

Reti neurali in meteorologia: un'applicazione su dati della regione Lombardia

Michele de Rosa*, Alessandro de Rosa*, Raffaello Di Martino*, Paolo Marrone*

Lavoro pervenuto in redazione il 02/11/2004

Sommario - Le reti neurali, nel seguito indicate anche con N.N. (Neural Networks), vengono utilizzate oramai in tantissimi campi che vanno dall'informatica, alla medicina, alla fisica, ecc.. La meteorologia, quale branca della fisica, si presta, dunque, all'utilizzo di tali modelli, la cui teoria ha fatto passi da gigante da quando, nel 1962, Rosenblatt enunciò la teoria sul perceptrone (5) e ne illustrò le proprietà. Nel presente lavoro vengono proposti i risultati della sperimentazione delle reti neurali bayesiane con shortcut (4), per prevedere l'andamento a 4 ore di alcune variabili meteorologiche relative ad alcune zone della regione Lombardia.

Summary - The neural networks are used in many scientific fields, like computer science, medicine, physics etc.. These models, which theory was expressed by Rosenblatt in 1962 (5) with his perceptron model, can be applied to the meteorology, which is a branch of the physics. In this work we will show some results obtained by applying the Bayesian neural networks with shortcuts (4) to forecast the values of some meteorological parameters in the next 4 hours. The time series have been downloaded from the "regione Lombardia" web site.

Bayesian Neural Networks

Le reti neurali bayesiane, Bayesian N.N., sono apparse in letteratura da una decina d'anni circa a seguito del lavoro svolto da David J. C. MacKay (1). Le N.N. vengono utilizzate principalmente per problemi di classificazione e di fitting di dati, in special modo nei casi in cui vi è una relazione fortemente non lineare tra input ed output. In generale per questi tipi di problemi vengono utilizzati approcci di tipo convenzionale, quali modelli autoregressivi, statistici ecc.. Le reti neurali offrono la possibilità di "apprendere" la legge che genera i dati in modo tale da dare luogo ad output "ammissibili", rispondenti cioè alla realtà con un certo grado di confidenza. Il problema principale dei metodi di approssimazione, dal quale sono affette anche le N.N. convenzionali, quali ad esempio MLP (Multi Layer Perceptron) (2) ed RBF (Radial Basis Function) (2), è il problema della complessità del modello utilizzato. La questione si pone quando si sceglie il modello che dovrebbe meglio rappresentare i dati tra tutti i modelli ammissibili. Ebbene ciò che si dovrebbe preferire è il modello più semplice e scartare quelli troppo complicati (tale principio è noto con il nome di Occam razor (rasoio di Occam). Questo può essere fatto aggiungendo semplicemente un termine di penalità direttamente proporzionale alla complessità di ciascun modello. Le Bayesian

N.N. hanno insiti i valori di penalità, i cosiddetti iperparametri, che automaticamente durante la fase di apprendimento, penalizzano i modelli troppo complessi. Questo avviene poiché durante la fase di apprendimento, detta di training, vengono assegnati alla rete i pesi più "probabili" per i dati di input. In questa fase una parte del dataset viene scelta per addestrare la rete: vengono forniti ad essa i dati di input ed i corrispondenti dati di output. Da notare che l'output di una rete può essere espresso nella seguente forma:

$$y_k(\underline{w}_k; \underline{h}) = g \left(\sum_{j=1}^M w_{k,j} g' \left(\sum_{i=1}^N h_{i,j} x_i \right) \right)$$

dove $k=1...T$ è il singolo output, T è il numero di output, M è il numero di hidden, N è il numero di input e g e g' sono le funzioni di attivazione dei neuroni output ed hidden rispettivamente. Le funzioni di attivazione più comunemente utilizzate sono la funzione logistica e la funzione lineare:

$$g(x) = \frac{1}{1 + e^{-x}} \quad \text{funzione logistica}$$

$$g(x) = cx \quad \text{funzione lineare}$$

dove $c \in \mathfrak{R}$. Nella fase di apprendimento i pesi w

* KWOS, Centro per lo studio della strumentazione meteorologica

della N.N. vengono aggiornati in modo tale da minimizzare una funzione di errore quale, ad esempio, l'errore quadratico medio tra l'output fornito dalla rete neurale ed i dati reali:

$$E(\underline{w}, t) = \frac{\sum_{z=1}^C [y_z(\underline{w}) - t_z]^2}{C}$$

dove C è il numero di dati in ingresso e t_z è il dato target z-esimo.

Reti neurali con shortcut

Le reti neurali con shortcut (4) rappresentano una ulteriore evoluzione, essendo costituite da una rete neurale di tipo MLP, come descritta in precedenza, e da connessioni che vanno dai neuroni di input ai neuroni di output, detti shortcut. La presenza di queste connessioni lineari fa sì che la rete riesca a modellare anche eventuali parti lineari presenti nel sistema. In conseguenza di ciò l'equazione che descrive l'output di una rete neurale è la seguente:

$$y_k(\underline{w}_k; \underline{h}; \underline{w}_k^{Sc}) = g \left(\sum_{j=1}^M w_{k,j} g' \left(\sum_{i=1}^N h_{i,j} x_i \right) \right) + \sum_{i=1}^N w_{k,i}^{Sc} x_i$$

dove $k=1...T$ è il singolo output, T è il numero di output, M è il numero di hidden, N è il numero di input, i valori $w_{i,j}$ sono i pesi delle connessioni shortcut e g e g' sono le funzioni di attivazione dei neuroni di output e dei neuroni hidden rispettivamente. L'equazione relativa al calcolo della funzione di errore $E(\underline{w}, t)$ varierà di conseguenza.

L'utilizzo di reti neurali con shortcut porta una serie di vantaggi che possono essere riassunti principalmente nella capacità di modellare un sistema lineare (per il quale i pesi $w_{k,j}$ risulteranno tendere a zero con alta probabilità, escludendo di fatto la componente non lineare della rete) ed un sistema non lineare (per il quale i pesi $w_{i,j}^{Sc}$ tenderanno a zero con probabilità molto alta, annullando la componente lineare della rete). La fase di apprendimento, come descritto in (4), consiste nella identificazione di un Generalized Linear Model (GLM), che rappresenterà la parte lineare del sistema e che darà in input alla rete neurale MLP alcuni parametri sulla distribuzione a priori dei pesi e sui valori iniziali degli iperparametri, dai quali poi partirà il processo di apprendimento. Tutto questo fa in modo che, prima di iniziare il processo di apprendimento della rete MLP, si abbiano già informazioni sul tipo di sistema che si sta analizzando.

Il metodo di Cao

Il metodo di Cao(3) è un algoritmo che permette di determinare la dimensione minima di "embedding" di una serie temporale. Tale metodo ha i seguenti vantaggi:

1. non necessita di parametri soggettivi eccetto che del ritardo temporale
2. non dipende in modo eccessivo dal numero di dati disponibili
3. è efficiente dal punto di vista computazionale

I metodi di scelta della dimensione minima di embedding, quali singular value decomposition o il false neighbours method, sono afflitti da svantaggi quali inefficienza computazionale o la necessità di utilizzare criteri soggettivi nel loro utilizzo. Tali svantaggi vengono superati dal metodo di Cao.

Sia x_1, x_2, \dots, x_n una serie temporale, allora i vettori ritardati possono essere costruiti come segue:

$$y_i(d) = (x_i, x_{i+\tau}, \dots, x_{i+(d-1)\tau})$$

dove $i=1, 2, \dots, N-(d-1)\tau$ è la dimensione di embedding e τ è il ritardo temporale.

In maniera del tutto simile al false neighbours method, viene definita la quantità:

$$a(i, d) = \frac{\|y_i(d+1) - y_{n(i,d)}(d+1)\|}{\|y_i(d) - y_{n(i,d)}(d)\|}$$

per $i=1, 2, \dots, N-d\tau$. Definiamo con $E(d)$ il valore medio rispetto ad i di tutti gli $a(i, d)$, che dipende solo dalla dimensione d e dal ritardo τ . Da ciò si definisce la quantità:

$$EI(d) = E(d+1)/E(d)$$

Il metodo di Cao consiste nel trovare il valore d_0+1 tale che $\forall d > d_0$. La figura seguente mostra un esempio di calcolo della quantità $EI(d)$ applicata ai dati prelevati dalla stazione di via Garibaldi a Bergamo. Si può notare come il valore $EI(d)$ si stabilizzi per valori di d pari a circa 20. Come si vedrà in seguito, il numero di input scelti per prevedere la temperatura rilevata dalla suddetta stazione sarà appunto 20.

L'architettura

L'architettura, cioè il modello, da scegliere è un problema di difficile soluzione nel caso delle N.N., poiché implica la scelta del numero di neuroni di input, del numero di strati e del numero di neuroni hidden. Nel caso in cui si vogliono utilizzare le N.N. per effettuare previsioni su serie temporali, è necessario scegliere l'ammontare dei dati appartenenti al passato che la rete dovrà ricevere in input per fornire una previsione sui dati futuri. Per fissare questo orizzonte temporale è stato scelto il metodo di Cao (3), pre-

cedentemente descritto, che permette di stabilire il numero di input necessari alla N.N. Nel caso delle serie temporali, detta $\{x_i\}_{i=1...L}$ una serie temporale ed N il numero di input della N.N., il dato target è fissato come $t_i=x_i$ per $i=1...L$, mentre i dati in input alla N.N. sono i valori della serie temporale $(x_{i-N+1}, \dots, x_{i-1})$ per $i=1...L$. Ogni rete, definita in seguito, sarà caratterizzata da 2 strati di neuroni (hidden e output), dalla funzione di attivazione tangente iperbolica per i neuroni hidden e da quella lineare per i neuroni di output. Da ciò risulta chiaro che per fissare l'architettura è necessario avere a disposizione la serie temporale. Nel nostro caso, le serie temporali sono state prelevate dal sito della regione Lombardia (www.ambiente.regione.lombardia.it) e

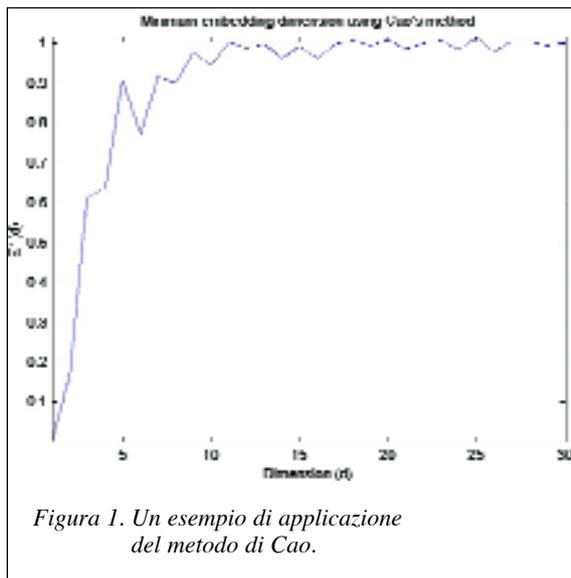


Figura 1. Un esempio di applicazione del metodo di Cao.

consistono di medie orarie dei seguenti parametri:

1. Temperatura
2. Pressione
3. Umidità
4. Radiazione solare
5. Velocità del vento
6. Direzione del vento
7. Precipitazioni piovose
8. Concentrazione di ozono
9. Concentrazione di polveri sospese nell'aria
10. Concentrazione di monossido di carbonio

Le serie temporali sono relative agli anni che vanno dal 1999 al 2002. I dati sono campionati ogni 4 ore e sono stati utilizzati dei metodi di interpolazione per ovviare ai dati mancanti. Per gli esperimenti è stato scelto di analizzare i parametri di temperatura, umidità relativa, pressione atmosferica, radiazione solare totale e velocità del vento re-

lativamente alle stazioni di:

- Bergamo
- Cassano D'Adda
- Morbegno
- Agrate Brianza

Gli esperimenti si riferiranno all'analisi di alcune delle grandezze scelte per ciascuna stazione.

I risultati

Di seguito verranno riportati i risultati relativi agli esperimenti effettuati sui parametri meteo scelti rilevati ogni 4 ore, riguardanti le centraline indicate in precedenza. Negli esperimenti il 70% del data set è stato utilizzato per il training della rete ed il restante 30% per la validazione della stessa. Si suppone che il sistema, che si tenterà di modellare, sia del tipo:

$$y(n) = \sum_{i=1}^N c_i y(n-i) + \sum_{j=1}^M d_j h(y(n-j)) + \varepsilon(n)$$

dove $N < n, M < n$, h è una funzione non lineare ed $\varepsilon(n)$ è rumore gaussiano. Nel seguito definiremo:

$$R(n) = y(n) - y^{NN}(n)$$

il residuo n-esimo, dove $y^{NN}(n)$ è la previsione del pattern n-esimo data in output dalla rete.

$R(n)$ rappresenta, quindi, il residuo n-esimo. Nel caso in cui la rete abbia modellato correttamente il sistema otterremo la relazione $\bar{R} \approx \bar{\varepsilon}$, ovvero il vettore dei residui \bar{R} tenderà ad essere gaussiano. I risultati ottenuti per ogni grandezza meteo verranno confrontati con l'output di un semplice previsore, quale un estrapolatore lineare.

La stazione di Bergamo

In questa sezione vengono presentati i risultati relativi alla previsione di temperatura e velocità del vento relativamente alla stazione sita in via Garibaldi a Bergamo. Per prevedere la temperatura, seguendo il metodo di Cao, è stata scelta un'architettura di rete con 20 neuroni di input, 3 neuroni hidden e 1 neurone di output. Di seguito vengono riportate le figure riguardanti la previsione di temperatura. Come si può notare, in Figura 2 viene mostrato un particolare del grafico di previsione della temperatura. In particolare viene confrontato il valore reale della temperatura (in rosso) con quello previsto dalla rete neurale (in blu) e da un previsore elementare quale l'estrapolatore lineare (in magenta). Sul totale delle previsioni, la rete neurale ha dato luogo ad un errore medio di previsione pari a -0.00034945 °C con una deviazione standard pari a 0.025711 °C, mentre l'errore medio di previsione

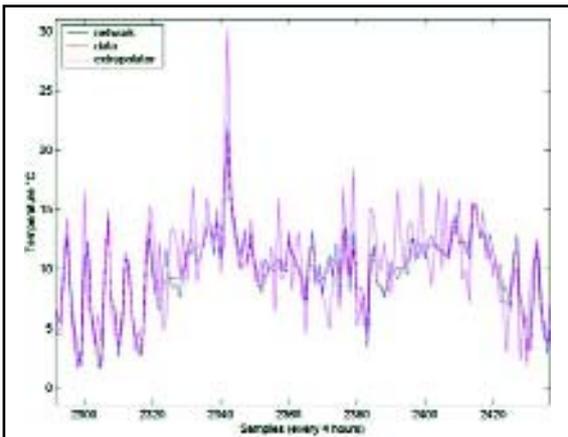


Figura 2. Un particolare della previsione di temperatura. Si può notare come l'estrapolatore abbia sovrastimato decisamente il picco a 20 °C circa

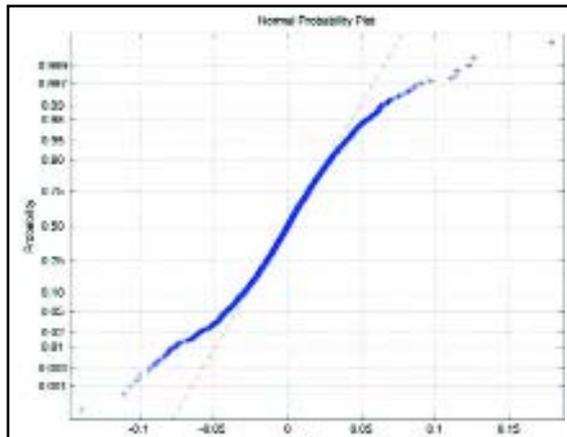


Figura 3. Analisi dei residui relativi all'errore di previsione della rete neurale.

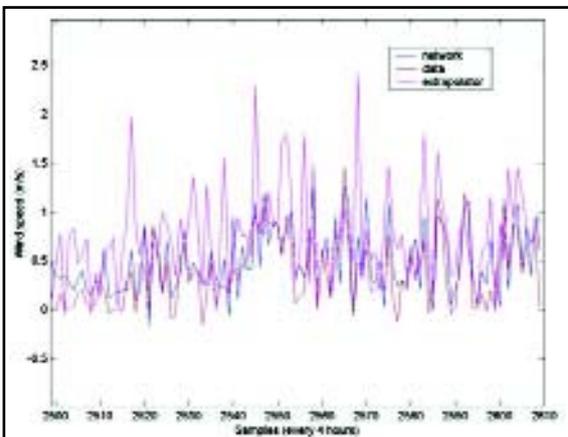


Figura 4. Un particolare della previsione della velocità del vento. L'estrapolatore sovrastima abbondantemente le velocità sopra 1 m/s.

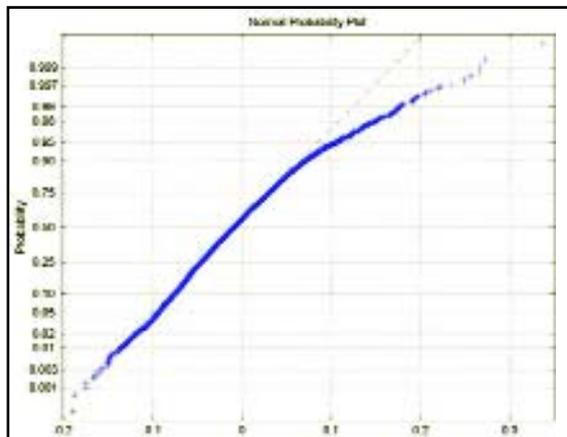


Figura 5. Analisi dei residui relativi all'errore di previsione della rete neurale.

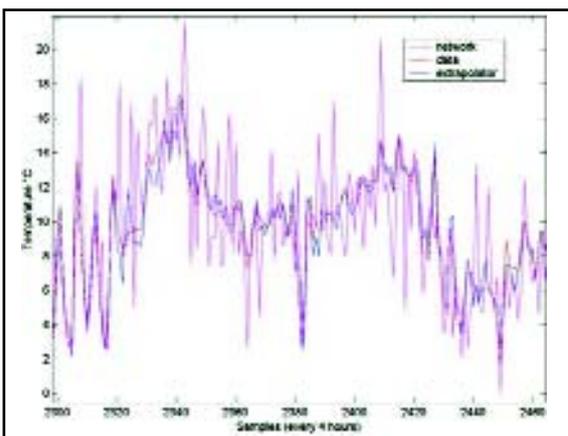


Figura 6. Previsione di temperatura. Si può notare la discrepanza tra la previsione data dall'estrapolatore e quella fornita dalla rete neurale

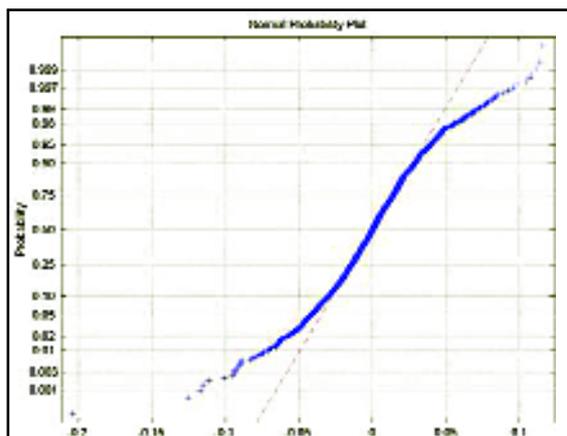


Figura 7. Analisi dei residui relativi all'errore di previsione della rete neurale.

dell'extrapolatore lineare è stato pari a 2.1725 °C con deviazione standard pari a 1.7345 °C. La Figura 3 mostra come i residui tendano ad essere gaussiani, a riprova del corretto comportamento della rete. Le figure seguenti riportano i risultati riguardanti la previsione della velocità del vento. Nel caso della velocità del vento, l'architettura, dettata dall'applicazione del metodo di Cao, è stata fissata a 20 neuroni di input, 3 hidden e 1 output.

In Figura 4 viene mostrato un particolare relativo al grafico di previsione della velocità del vento, in cui vengono confrontati il dato reale (rosso), la previsione della rete neurale (blu) e la previsione dell'extrapolatore (magenta). L'errore medio commesso dalla rete neurale sul totale dei dati è risultato pari a -0.0045 m/s con deviazione standard pari a 0.061785 m/s, mentre l'errore medio commesso dall'extrapolatore è risultato pari a 0.47044 m/s con deviazione standard di 0.86672 m/s.

La stazione di Cassano D'Adda

I risultati, che verranno presentati relativamente alla stazione di Cassano D'Adda, si riferiranno alla previsione a 4 ore della temperatura e dell'umidità relativa. Nelle figure che seguono vengono presentati i risultati relativi alla temperatura. L'errore medio di previsione commesso dalla rete neurale è stato pari a -0.00034006 °C con deviazione standard pari a 0.025888 °C, mentre l'extrapolatore ha dato la sua previsione commettendo un errore medio di previsione pari a 2.2987 °C con deviazione standard pari a 3.0356 °C. Per ottenere tali risultati, è stata utilizzata una rete con 20 neuroni di input, 3 neuroni hidden e 1 neurone di output. I residui in Figura 7 tendono ancora ad essere gaussiani. Per quanto riguarda la previsione dell'umidità relativa, è stata utilizzata una rete neurale con 23 neuroni di input, 3 neuroni hidden e 1 neurone di output. L'umidità è una variabile con una componente altamente non lineare. L'errore medio di previsione dato dalla rete neurale è stato pari a 0.0018937% con deviazione standard 0.050843%, mentre nel caso dell'extrapolatore l'errore medio di previsione è stato pari a 8.383% con deviazione standard pari a 8.3978%.

La stazione di Morbegno

Gli esperimenti relativi alla stazione di Morbegno (So), sita in via Cortivacci, si riferiscono alla previsione della pressione e della radiazione solare totale. Le figure che seguono mostrano il risultato ottenuto con la pressione, utilizzando una rete neurale con 30

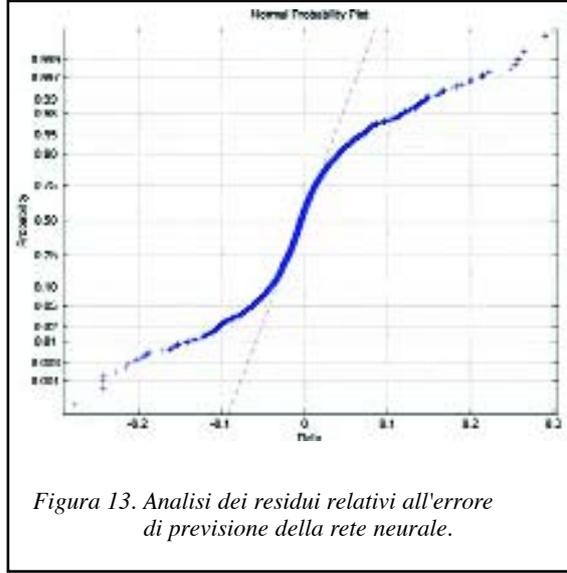
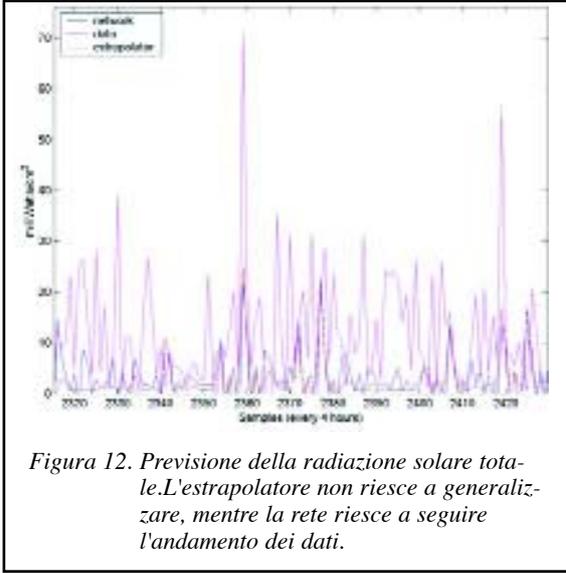
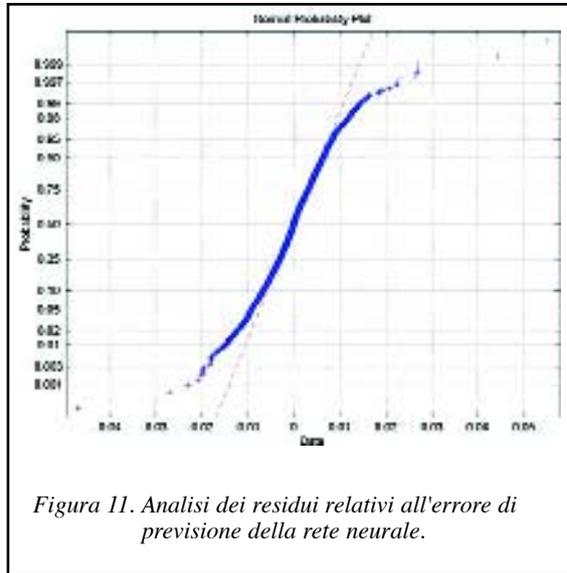
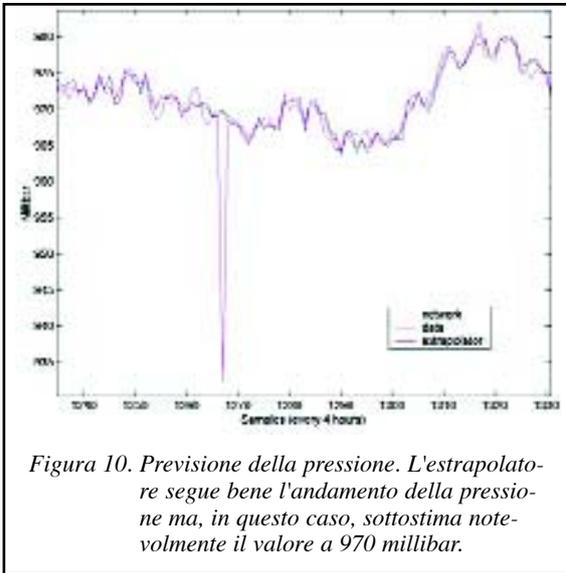
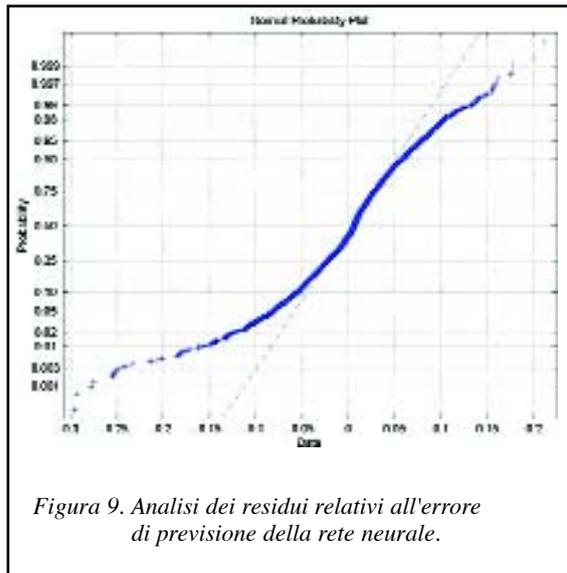
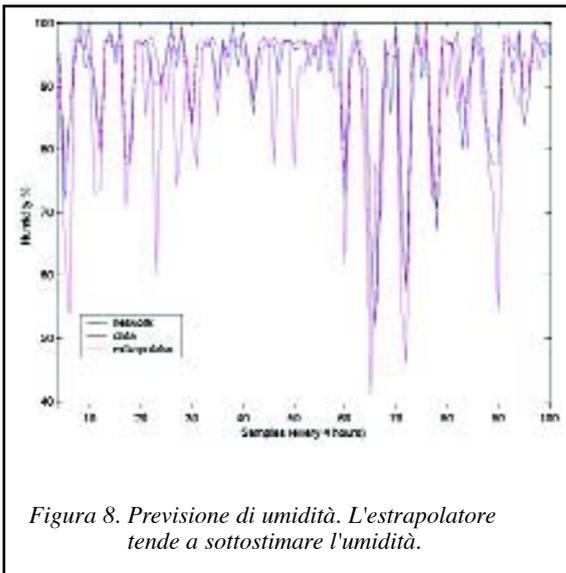
neuroni di input, 3 neuroni hidden e 1 neurone di output. Come si può notare, la rete neurale segue molto da vicino l'andamento della pressione, che risulta anche essere la grandezza che varia con maggiore "lentezza" nel corso del tempo, al contrario dell'umidità o della temperatura. L'errore medio commesso dalla rete è pari a 0.000072546 millibar con deviazione standard pari a 0.0057303 millibar. L'extrapolatore effettua la previsione commettendo un errore medio pari a 0.96864 millibar con deviazione standard pari a 2.225 millibar. I residui, anche in questo caso, risultano tendere ad essere gaussiani. Come si può notare in Figura 12, l'extrapolatore non riesce ad effettuare una previsione di buona qualità, probabilmente per la natura altamente non lineare della grandezza presa in esame. L'errore medio di previsione dell'extrapolatore è pari a 14.3406 milliWatts/cm² con una deviazione standard pari a 13.2205 milliWatts/cm². La rete utilizzata per effettuare la previsione è costituita da 27 neuroni di input, 3 neuroni hidden e 1 neurone di output e l'errore medio di previsione è risultato pari a -0.0026215 milliWatts/cm² con deviazione standard pari a 0.045158 milliWatts/cm². I residui di Figura 13 tendono ad essere gaussiani.

La stazione di Agrate Brianza

I parametri meteo utilizzati per l'esperimento, che verrà descritto, sono la velocità del vento e la radiazione solare totale. La rete neurale utilizzata in questo esperimento è costituita da 20 neuroni di input, 3 neuroni hidden e 1 neurone di output. L'errore medio di previsione nel caso della rete neurale è stato pari a -0.00099445 m/s e deviazione standard pari a 0.043226 m/s, mentre l'extrapolatore ha prodotto un errore medio di previsione pari a 0.36194 m/s e deviazione standard pari a 0.34474 m/s. Nel caso della previsione della radiazione solare totale si è utilizzata una rete con 25 neuroni di input, 3 neuroni hidden e 1 neurone di output. L'errore medio di previsione è risultato pari a -0.00083528 milliWatts/cm² con deviazione standard pari a 0.044515 milliWatts/cm². Nel caso dell'extrapolatore, l'errore medio di previsione è stato pari a 14.5401 milliWatts/cm² con deviazione standard pari a 14.472 milliWatts/cm².

Conclusioni

Il presente lavoro dimostra l'applicabilità delle reti neurali bayesiane con shortcut per applicazioni di tipo meteorologico. I risultati ottenuti mostrano l'estrema flessibilità del modello utilizzato, che si pro-



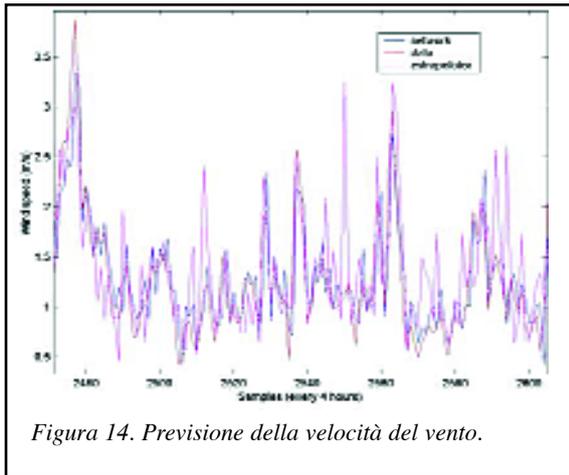


Figura 14. Previsione della velocità del vento.

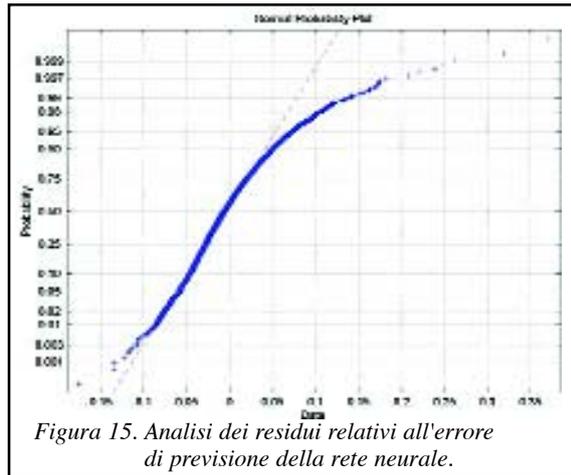


Figura 15. Analisi dei residui relativi all'errore di previsione della rete neurale.

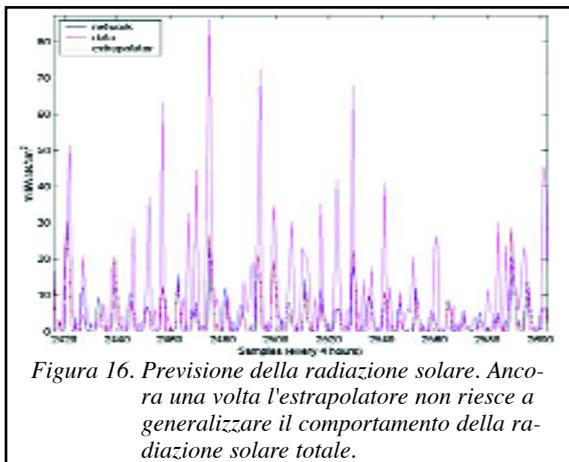


Figura 16. Previsione della radiazione solare. Ancora una volta l'extrapolatore non riesce a generalizzare il comportamento della radiazione solare totale.

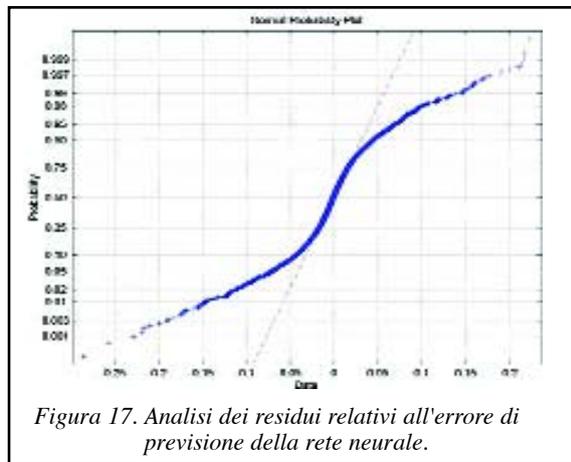


Figura 17. Analisi dei residui relativi all'errore di previsione della rete neurale.

pone come modello alternativo ai modelli fisici convenzionali nei casi in cui tali modelli risultino scarsamente applicabili. Inoltre per la non elevata richiesta di risorse computazionali necessarie nella fase di “forwarding” dei dati (la fase di formulazione della previsione), tali modelli possono essere facilmente utilizzati anche qualora non siano disponibili risorse di calcolo ingenti. Sviluppi futuri prevedono l'utilizzo delle stesse per la previsione di altri parametri più complessi quali copertura nuvolosa, visibilità o eventi piovosi tramite l'ausilio di più reti neurali connesse tra loro, la possibilità di effettuare previsioni oltre le 4 ore e la possibilità di ottenere più pre-

visioni simultanee (4 e 8 ore contemporaneamente ad esempio), aumentando il numero di output.

Ringraziamenti

Un particolare ringraziamento va alla regione Lombardia per aver messo a disposizione i dati delle stazioni meteorologiche.

To contact:

- michele.derosa@kwos.org
- alessandro.derosa@kwos.org
- raffaello.dimartino@kwos.org
- paolo.marrone@kwos.org

Bibliografia

1. MacKay, David J.C. Probable Networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks.
2. Bishop, Christopher M., Hinton G. Neural Networks for Pattern Recognition. Clarendon Press. 1995.
3. Cao, L. Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series.
4. Eleuteri, A., Acernese, F., Barone, F., De Rosa, R., Milano, L., Tagliaferri, R. A hierarchical Bayesian learning scheme

- for autoregressive neural networks: application to the CATS benchmark. IJCNN 2004.
5. Rosenblatt, F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms. Washington D.C.: Spartan 1962.

Altre informazioni possono essere ricavate dai siti:
www.inference.phy.cam.ac.uk/mackay;
www.kwos.org