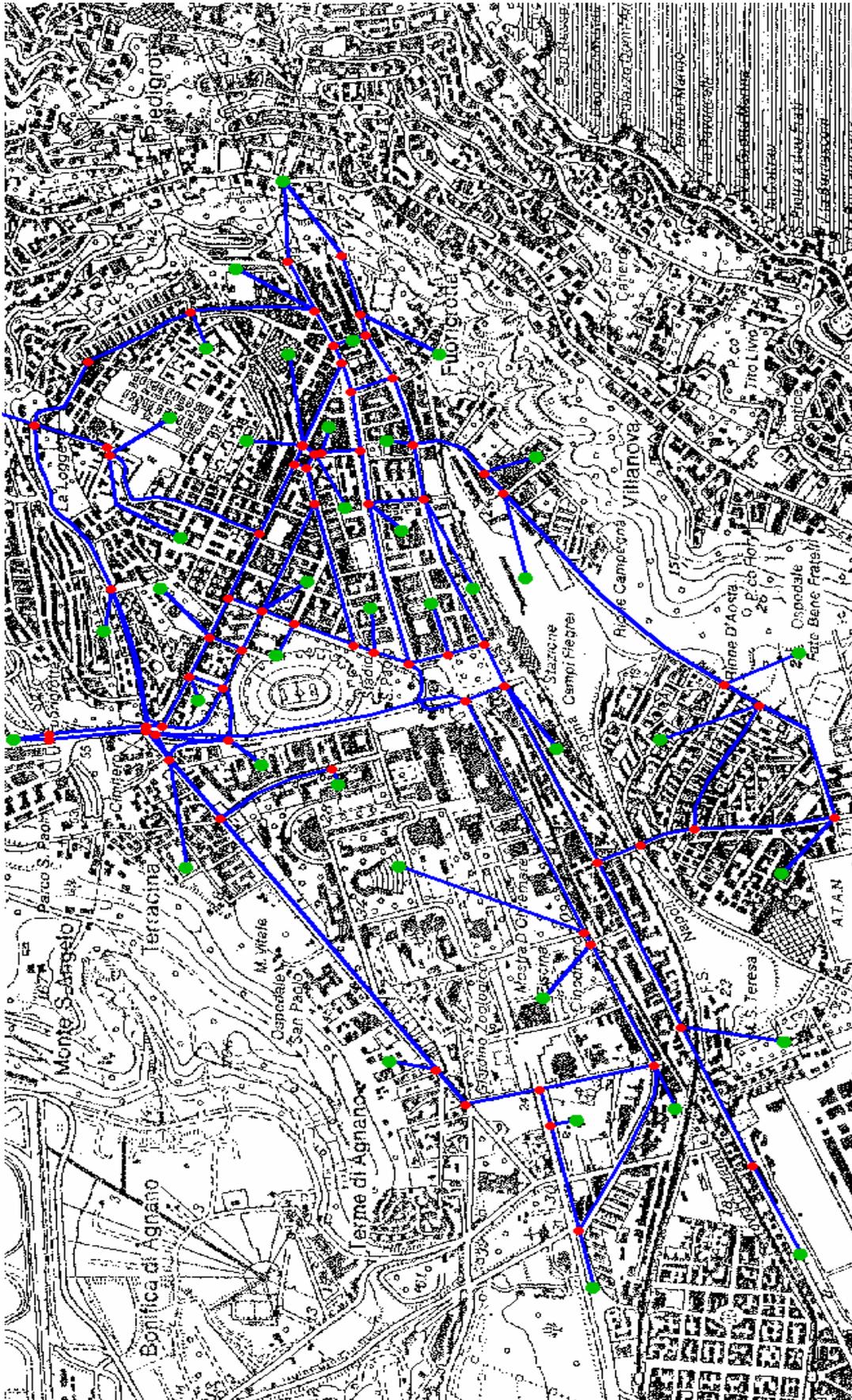
Tabella riassuntiva sui dati di Fuorigrotta

	COEFFICIENTI	MEDIA	DEV. STD.	DESCRIZIONE VARIABILE
INTERNI	A1	0.65	0.108333	generazione Casa lavoro
	A2	0.28	0.108333	generazione Casa Altri motivi
	B1	1.21	0.172857	distribuzione addetti Casa lavoro
	B2	1.83	0.653571	distribuzione costo Casa lavoro
	B3	0.68	0.172166	distribuzione addetti Casa altri motivi
	B4	3.21	0.243182	distribuzione costo Casa altri motivi
ESTERNI	C1	170.1	41.351441	flussi in entrata dal cordone sez. 1042
	C2	329.1	70.413465	flussi in entrata dal cordone sez. 1041
	C3	599.7	124.754442	flussi in entrata dal cordone sez. 1040
	C4	174.9	39.428790	flussi in entrata dal cordone sez. 1039
	C5	337.4	72.225740	flussi in entrata dal cordone sez. 1038
	D1	424.8	87.738966	flussi in uscita dal cordone sez. 1038
	D2	272.3	61.077822	flussi in uscita dal cordone sez. 1039
	D3	383.8	78.569913	flussi in uscita dal cordone sez. 1040
	D4	225	45.253037	flussi in uscita dal cordone sez. 1041
	D5	101.6	24.893592	flussi in uscita dal cordone sez. 1042

Tabella dati di Fuorigrotta” modificata”

	COEFFICIENTI	MEDIA	DEV. STD.	DESCRIZIONE VARIABILE
INTERNI	A1	0.3584	0.059733	generazione Casa lavoro addetti industria
	A2	0.4358	0.072625	generazione Casa lavoro addetti servizi pubblici
	A3	0.3259	0.054308	generazione Casa lavoro addetti servizi privati
	A4	0.7056	0.117600	generazione Casa scuola elementari
	A5	0.7308	0.121800	generazione Casa scuola medie inferiori
	A6	0.7224	0.120400	generazione Casa scuola medie superiori
	A7	0.2850	0.110465	generazione Casa altri motivi
	B1	1.1000	0.157143	distribuzione addetti Casa lavoro industria
	B2	0.9300	0.132857	distribuzione addetti Casa lavoro servizi pubblici
	B3	0.9300	0.132857	distribuzione addetti Casa lavoro servizi privati
	B4	0.9000	0.128571	distribuzione alunni Casa scuola elementari
	B5	0.9500	0.135714	distribuzione alunni Casa scuola medie inferiori
	B6	1.0000	0.142857	distribuzione alunni Casa scuola medie superiori
	B7	0.9100	0.227500	distribuzione addetti Casa altri motivi
	C1	0.7000	0.250000	distribuzione costo Casa lavoro industria
	C2	0.5800	0.207143	distribuzione costo Casa lavoro servizi pubblici
	C3	0.8300	0.296429	distribuzione costo Casa lavoro servizi privati
	C4	2.5200	0.900000	distribuzione costo Casa scuola elementari
	C5	2.2400	0.800000	distribuzione costo Casa scuola medie inferiori
	C6	0.3500	0.125000	distribuzione costo Casa scuola medie superiori
	C7	0.7800	0.059091	distribuzione costo Casa altri motivi
ESTERNI	D1	630.563	107.893900	flussi in uscita dal cordone sez. 81
	D2	135.744	25.653800	flussi in uscita dal cordone sez. 82
	D3	246.766	17.002500	flussi in uscita dal cordone sez. 83
	D4	142.427	21.601400	flussi in uscita dal cordone sez. 84
	D5	220.161	38.895700	flussi in uscita dal cordone sez. 85
	D6	133.298	24.555400	flussi in uscita dal cordone sez. 86
	D7	156.724	21.353200	flussi in uscita dal cordone sez. 87
	D8	232.352	38.114000	flussi in uscita dal cordone sez. 88
	E1	430.599	66.511300	flussi in entrata dal cordone sez. 81
	E2	129.141	20.012500	flussi in entrata dal cordone sez. 82
	E3	194.872	33.355100	flussi in entrata dal cordone sez. 83
	E4	95.064	18.196700	flussi in entrata dal cordone sez. 84
	E5	213.801	33.135700	flussi in entrata dal cordone sez. 85
	E6	108.742	22.247500	flussi in entrata dal cordone sez. 86
	E7	94.467	16.632500	flussi in entrata dal cordone sez. 87
	E8	275.938	52.332700	flussi in entrata dal cordone sez. 88

Tabella dati di Acerra



Rete urbana del quartiere di fuorigrotta



Rete urbana della città di Acerra

4.1 Le reti neurali artificiali utilizzate

La definizione degli aspetti legati all’utilizzo delle reti neurali artificiali segue quanto riportato in letteratura e nelle esperienze nel campo dei trasporti. In effetti la scelta di una particolare tipologia e in seguito architettura di rete neurale è, per stessa natura dello strumento, un processo basato su esperienze simili, considerazioni generali e tentativi sul caso particolare.

Le reti neurali artificiali considerate hanno tutte un funzionamento di tipo Feed Forward ovvero non hanno collegamenti tra neuroni che generano ciclicità. Le reti che generano ciclicità, le cosiddette Recurrent, non sono presenti in letteratura applicate a casi assimilabili al presente e questo è dovuto al fatto che la loro la loro maggior difficoltà nella fase di definizione e i maggiori tempi computazionali necessari alla fase di addestramento si giustificano in casi in cui si ha necessità di svolgere compiti di previsione di sequenze temporali che sono difficilmente alla portata delle reti Feed Forward.

In base a questa considerazione sono state selezionate le architetture Feed Forward più versatili e maggiormente utilizzate in letteratura in modo da paragonare la loro risposta.

In particolare sono state selezionate le seguenti architetture: il Multi Layer Perceptron (MLP), la Radial Basis Function network (RBF), e la Generalized Regression network (GR). Ognuna ha caratteristiche proprie, oltre alle differenze modellistiche viste in precedenza all’interno di ciascuna rete, e viene utilizzata in contesti ben definiti.

4.1.1 Multi Layer Perceptron

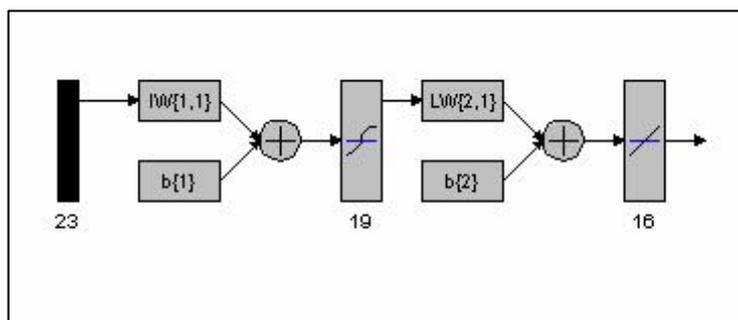
Il Multi Layer Perceptron è di gran lunga l’architettura neurale più utilizzata. Benché ci siano altre architetture in grado di assicurare prestazioni migliori in certe condizioni, il MLP risulta essere ancora largamente preferito per la sua adattabilità e modulabilità ai diversi casi, al maggior numero di controlli che si possono attuare su di esso, e per la sua migliore risposta nei casi di estrapolazione; difatti il MLP risulta avere le migliori capacità predittive qualora l’input che gli si presenta è in valore esterno al range dei casi presentati in fase di addestramento. Diversamente, nei casi in cui l’input presentato è all’interno del range di valori dei dati di addestramento, si parlerà di interpolazione.

La possibilità di inserire più hidden layer, strati di neuroni nascosti, al suo interno dà grosse possibilità di utilizzo ma si fa notare come già l’utilizzo di un MLP con un solo hidden layer e di funzioni di attivazione nonlineari e continue ha in sé la proprietà di approssimatore universale.

Oltre alla definizione dei numeri di strati nascosti, l’architettura "ottima" per la risoluzione di un determinato problema con il MLP viene determinata anche dal numero di neuroni negli strati nascosti e dal numero di connessioni. A tal proposito sono state elaborate diverse procedure e altrettanti algoritmi in modo da determinare anche questi ultimi due aspetti, che altrimenti dovrebbero essere definiti seguendo alcune regole empiriche, attraverso un processo di “trial and error”. Il pruning (letteralmente “potatura”) è una di queste e consiste nella eliminazione di quei pesi della rete che non

servono all'apprendimento. In particolare algoritmi basati sulla tecnica del pruning consentono di stimare la rilevanza di un generico peso ed eventualmente eliminarlo durante l'apprendimento.

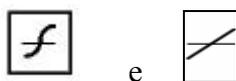
Il diagramma di funzionamento dell’architettura prescelta viene riportato qui sotto:



Schema funzionale dell’architettura scelta per la rete urbana di Fuorigrotta

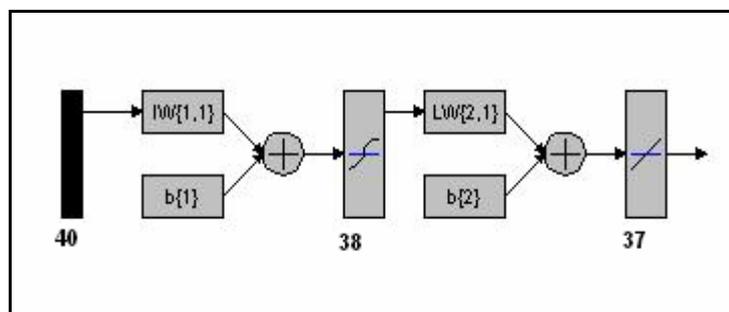
La simbologia adottata è la seguente: **IW** è l’acronimo di Input Weights ed indica i pesi associati al piano di input, **LW** (Layer Weights) indica i pesi associati agli strati hidden, **b** è il vettore dei pesi dei bias, **23**, **19** e **16** sono il numero di neuroni negli strati di neuroni, **23** è la numerosità del vettore di input, **16** è la numerosità del vettore di output.

I simboli :



Rappresentano le funzioni di trasferimento, e quindi rispettivamente una funzione sigmoide di tipo tangente iperbolica che dà alla rete la non linearità dell’associazione e una funzione *pureline* o semplicemente lineare che “raggruppa” le elaborazioni dell’hidden layer. Questa combinazione di funzioni è frutto di una serie di tentativi fatti per la ricerca della migliore combinazione di funzioni, peraltro è confermata in molte

esperienze pubblicate. L’architettura neurale per la cittadina di Acerra è simile a quella utilizzata per Fuorigrotta e differisce solo per la numerosità di neuroni all’interno dei vari layer.



Schema funzionale dell’architettura scelta per la rete urbana di Acerra

4.1.2 Radial Basis Function Network

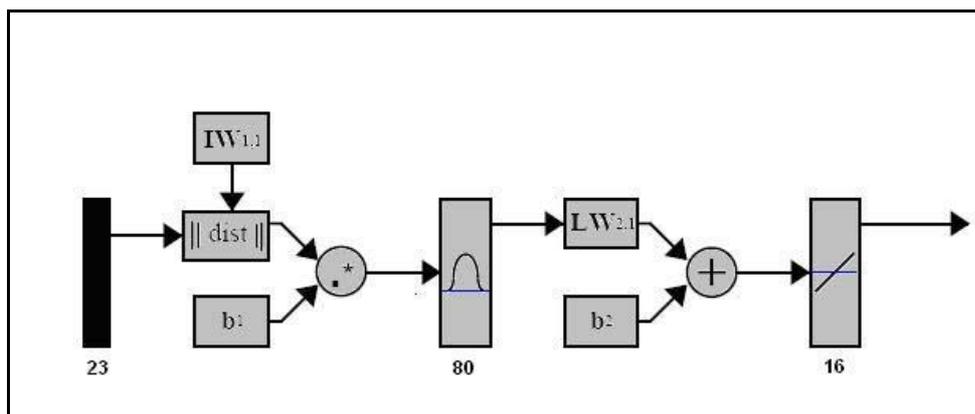
La rete RBF viene così chiamata in quanto le funzioni di trasferimento nelle unità nascoste sono Funzioni a Base Radiale, in modo da tener conto della multidimensionalità dello spazio di valori nei quali si ricerca la risposta.

Come detto precedentemente, il risultato è definito dalla distanza euclidea del vettore di input da un vettore prototipo, definito centro dell’ m -esima funzione radiale. Il numero centri di funzioni a base radiale non è necessariamente uguale al numero N dei dati a disposizione per il training, anzi è tipicamente è molto minore e la determinazione di tali centri è parte del processo di addestramento. La funzione radiale può essere qualsiasi, ma la quasi totalità dei casi presenti in letteratura si riferisce a funzioni di tipo Gaussiano, in quanto è fortemente rappresentativa dell’effetto della distanza del singolo

valore dal centro della funzione. La funzione di attivazione per le funzioni di trasferimento Gaussiano può essere di due tipi: esponenziale o softmax.

La scelta di un numero elevato di centri, e quindi un elevato numero di parametri da definire, può portare ad una sensibilità troppo elevata da parte della rete nei confronti dei dati di allenamento. Questa sensibilità si traduce in una “memorizzazione” dei dati del training set e una diminuzione della capacità interpolante dei dati di allenamento da parte della rete perdendo la caratteristica fondamentale della generalizzazione (overfitting). Il caso estremo di una scelta elevata di centri è quello in cui essi sono posti uguali agli ingressi dell’insieme dell’allenamento. La matrice che si viene a formare è nota come *full design matrix*. Le reti RBF sono abitualmente utilizzate con un solo hidden layer e non sono state riscontrate esperienze significative che riportassero l’utilizzo di tali architetture con più di un hidden layer, tuttavia è bene chiarire che l’utilizzo di un ulteriore strato nascosto prima dello strato con funzioni radiali viene utilizzato per dare modo alla rete di ignorare gli input irrilevanti. Con questa introduzione di un ulteriore strato nascosto, la forma delle funzioni assume non più una base radiale bensì ellittica (Elliptical Basis Function – EBF).

Tuttavia aumentano molto i tempi di addestramento della rete e le richieste di prestazioni dei processori, pertanto si ritiene più utile effettuare un’analisi sul training set e sulla significatività dei parametri nella risoluzione del problema posto che ricorrere a tale tipo di rete.



Schema dell'architettura RBF utilizzata

I simboli utilizzati sono simili a quelli utilizzati per le MLP, così come il funzionamento

che differisce per il calcolo della distanza $\|\text{dist}\|$ tra il vettore generico e il vettore

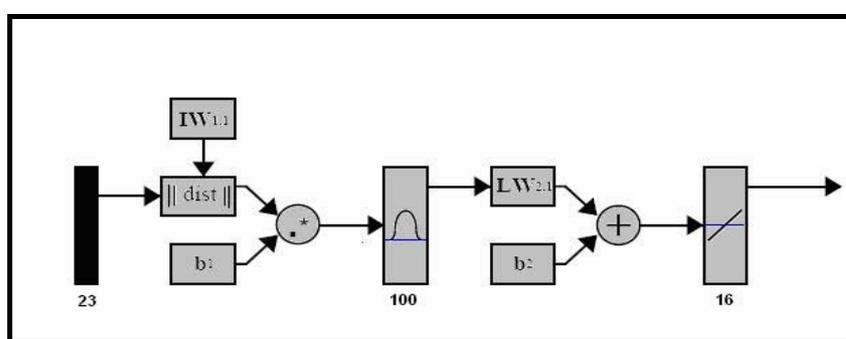
prototipo e per la funzione radiale  di trasferimento.

4.1.3 Generalized Regression Network

Il funzionamento di questo tipo di architettura neurale è molto simile al funzionamento di una rete RBF con la differenza che i centri sono definiti dagli stessi esempi e che le funzioni di attivazione sono funzioni di tipo softmax tali che la somma dei valori di attivazione sia 1.

Le potenzialità di questo tipo di rete neurale, come riportato nel capitolo 1, sono sorprendenti in termini interpolativi ma la risposta di questa rete è ancor più legata alla bontà della base dati rispetto alle altre reti. Tale base deve avere un range che esaurisca tutte i possibili valori che le variabili possono assumere e deve essere uniformemente distribuita lungo tutto il range dei valori assunti. Inoltre è una rete che risulta essere

adatta alla simulazione di funzioni dolci (smooth functions) dimostrandosi in questi casi molto precisa, viceversa necessita di una quantità di dati molto elevata per funzioni che presentano picchi, aumentando di conseguenza i tempi computazionali e le richieste prestazionali nei confronti dei processori. Non si riscontrano casi di utilizzo di più layer all’interno di tale architettura, anzi un ulteriore accorgimento alla base dati è dato dall’impossibilità di non tener conto di casi giudicati irrilevanti.



Schema dell'architettura GR utilizzata

Le notazioni sono uguali a quelle della rete Radial Basis Function, ma cambia fondamentalmente la funzione di attivazione che è di tipo softmax, e il numero di centri attivati per esempio, in questo caso 100, quanti ne sono gli esempi presentati.

4.2 Confronto con la metodologia tradizionale

In questa sezione vengono presentati i risultati relativi al confronto tra la metodologia tradizionale di creazione del training set e la metodologia proposta. Il confronto è stato effettuato sui dati del quartiere di Fuorigrotta, in termini di RMSE% (Root Mean Square Error in Percentage) sui coefficienti stimati in fase di simulazione,

utilizzando le 3 reti neurali in oggetto. La formulazione matematica dell’RMSE%

calcolato sui parametri è la seguente:

$$RMSE\% = \frac{\sqrt{n \cdot \sum (x_i^{st} - x_i)^2}}{\sum x_i^{st}}$$

Dato il diverso funzionamento delle diverse reti, è necessario individuare la migliore architettura neurale per ogni singola tipologia. Tale individuazione è necessaria anche rispetto al contenuto informativo che viene presentato in fase di addestramento, pertanto non deve stupire l’utilizzo di reti neurali diverse per ogni singolo caso.

Il training set creato secondo le modalità proposte, così come quello creato con la metodologia tradizionale, è stato contemporaneamente validato seguendo una ormai consueta procedura che vede l’intero data set diviso in due parti, uno per l’addestramento vero e proprio e l’altro per la validazione dell’addestramento in corso.

In tal modo la rete neurale apprende da alcuni esempi e viene validata con altri dati differenti dai primi. Sono stati creati 300 esempi composti da due vettori, uno di output di numerosità 16 e uno di input di numerosità 23, dedicati all’addestramento per metodologia utilizzata. Il vettore di output è composto da 2 parametri di generazione, per i motivi casa-lavoro e casa-altri motivi, per la matrice interna, 4 parametri di distribuzione, per i motivi casa-lavoro e casa-altri motivi, per la matrice interna e 10 parametri relativi al numero di macchine in ingresso/uscita dall’area di studio nelle 5 sezioni al cordone. Il vettore di input è costituito dai flussi sugli archi di monitoraggio,

in modo da simulare l’acquisizione dati in tempo reale con, ad esempio, delle normali schede VMI di conteggio.

flussi input Fuorigrotta	
ID arco	flusso rilevato
38	2387.1
39	4472
40	5997
42	5541
49	4255.7
59	2594.9
70	1169.2
75	3321.8
80	2976.1
105	2157.9
106	2728.8
114	4585.3
122	3139.3
128	5234.9
134	2331.8
160	2496.5
167	3910.5
168	2429.5
176	3838
223	1081.6
229	2339.1
230	3069.9
231	2717

Vettore flussi di input

La simulazione viene fatta attraverso un vettore di flussi reale basato su osservazioni note a priori con le stesse caratteristiche informative dei dati con cui si è costruito il training set e, per semplificazione, si ipotizza che tale flusso di traffico sia comunque il risultato statistico di un buon numero di osservazioni, in modo da simulare l’effettivo aggiornamento della matrice OD attraverso dei rilevamenti reali. Si riportano di seguito le tabelle dei risultati delle elaborazioni fatte sulla rete di Fuorigrotta utilizzando la metodologia tradizionale e la metodologia proposta.

Coefficienti stimati con la metodologia tradizionale				
Coef.	GR	MLP	RBF	Obiettivo
B1	1.34361966	1.400856083	1.313838714	1.21
B2	2.118873553	2.421913254	1.530043711	1.83
B3	0.788384061	0.862861238	0.737694	0.68
B4	3.647164503	3.700019909	3.58090216	3.21
A1	0.800193091	0.814834037	0.656242268	0.65
A2	0.285700551	0.280182512	0.3133316	0.28
C1	1664.805046	1677.342809	1656.081315	1701
C2	3220.237456	3244.798564	3203.363117	3291
C3	5868.680246	5913.537615	5837.927823	5997
C4	1710.73829	1722.445351	1701.773865	1749
C5	3301.763935	3327.014947	3284.462389	3374
D1	4154.097347	4185.22331	4132.329496	4248
D2	2663.223897	2682.887534	2649.26836	2723
D3	3754.334274	3783.944455	3734.661218	3838
D4	2201.302799	2184.351179	2189.767771	2250
D5	992.8164845	988.0683714	987.6140351	1016
	GR	MLP	RBF	
RMSE%	0.082595509	0.111073439	0.073292944	
RMSE% I-I	0.1273916	0.17400511	0.110250881	
RMSE% Sc-At	0.022267812	0.018419999	0.027651935	

Risultati ottenuti con la metodologia tradizionale

Coefficienti stimati con la metodologia proposta				
Coef.	GR	MLP	RBF	Obiettivo
B1	1.254231964	1.365674173	1.272378154	1.21
B2	1.987080445	2.189590055	2.084177977	1.83
B3	0.651931632	0.582129933	1.013596684	0.68
B4	3.194882874	3.08269258	3.274277333	3.21
A1	0.676018305	0.709562314	0.692093478	0.65
A2	0.2732732	0.22387208	0.302237653	0.28
C1	1680.532037	1689.496965	1711.23407	1701
C2	3280.096781	3223.562544	3342.577872	3291
C3	5944.396368	5979.616533	6041.759686	5997
C4	1733.409448	1744.455569	1761.875252	1749
C5	3362.992898	3387.94068	3381.321605	3374
D1	4241.140039	4248.645259	4250.082261	4248
D2	2720.246465	2721.206705	2726.075623	2723
D3	3832.242257	3813.515151	3851.080363	3838
D4	2247.894198	2246.486671	2258.801472	2250
D5	1014.00192	1018.299136	1018.649049	1016
	GR	MLP	RBF	
RMSE%	0.028042802	0.096748806	0.085953846	
RMSE% I-I	0.043695952	0.152664976	0.135646056	
RMSE% Sc-At	0.005830645	0.007519331	0.006494916	

Risultati ottenuti con la metodologia proposta

I risultati ottenuti hanno mostrato come le prestazioni delle reti addestrate con differenti training set siano grossomodo soddisfacenti ed equivalenti, con una leggera predominanza sulla risposta da parte delle reti neurali addestrate con la metodologia proposta. Le risposte delle tre tipologie di rete neurale presentate sono sicuramente accettabili con una prevalenza della Generalized Regression Neural Network (GRN) che per sua stessa natura è più precisa delle altre tipologie di rete, ma richiede più tempo per l’addestramento (sull’ordine della decina di minuti in più).

Il comportamento di questa rete risulta significativo per due aspetti: è la rete che più risente della bontà dei dati presentati in fase di addestramento e quindi della funzione da “imparare”, e inoltre, se paragonato al risultato della rete RBF, si può notare come il numero di centri della funzione di trasferimento sia un parametro da valutare secondo la precisione che si vuole ottenere e i tempi di calcolo da ritenere accettabili.

Risulta, a nostro avviso, fondamentale analizzare la risposta delle reti neurali con le due differenti metodologie di creazione del training set all’input presentato. Si è quindi preferito procedere alla distinzione dei parametri di output in parametri interni e parametri di scambio e attraversamento e si è calcolato l’RMSE% per i due sottovettori.

Ne è risultato che i due sottovettori di risposta dell’ ANN addestrata con training set creato con metodologia tradizionale hanno un medesimo RMSE%, mentre al contrario i sottovettori di output delle ANN addestrate con un training set creato con la metodologia proposta hanno diverso RMSE%, in particolare i parametri relativi allo

scambio attraversamento hanno un indice molto basso, al contrario dei parametri di domanda interna che hanno un indice appena sopra la media.

Questo risultato permette di fare un’importante considerazione sul modo in cui l’ANN ha appreso l’associazione tra flussi e parametri di domanda a seconda del tipo di training set utilizzato. Tenendo presente la particolare domanda di mobilità presente nell’area di Fuorigrotta, che interessa maggiormente flussi di scambio e attraversamento, è naturale che la variabilità delle osservazioni dipenda maggiormente dalla domanda di scambio attraversamento piuttosto che dalla domanda interna; questa aspettativa viene confermata dalle reti addestrate con la metodologia proposta, e quindi a piccole variazioni di flussi osservati corrispondono piccole variazioni di parametri di domanda di scambio e attraversamento, mentre i parametri di domanda interni possono avere anche variazioni più consistenti. Viceversa questa caratteristica non viene rilevata sulle reti addestrate con la metodologia tradizionale.

In altre parole la bontà di creazione degli esempi con la metodologia proposta risiede oltre che nei risultati ampiamente accettabili, anche nella risposta della rete in dipendenza del tipo di domanda di mobilità di cui è caratterizzata l’area in oggetto. Questo risultato mette ancora di più in evidenza come l’informazione creata per la simulazione con le reti neurali artificiali deve essere attentamente studiata sia sotto il profilo dell’informazione del singolo parametro sia sotto il profilo di qualità degli

esempi utilizzati per l’addestramento, che devono essere “specchio della realtà” che si vuole riprodurre.

4.3 Confronto delle prestazioni di ANN su diversi test site.

La conferma dei concetti su esposti trova risposta nell’applicazione della metodologia proposta su diversi test site che hanno quindi diverse caratteristiche della domanda.

In particolare si è scelto di utilizzare una sola tipologia di ANN e testarla su diverse reti urbane e domande di mobilità, in modo da rendere più facile il confronto. Si è scelta la tipologia MLP che consente un maggior numero di regolazioni e un maggior controllo sul processo.

I test site sono il quartiere di Fuorigrotta con domanda reale e domanda modificata dall’operatore, e la cittadina di Acerra. Viene quindi utilizzato il quartiere di Fuorigrotta in due modi: con la sua domanda reale, già vista in precedenza, e con una modificata in modo da rendere la domanda di scambio e attraversamento paragonabile con la domanda interna. In aggiunta a tale applicazione, viene proposta un’ulteriore verifica su base dati reale riguardo la cittadina di Acerra, i cui dati sono più numerosi sia qualitativamente che numericamente.

Il paragone verrà fatto sia sulla risposta della ANN, quindi sui parametri simulati, sia sull’effettiva corrispondenza ai flussi ricercati, in modo da verificare la bontà della metodologia applicata sia dal punto di vista della sola risposta della ANN che dal punto di vista generale di possibili applicazioni ITS.

flussi input Fuorigrotta con domanda reale		flussi input Fuorigrotta con domanda modificata		flussi input Acerra	
ID arco	flusso rilevato	ID arco	flusso rilevato	ID arco	flusso rilevato
38	3734.1282	38	335.1818	2	442.5569
39	4176.7050	39	447.5260	5	331.3494
40	6604.3177	40	600.4016	10	207.1293
42	5617.1255	42	554.7104	14	605.7201
49	7502.1639	49	1265.5176	18	427.7848
59	5549.6689	59	1367.0072	24	1080.5939
70	2468.9936	70	79.7665	30	1370.7989
75	6094.3666	75	1173.5974	36	1075.1428
80	3318.7583	80	972.1331	42	455.1274
105	3996.2240	105	761.5559	44	421.3104
106	4073.5474	106	679.2010	45	360.7365
114	7101.4542	114	862.3503	47	520.7057
122	5121.3310	122	1460.0021	51	451.6707
128	7804.8467	128	427.7190	59	1164.3013
134	4173.1138	134	638.6806	62	1205.3115
160	4658.3994	160	366.0133	63	1164.3013
167	7062.2436	167	563.5397	64	443.863
168	4196.6328	168	242.5222	66	465.2848
176	6307.8992	176	384.1296	68	293.8039
223	1272.1726	223	362.7696	72	729.0969
229	4535.5215	229	1075.4625	75	781.9623
230	5232.6178	230	560.8376	78	29.9918
231	4695.7100	231	271.8939	80	772.735
				88	1454.9045
				92	1075.8552
				93	343.9853
				96	1596.1464
				111	488.4454
				114	768.5441
				116	772.8534
				135	128.6413
				143	18.3179
				145	48.3097
				186	131.7412
				190	18.3179
				227	4.3092
				269	463.2856
				286	484.0631
				292	484.0631
				295	18.3179

Vettori di input dei flussi nel quartiere di Fuorigrotta e nella città di Acerra

Quartiere di Fuorigrotta					
Stima dei parametri con domanda modificata			Stima dei parametri con domanda reale		
Coef.	MLP	Obiettivo	Coef.	MLP	Obiettivo
B1	1.331438	1.316617	B1	1.2434	1.162
B2	1.988438	2.026457	B2	2.97771	2.7408
B3	0.627941	0.644162	B3	0.70451	0.76202
B4	3.467678	3.456444	B4	3.61876	3.4761
A1	0.677091	0.682542	A1	0.71699	0.80088
A2	0.394829	0.389786	A2	0.30972	0.39852
C1	170.5800	170.3066	C1	3164.4121	3244
C2	330.1887	329.4879	C2	5581.6378	5618
C3	601.4864	600.4016	C3	6656.650	6604
C4	175.3318	175.0176	C4	3011.1472	3050
C5	338.7208	337.8759	C5	5082.8353	5114
D1	425.8908	425.0249	D1	6372.4262	6423
D2	273.1231	272.5084	D2	4138.6324	4177
D3	384.9306	384.1296	D3	6387.1127	6307
D4	225.6739	225.2226	D4	3560.9762	3511
D5	101.8154	101.5873	D5	1432.9142	1455
	MLP			MLP	
RMSE%	0.0308027		RMSE%	0.08809412	
RMSE% I-I	0.0170622		RMSE% I-I	0.13827880	
RMSE% Sc-At	0.0376777		RMSE% Sc-At	0.01297086	

Risultati ottenuti con la MLP utilizzando la rete di Fuorigrotta modificata

Stima dei parametri della cittadina di Acerra		
Coef.	MLP	Obiettivo
A1	0.370354835	0.36072291
A2	0.44809668	0.4368434
A3	0.334640769	0.32660243
A4	0.725747396	0.70644582
A5	0.750749387	0.73144582
A6	0.742745406	0.72344582
A7	0.320741283	0.28632534
B1	1.127819682	1.10192776
B2	0.952603492	0.93156631
B3	0.993274505	0.96905102
B4	0.922603492	0.90156631
B5	0.974342222	0.95168679
B6	1.024342222	1.00168679
B7	0.949990793	0.91277116
C1	0.763020371	0.70301213
C2	0.635743649	0.58253019
C3	0.909133508	0.83361456
C4	2.746873334	2.53084368
C5	2.441665186	2.24963882
C6	0.387681264	0.35156631
C7	0.790432381	0.78072291
D1	661.1680708	618.732981
D2	115.2739511	146.09817
D3	251.3451656	241.950767
D4	181.4595766	131.825902
D5	221.9329585	207.731249
D6	62.93148823	130.11449
D7	161.0655522	152.859527
D8	231.5048072	239.19418
E1	446.0445756	429.61914
E2	127.8153998	148.858321
E3	164.8828808	164.143002
E4	145.0139067	119.086736
E5	224.2493816	217.292213
E6	117.5793755	127.199846
E7	61.97129801	106.247326
E8	263.2039100	269.153742
	MLP	
RMSE%	0.146131457	
RMSE% I-I	0.058827546	
RMSE% Sc-At	0.211754372	

Risultati ottenuti con la MLP utilizzando la rete di Acerra

Si riportano quindi i primi 40 flussi derivati dalla domanda stimata, per la rete di Fuorigrotta con domanda reale e domanda modificata e per la rete di Acerra.

Fuorigrotta	STIMATI		ORIGINALI			
	AB FLOW	BA FLOW	AB FLOW	BA FLOW	ERR AB	ERR BA
1	126.6484	800.8101	106.0677	806.2786	0.1940336	0.0067824
3	173.5265	1028.1280	143.9024	1012.2680	0.2058624	0.0156678
5	1901.1243	1323.3669	1918.6427	1361.5462	0.0091306	0.0280411
8	338.1490	244.5319	333.9398	241.0059	0.0126047	0.0146303
11		2162.5463		2125.2362		0.0175557
12	658.7973	2546.2871	640.1573	2557.7211	0.0291178	0.0044704
13	465.9019	100.8620	452.0044	107.6567	0.0307464	0.0631145
15	817.4014	140.7851	801.7095	147.2610	0.0195731	0.0439757
17	90.3447	3.3357	89.6326	4.0098	0.0079447	0.1681131
18	224.1552	1399.8724	229.4283	1391.7368	0.0229837	0.0058456
19	201.5305	203.3324	192.2202	206.6963	0.0484356	0.0162746
20	336.6806	180.6225	326.5682	184.6148	0.0309657	0.021625
21	591.0488	889.7602	593.6863	840.7724	0.0044426	0.0582652
24	479.0774	513.6907	453.2681	528.6092	0.0569405	0.0282222
25	235.4958	178.1634	242.1421	185.8062	0.0274479	0.0411332
32	1017.3772	693.5277	911.7693	707.4039	0.1158274	0.0196157
33	522.0597	370.6596	541.6385	353.2205	0.0361474	0.0493717
38	3781.6115		3734.1282		0.012716	
39	3052.4059	4138.8781	3048.1241	4176.7050	0.0014047	0.0090566
40	6655.7453		6604.3177		0.007787	
42	5697.9139	3553.1648	5617.1255	3509.1046	0.0143825	0.012556
43	849.4376	997.2236	884.7224	1004.4174	0.0398823	0.0071622
44	1091.4842	3105.7417	1133.5093	3093.8630	0.0370752	0.0038394
46	2624.8109	2574.3192	2660.4422	2512.7751	0.013393	0.0244925
54	2559.7409	2039.8714	2525.8099	2089.7163	0.0134337	0.0238525
79	153.7725	118.9784	158.5209	115.5393	0.0299544	0.0297656
80	3301.3928	2410.8220	3318.7583	2479.1705	0.0052325	0.0275691
91	127.1635	323.7528	135.4589	331.5647	0.0612392	0.0235607
94	57.2259	278.8177	68.4091	237.1405	0.1634753	0.175749
95	348.5261	31.5862	327.9927	29.8863	0.0626032	0.0568789
97	1613.3895		1611.4999		0.0011726	
100	11629.4352		11901.0624		0.0228238	
101	7744.9173		7827.5150		0.0105522	
118	2783.5646		2753.3707		0.0109662	
127	1816.7948	2151.8687	1902.1433	2175.6022	0.0448696	0.0109089
129	193.4170	235.8978	205.1611	247.8606	0.0572433	0.0482642
135	15.0319	419.6816	19.2992	390.1051	0.2211128	0.0758167
159	2426.8165	1787.3661	2433.6677	1739.4461	0.0028152	0.027549
223	1300.8696	1239.2291	1272.1726	1257.9158	0.0225575	0.0148553
234	2886.8387	2594.7553	2821.0493	2504.6507	0.0233209	0.0359749

Confronto tra flussi stimati e flussi originali del quartiere di Fuorigrotta con domanda reale

Fuorigrotta	STIMATI		ORIGINALI		ERR AB	ERR BA
	AB FLOW	BA FLOW	AB FLOW	BA FLOW		
1	57.8762	431.2452	57.3757	414.6418	0.008647769	0.0385011
3	80.0572	447.0885	78.9082	449.5603	0.014352238	0.0055287
5	1063.9986	733.0774	1044.9987	741.0335	0.017857072	0.010853
8	206.1587	110.0344	201.9148	105.4348	0.020585597	0.0418015
11		906.9638		944.0987		0.0409442
12	412.1474	1125.9008	405.741	1174.1229	0.015543953	0.0428298
13	254.0878	99.1932	252.0357	78.9102	0.008076342	0.2044797
15	489.1183	56.4435	483.11	54.3382	0.01228394	0.0372992
17	51.2624	0.613	51.1864	1.2769	0.001482568	1.0830343
18	144.8725	594.6402	142.2409	611.0108	0.018164938	0.0275303
19	112.5256	92.6126	110.9624	87.4249	0.01389195	0.0560151
20	202.4633	70.6853	201.1822	69.2261	0.006327567	0.0206436
21	372.1455	347.8013	366.5846	352.8815	0.014942811	0.0146066
24	283.8417	324.3854	279.2741	304.6779	0.016092068	0.0607534
25	130.8359	71.7153	129.4659	70.0448	0.010471132	0.0232935
32	543.4221	406.036	535.1447	390.0129	0.01523199	0.0394623
33	286.7103	243.7071	282.7195	221.7173	0.013919277	0.0902304
38	333.8881		335.1818		0.003859646	
39	450.1279	2450.1567	447.526	2450.918	0.005813952	0.0003107
40	604.5411		600.4016		0.006894508	
42	555.8899	2004.2509	554.7104	2015.1826	0.00212638	0.0054543
43	107.9032	226.9612	120.0481	294.4021	0.112553659	0.2971473
44	781.2831	1627.3347	738.6164	1651.6307	0.054611062	0.0149299
46	1225.0711	1222.9678	1206.8404	1276.5662	0.01488134	0.0438265
54	1341.3704	1133.7158	1339.9731	1145.5639	0.001041696	0.0104507
79	81.4382	61.0217	79.8341	55.9379	0.019697145	0.0833113
80	974.9073	1182.9377	972.1331	1159.5564	0.002853733	0.0197655
91	50.1879	178.9357	49.9222	176.0011	0.005294105	0.0164003
94	42.2411	128.7345	36.8272	126.6082	0.128166643	0.0165169
95	162.1973	12.0031	155.7822	11.372	0.039551213	0.0525781
97	657.6899		681.1401		0.035655405	
100	6313.9027		6336.3797		0.003559922	
101	3813.9468		3832.9104		0.004972172	
118	1645.7498		1615.0016		0.018683399	
127	891.3399	1136.1353	893.6572	1092.0409	0.002599794	0.0388109
129	148.2531	106.5079	129.639	105.0702	0.125556228	0.0134985
135	6.3582	218.465	6.5287	214.5064	0.026815765	0.0181201
159	1411.897	904.5775	1438.4509	884.5178	0.01880725	0.0221758
223	363.1604	542.8642	362.7696	540.4674	0.001077373	0.0044151
234	1340.8141	1252.2174	1320.8887	1300.8335	0.014860673	0.038824

Confronto tra flussi stimati e flussi originali del quartiere di Fuorigrotta con domanda modificata

Acerra	FLUSSI STIMATI		FLUSSI ORIGINALI			
	AB FLOW	BA FLOW	AB FLOW	BA FLOW	ERR AB	ERR BA
1	235.7486	537.0037	236.7614	530.0248	0.004278	0.013167
2	229.0145	218.9484	231.9389	210.6180	0.012608	0.039552
3	334.2101	822.3551	338.6508	829.7551	0.013113	0.008918
4	247.5083		260.5250		0.049963	
5	175.8819	160.0812	175.3593	155.9901	0.00298	0.026227
6	127.0726	151.4685	127.8822	153.1184	0.006331	0.010775
7	173.2571	239.6772	172.4618	251.3931	0.004611	0.046604
8	512.7344	1223.0619	518.7069	1232.1928	0.011514	0.00741
9	591.0879	346.7269	593.9331	350.5815	0.00479	0.010995
10	207.7266		207.1293		0.002884	
11	571.1330	278.1487	572.5584	282.2822	0.00249	0.014643
12	435.0971	259.0945	442.0438	271.0821	0.015715	0.044221
13		78.8944		77.5392		
14	400.6360	188.5021	406.1028	199.6173	0.013462	0.017478
15	417.2110		408.9538		0.020191	0.055683
16		224.2791		226.7084		
17		151.0306		156.0471		
18	417.2372		427.7848		0.024656	0.010716
19		316.5663		308.8367		
20	387.8873	165.5451	391.7146	174.8639	0.009771	0.032147
21	474.1496		465.7942		0.017938	0.025028
22	458.2563	400.0435	447.2074	406.1971	0.024706	0.053292
23	314.4165		300.4512		0.046481	0.015149
24	339.9124	744.2735	337.2358	743.3582	0.007937	0.001231
25	633.6087	1211.6775	634.3473	1225.4140	0.001164	0.011121
26		258.0466		276.6064		0.067098
27		155.4522		156.6751		0.007805
28	321.9827		320.5771		0.004385	
29	356.7304		352.4357		0.012186	
30		1356.1143		1370.7989		0.010712
31		593.7875		609.1162		0.025165
32		410.3980		424.6596		0.033584
33		1781.4553		1790.0197		0.004785
34	341.5701		349.1009		0.021572	
35	410.8971		424.8536		0.03285	
36	1088.3837		1075.1428		0.012315	
37	377.5819		378.0629		0.001272	
38	495.4800		508.7018		0.025991	
39	114.6662	177.9758	120.4932	180.0064	0.04836	0.011281
40	123.1313	120.2061	129.4348	119.5432	0.0487	0.005545

Confronto tra flussi stimati e flussi originali della città di Acerra

Errore medio su arco	ERR
Acerra	0.044201
Fuorigrotta con domanda modificata	0.043899
Fuorigrotta con domanda reale	0.040312

Errore medio su tutti i flussi stimati

L’errore medio viene semplicemente calcolato come media dei discostamenti tra i valori stimati e i valori reali, riportati nelle tabelle qui sopra (ERR_AB e ERR_BA). Ovvero, semplicemente:

$$errore\ medio = \frac{1}{n} \sum_{i=1}^n \frac{|f_i^{st} - f_i|}{f_i}$$

Tali risultati sono sicuramente soddisfacenti, hanno messo in luce come l’informazione costruita a priori risulti essere determinante per la risposta dell’ANN e hanno confermato le ipotesi del lavoro alla base della presente tesi, evidenziando, nel caso di Acerra, come aumentando la numerosità la risposta sia in termini di parametri che di flussi sia assolutamente paragonabile con le precedenti.

5 Conclusioni

Il presente lavoro di tesi ha apportato un contributo scientifico sulla possibilità di utilizzo di una rete neurale artificiale per il controllo dei flussi di traffico su una rete urbana reale di grandi dimensioni.

Si è evidenziato come la scelta del tipo di dati da utilizzare per le applicazioni e la creazione dei dati sia di fondamentale importanza per una corretta implementazione della rete neurale ai problemi posti.

I dati creati con la metodologia proposta non hanno alcun legame tra loro se non il fatto che discendono da una statistica che fa riferimento ai probabili valori medi di accadimento, e quindi, evitando quindi la forte collinearità insita nella generazione di dati fatta con la metodologia tradizionale, si è al riparo da possibili risposte non corrette da parte della rete neurale. Inoltre la creazione di dati a partire dai modelli di generazione e distribuzione ha permesso di evitare alcune incongruenze, oggetto di critiche, con la metodologia tradizionale che modifica i singoli valori della matrice OD senza alcun controllo. La scelta fatta sul tipo di dati a cui far riferimento, in particolare il vettore dei parametri di domanda, ha permesso l’applicabilità su reti di grandi dimensioni con risultati più che accettabili eludendo le problematiche insite nei limiti tecnologici delle apparecchiature utilizzate e della difficile reperibilità dei dati richiesti in tali applicazioni. I confronti effettuati, in termini di differenze di risposta tra diverse reti neurali e in termini di capacità previsionali tra diverse reti urbane hanno sempre ampiamente confermato la validità dell’approccio seguito. Ulteriori sviluppi sono da attendersi con un’analisi della stabilità della risposta, che necessariamente richiede una base dati decisamente più ampia e differenziata, e con l’integrazione della logica fuzzy nella creazione dei dati in alternativa alla più classica ma più consolidata statistica.

Relativamente alle reti neurali utilizzate, l’associazione tra flussi e coefficienti è risultata nettamente più facile da implementare, e di conseguenza la risposta è stata notevolmente più veloce, benché si è evidenziato un limite temporale di apprendimento nell’applicazione della metodologia proposta relativamente alla città di Acerra dovuto, è bene ricordarlo, anche al supporto hardware utilizzato. Le esperienze riportate in letteratura fanno rilevare infatti che l’associazione di input di un ordine di grandezza inferiore all’output, come sarebbe stato nel nostro caso se si fosse associato al vettore flussi, di qualche decina di elementi, il vettore matrice che facilmente supera il migliaio di unità, non è di facile soluzione e necessita di un numero di dati dedicati all’apprendimento dell’ordine delle decine di migliaia di unità. Il lavoro necessario per approntare una rete che possa rispondere a queste esigenze necessita di molto tempo e risorse, oltre al fatto che la rete potrebbe rispondere non velocemente e non col giusto grado di approssimazione che le si richiede; in questi termini si parla solitamente di un problema mal posto. L’associazione proposta, quindi ci ha permesso lo studio di reti urbane di grandi dimensioni con le reti neurali ed è interessante notare come, tenendo in conto la fase di training che, a seconda della rete urbana considerata e della tipologia di rete neurale adottata, varia da qualche minuto a poche decine di minuti, la risposta di una rete neurale addestrata a tempi stimabili in qualche secondo; prestazioni assolutamente non paragonabili con l’attuale procedura di calcolo che impiega alcune ore per dare la stessa risposta. Si ritiene che la risposta possa ulteriormente migliorata

prendendo in considerazione i parametri posti come invarianti in questo elaborato, quali ad esempio gli attributi stessi delle alternative dei modelli. Le architetture neurali utilizzate hanno dimostrato, come affermato in letteratura, la loro versatilità, ma le continue proposte di miglioramenti computazionali impongono una costante attenzione alle novità, che sempre più si rivolgono ai particolari problemi in cui sono adoperate.

Concludendo, e sempre ricordando le parole di Norberto Bobbio riportate all’inizio del presente lavoro, le applicazioni di reti neurali al settore dei trasporti sono sicuramente interessanti, all’avanguardia e, come ogni innovazione tecnologica, suscitano curiosità e diffidenza. Ciò che si rende necessario, soprattutto in questo periodo storico, e che si vuole sottolineare è l’attenzione che si deve avere sugli sviluppi del settore tecnologico, monitorandolo e testandolo continuamente, rilevandone gli effetti, senza abbandonare o preferire “sine quaestione et contradictione” la teoria che finora è stata faticosamente costruita.

6 Bibliografia

1. Abdulhai, B., Porwal, H., Recker, W., (1999). **“Short-term freeway traffic flow prediction using genetically-optimized time-delay-based neural networks.”** Transportation Research Board, 1999.
2. Adeli, H., Jiang, X., (2003) **“Neuro-fuzzy logic model for freeway work zone capacity estimation”** Journal of Transportation Engineering, ASCE 129 (5), 484–493.

3. Adeli, H., Karim, A., (2001) “**Fuzzy-wavelet RBFNN model for freeway incident detection**” Journal of Transportation Engineering, ASCE 126 (6), 464–471.
4. Ben Akiva, M., Morikawa, T., (1988) “**Data combination and updating methods for travel surveys**” Transportation Research Record vol.1203 pp.40–47.
5. Berthold, M.R., Diamond, J., (1998) “**Constructive training of probabilistic neural networks**” Neurocomputing vol.19 pp.167-183.
6. Bierlaire, M., (2002) “**The total demand scale: a new measure of quality for static and dynamic origin destination trip tables**” Transportation Research-B vol.36 pp.837-850.
7. Camus, R., Cantarella, G.E., Inaudi, D., (1997) “**Real-time estimation and prediction of origin destination matrices per time slices**” International Journal of Forecasting vol.13 pp.13-19.
8. Chang, G.L., Su, C.C., (1995) “**Predicting intersection queue with neural network models**” Transportation Research-C vol.3 n.3 pp.175-191.
9. Chen, H., Grant-Muller, S., Mussone, L., Montgomery, F., (2001). “**A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting**” Neural Computing and Applications 10, 277–286.
10. Chen H., Grant-Muller, S., (2001) “**Use of sequential learning for short-term traffic flows forecasting**” Transportation Research-C vol.9 pp.319-336.
11. Codina, E., Barcelò, J., (2004) “**Adjustment of O-D trip matrices form observe volumes: an algorithmic approach based on conjugate directions**” European Journal of Operational Research vol.155 pp.535-557.
12. Costa, A., Markellos, R.N., (1997) “**Evaluating public transport efficiency**

- with neural network models”** Transportation Research-C vol.5 n.5 pp. 301-312
13. Dougherty, M.S., (1995) **“A review of neural networks applied to transport”** Transportation Research vol.3 n.4 pp. 247-260.
 14. Dougherty, M.S., Cobbett, M.R., (1997) **“Short-term inter-urban traffic forecasts using neural networks”** International Journal of Forecasting vol.13 pp.21-31.
 15. Haykin, Simon (1999) **“Neural Networks: a comprehensive foundation”** Prentice Hall, Upper Saddle River, New Jersey
 16. Harikishan Reddy, K., Chakroborty, P., (1998) **“A fuzzy inference based assignment algorithm to estimate O-D matrix from link volume counts”** Comput., Environ. And Urban System vol.22 n.5 pp.409-423.
 17. Hazelton, M.H., (2003) **“Some comments on origin-destination matrix estimation”** Transportation Research-A vol.27 pp.811-822.
 18. Hazelton, M.L., (2001) **“Inference for origin-destination matrices: estimation, prediction and reconstruction”** Transportation Research-B vol.35 pp.667-676.
 19. Hefny H.A., Bahnasawi A.A., Abdel Wahab A.H., Shaheen S.I. (1999) **“Logical Radial basis Function Network: a hybrid intelligent model for function approximation”** Advances in Engineering Softwares 30 pp 407- 417.
 20. Henn V. (2005) **“What is the meaning of fuzzy cost in fuzzy traffic assignment models?”** Transportation Research part C 13 pp.107-119
 21. Henry, J.J., Farges, J.L., Gallego J.L., (1998) **“Neuro fuzzy techniques for traffic control”** Control Engineering Practise vol.6 pp.755-761.
 22. Hensher, D.A., Ton, T.T., (2000) **“A comparative of the potential of artificial**

neural networks and nested logit models for commuter mode choice”

Transportation Research-E vol. 36 pp. 155-172.

23. Ishak, S., Kotha, P., Alecsandru, C., (2003) **“Optimization of dynamic neural networks performance for short-term traffic forecasting. In”** Transportation Research Board 82nd Annual Meeting, Washington, DC.
24. Jayakrishnan, R., Tsai, W.K., Chen, A., (1995) **“A dynamic traffic assignment model with traffic flows relationships”** Transportation Research-C vol.3 n.1 pp. 51-72.
25. Khan, S., Ritchie, S., (1998) **“statistical and neural classifiers to detect traffic operational problems on urban arterials”** Transportation Research-C vol.6 pp.291-314.
26. Kirby, H.R., Watson, S.M., Dougherty, M.S., (1997) **“Should we use neural networks or statistical models for short-term motorway traffic forecasting?”** International Journal of Forecasting vol. 13 pp.43-50
27. Lan, C.J., Davis, C.A., (1999) **“Real-time estimation of turning movement proportions from partial counts on urban networks”** Transportation Research-C vol.7 pp.305-327.
28. Ledoux, C., (1997) **“An urban traffic flows model integrating neural networks”** Transportation Research-C vol.5 n.5 pp.297-300.
29. Levendovszky, J., Vegso, C.S., Van der Meulen, E.C., (2000) **“Non parametric decision algorithms for CAC in ATM networks”** Performance Evaluation vol.41 pp.133-147.
30. Li, B., De Moor, B., (1999) **“Recursive estimation based on equality-constrained optimization for origin-destination matrices”** Transportation Research-B vol.33 pp.203-214.

31. Li, M.Z.F., (1999) **“Estimating congestion toll by using traffic count data – Singapore’s area licensing scheme”** Transportation Research-E vol.35 pp.1-10.
32. Messai, N., Thomas, P., Lefebvre, D., El Moudni, A., (2002) **“Optimal neural network architectures for the flow-density relationship of traffic models”** Mathematics and Computers in Simulation vol.60 pp.401-409.
33. Narayanan, A., (1995) **“A linear programming approach for synthesizing origin destination trip tables based on a partial set of link volumes”** thesis performed at the Virginia Polytechnic Institute-State University, Blacksbourg, Virginia.
34. Nijkamp, P., Reggiani, A., Tritapepe, T., (1996) **“Modelling inter-urban transport flows in Italy: a comparison between neural network analysis and logit analysis”** Transportation Research-C vol.4 n.6 pp.323-338.
35. Nijkamp, P., Reggiani, A., Tsang, W.F., (2004) **“Comparative modelling of interregional transport flows: application to multimodal European freight transport”** European Journal of Operational Research vol.155 pp.584-602.
36. Palacharla, P.V., Nelson, P.C., (1999) **“Application of fuzzy logic and neural networks for dynamic travel time estimation”** International Transaction in Operational Research vol.6 pp.146-160.
37. Potvin, J.Y., Robillard, C., (1995) **“Clustering for vehicle routing with a competitive neural network”** Neurocomputing vol.8 pp.125-139.
38. Qiao, F., Yang, H., Yam, W.H.K., (2001) **“Intelligent simulation and prediction of traffic flow dispersion”** Transportation Research-B vol.35 pp.843-863.
39. Ridwan M. (2004) **“Fuzzy preference based traffic assignment problem”**

Transportation Research part C

40. Rodrigue, J.P., (1997) “**Parallel modelling and neural networks: an overview for transportation/land use system**” Transportation Research-C vol.5 n.5 pp.259-271.
41. Sadek, A.W., Charles, M., (2003) “**Modular artificial neural networks for solving the inverse transportation planning problem**” Transportation Research Board 82nd Annual Meeting.
42. Schittenkopf, C., Deco, G., Brauer, W., (1997) “**Two strategies to avoid overfitting in feedforward networks**” Neural Networks vol.10 n.3 pp.505-516.
43. Sherali, D., Silvanandan, R., Hobeika, A.G., Narayanan, A., (1994) “**Estimate missing link volumes in a traffic network: a linear programming approach**” paper presented at the Transport Research Annual Meeting, Washington D.C.
44. Sherif, I., Prashanth, K., Ciprian, A., (2003) “**Optimization of dynamic neural networks performance for short-term traffic prediction**” Transportation Research Board 82nd Annual Meeting.
45. Smith, B.L., Williams, B.M., Oswald, R.K., (2002) “**Comparison of parametric and non-parametric models for traffic flows forecasting**” Transportation Research-C vol.10 pp.303-321.
46. Spall, J., Chin, D., (1997) “**Traffic-responsive signal timing for system wide traffic control**” Transportation Research-C vol.5 n.3 pp.153-163.
47. Van der Vort, M., Dougherty, M.S., Watson, S.M., (1996) “**Combining Kohonen maps with ARIMA time series models for traffic flows prediction**” Transportation Research-C vol.4 n.5 pp.307-318.
48. Van der Zijpp, N.J.,(1995) “**Dynamic origin destination matrix estimation on**

motorway networks” thesis performed at the Transportation Planning and Traffic Engineering department of the Faculty of Civil Engineering of Delft University of Technology, Delft, Netherland.

49. Vithoulkas P., Koutsopoulos H. (2003) “**Modeling discrete choice behaviour using concepts from Fuzzy theory, approximate reasoning and neural network**” Transportation Research part C 11 pp.51-73
50. Vlahogianni, Karlaftis, Golias(2005) “**Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach**” Transportation research part C 13 pp.211-234
51. Vukadinovic, K., Teodorovic, D., Pavkovic, G., (1999), “**An application of neuro fuzzy modelling: the vehicle assignment problem**”, European Journal of Operational Research vol.114 pp.474-488.
52. Wong, S.C., Tong, C.O., (1998) “**Estimation of time-dependent origin-destination matrices for transit network**” Transportation Research-B vol.32 n.1 pp.35-48.
53. Xiong, D., (2001) “**Automated road network extraction from high resolution images**” Technical Report U.S. Department of Transportation vol.3 pp.1-14.
54. Yang, H., (1995) “**Heuristic Algorithms for the bi-level origin destination matrix estimation problem**” Transportation Research-B vol.29B n.4 pp.231-242.
55. Yang, H., Akiyama, T., Sasaki, T., (1998) “**Estimation of time-varying origin-destination flows from traffic counts: a neural network approach**” Mathl. Comput. Modelling vol.27 n.9 pp.323-334.
56. Yin, H., Wong, S.C., Xu, J., Wong, C.K., (2002) “**Urban traffic flow prediction using a fuzzy-neural approach**” Transportation Research-C

vol.10 pp.85-98.

57. Yun, S.Y., Namkoong, S., Rho, S.H., Shin, S.W., Choi, J.U., (1998) “**A performance evaluation of neural network models in traffic volume forecasting**” Mathl. Comput. Modelling vol.27 n.9 pp.310-323.
58. Zhang, X., Maher, M.J., (1998) “**The evaluation and application of a fully disaggregate method for trip matrix estimation with platoon dispersion**” Transportation Research-B vol.32 n.4 pp.261-276.
59. Zhong M., Lingras P., Sharma S., (2004) “**Estimation of traffic missing counts using factor, genetic, neural and regression techniques**” Transportation Research Part C pp.139-166.