

Sommario

Il Machine Learning si propone di ampliare la classe dei problemi trattabili laddove non risultino applicabili approcci algoritmici o analitici. Rilevante è dunque in questo ambito il trattamento di domini strutturati: poiché in numerosi campi le relazioni tra i dati possono essere rappresentate sotto forma di grafi, estendere la gamma di strategie utili ad apprendere trasduzioni strutturali può infatti dare risposta a problemi esistenti e permettere lo sviluppo di nuove applicazioni.

Nel contesto del Machine Learning, il trattamento di domini strutturati attraverso il paradigma neurale trova principalmente due ostacoli: l'alto costo computazionale e la presenza di vincoli sulla classe degli input trattabili. Il Reservoir Computing propone soluzioni ad entrambi i problemi, offrendo modelli efficienti in grado di apprendere trasduzioni strutturali definite su grafi generici.

Nel corso della tesi vengono introdotti nuovi modelli di Reti Neurali Ricorsive, nell'ambito del Reservoir Computing, per l'apprendimento supervisionato di trasduzioni su grafi. I modelli proposti si caratterizzano per l'adozione di un approccio costruttivo, innovativo nell'ambito del Reservoir Computing, che permette di individuare soluzioni ad alcuni dei problemi che caratterizzano i modelli esistenti. I modelli proposti permettono infatti la determinazione della topologia della rete attraverso un processo adattivo e la realizzazione di un meccanismo stabile di output-feedback per introdurre informazione supervisionata all'interno del processo di encoding del reservoir. La strategia costruttiva permette inoltre la realizzazione di modelli molto efficienti dal punto di vista computazionale.

L'analisi sperimentale svolta riguarda l'apprendimento di trasduzioni strutturali da dataset reali appartenenti all'ambito della Chemioinformatica.

Introduzione

In molti settori applicativi i dati trattati trovano una propria rappresentazione naturale in forma di grafi. Per citarne solo alcuni, si hanno esempi in questo senso nella Chimica [4, 34, 3], nella Proteomica [1], nel trattamento e riconoscimento di immagini [11] o nell’Ingegneria del Software [11]. Estendere le metodologie per affrontare problemi definiti su domini strutturati, rappresenta quindi un modo per dare risposte efficaci a problemi esistenti, aprendo spazio per lo sviluppo di nuove applicazioni.

L’ampliamento della gamma dei problemi trattabili è d’altra parte una delle ragioni d’essere del Machine Learning [35, 18]: laddove non risulti applicabile o efficace un approccio classico, algoritmico o analitico, l’Apprendimento Automatico si configura infatti come uno strumento utile per spostare la frontiera che delimita i problemi affrontabili.

Nell’ambito del Machine Learning il paradigma neurale ricopre un ruolo rilevante. Le Reti Neurali Artificiali [19, 5] rappresentano infatti un insieme vasto e variegato di modelli computazionali, efficacemente applicati a problemi reali. In analogia con il modello biologico, in una Rete Neurale Artificiale numerose unità (neuroni) sono collegate fra loro attraverso connessioni pesate

(sinapsi) che permettono ad un segnale (attivazione) di propagarsi attraverso la rete. La configurazione della topologia e dei pesi sulle connessioni ha dunque impatto sul modo in cui il segnale di input viene propagato e quindi sul segnale di output. Grazie a questo, le Reti Neurali Feedforward [5, 19], caratterizzate dal susseguirsi di più livelli di unità non lineari connesse fra loro secondo una topologia aciclica, risultano essere approssimatori universali [6], sono in grado cioè di approssimare con precisione arbitraria qualsiasi funzione continua. L'apprendimento supervisionato in una Rete Neurale Artificiale avviene dunque adattando i pesi delle connessioni con il fine di approssimare una funzione non nota, a partire da dati di esempio (i.e. coppie (\mathbf{x}, \mathbf{y}))).

L'idea di adottare un approccio costruttivo [42, 9, 27] nasce nel contesto delle Reti Neurali Feedforward con lo scopo di semplificare il processo di creazione e di allenamento dei modelli. La strategia proposta in questo caso sfrutta infatti un procedimento incrementale per la costruzione della rete nella sua totalità: iterativamente singole unità vengono aggiunte alla rete ed allenate, secondo un meccanismo guidato dalle caratteristiche del problema affrontato. La rete, dunque, cresce in maniera incrementale, sfruttando ad ogni iterazione tutte le informazioni precedentemente apprese.

L'adozione dell'approccio costruttivo offre il vantaggio di non dover stabilire a priori la topologia della rete, strettamente legata alla sua capacità computazionale, riducendo inoltre il costo computazionale dell'allenamento, che in questo caso risulta circoscritto a singole unità e può dunque essere svolto tramite algoritmi più semplici ed efficienti. Da un punto di vista più generale, la strategia costruttiva evidenzia inoltre come il paradigma neurale offra naturalmente gli strumenti necessari alla suddivisione di un problema in

sotto-problemi, che si rispecchia nella visione concettuale di una rete come formata da più parti che cooperano fra loro.

Benché sia stato a lungo legato all'apprendimento di funzioni definite su vettori, l'approccio neurale ha visto nascere paradigmi e modelli orientati all'apprendimento su domini più complessi. È il caso delle Reti Neurali Ricorrenti [48, 8], che generalizzano le Reti Neurali Feedforward al trattamento di sequenze, attraverso l'introduzione di cicli nella topologia delle connessioni fra le unità.

L'applicazione di reti di questo tipo trova tuttavia un ostacolo nell'alto costo computazionale generalmente richiesto per il loro allenamento e la loro analisi. Per dare soluzione a questo tipo di problema nasce, nell'ambito delle Reti Neurali Ricorrenti, il paradigma del Reservoir Computing [29, 49, 25]: una classe di modelli caratterizzati dalla separazione concettuale tra una parte ricorrente (i.e. con connessioni cicliche fra le unità) ed una parte non-ricorrente da cui ricavare l'output. L'opportunità di limitare l'apprendimento alla sola parte non-ricorrente del modello, lasciando inalterata la parte ricorrente, rappresenta il punto di forza dell'approccio individuato dal Reservoir Computing, consentendo l'impiego di strategie di allenamento molto efficienti dal punto di vista computazionale. A fronte di questo vantaggio, tuttavia, la presenza di connessioni che non vengono interessate da alcun tipo di adattamento rappresenta un limite per i modelli esistenti di Reservoir Computing: parte della rete risulta infatti fissata a priori e non ha la capacità di adattarsi al problema affrontato.

L'obiettivo dell'apprendimento di funzioni definite su domini strutturati (i.e. trasduzioni strutturali) è invece affrontato, in ambito neurale, attraverso

le Reti Neurali Ricorsive [11, 44], che generalizzano le Reti Neurali Ricorrenti. Il legame fra la topologia dell'input e le dinamiche interne della rete impone tuttavia in questo caso dei vincoli sulla classe degli input trattabili. Nella loro formulazione originale, infatti, le Reti Neurali Ricorsive risultano applicabili unicamente a dati strutturati che presentino un ordinamento topologico fra i vertici, limitando di conseguenza il dominio di input ai soli Grafi Diretti Aciclici. Tale assunzione ha impatto sulla capacità dei modelli [33], riducendo il numero di problemi efficacemente affrontabili.

Nella storia recente delle Reti Neurali Ricorsive si ritrovano vari modelli tesi all'ampliamento della classe degli input trattabili, orientati dunque all'apprendimento di trasduzioni strutturali generiche. Due approcci distinti, in particolare, lasciano emergere elementi interessanti per la realizzazione di modelli neurali in grado di trattare domini strutturati generici (i.e. grafi). Il modello *Neural Networks for Graphs* (NN4G) [32] sfrutta un approccio costruttivo per evitare i problemi legati alla presenza di cicli nell'input. L'adozione di una strategia costruttiva permette inoltre in questo caso di sfruttare localmente informazioni globali, legate alla topologia dell'input. Ad ogni ciclo del processo di costruzione della rete, infatti, nuove informazioni relative alla struttura dei dati vengono raccolte e rese disponibili per le unità aggiunte successivamente.

Un secondo esempio di Reti Neurali Ricorsive per il trattamento di grafi è dato dal modello *Graph Echo State Network* (GraphESN) [13], appartenente all'ambito del Reservoir Computing, che sfrutta le caratteristiche strutturali della rete per realizzare il trattamento di input generici. Ne risulta dunque un modello di Rete Neurale Ricorsiva efficiente per l'apprendimento di tra-

sduzioni strutturali definite su grafi, che tuttavia mantiene inalterate alcune delle criticità caratteristiche del Reservoir Computing.

Da quanto detto emergono dunque necessità e limitazioni legate all'apprendimento di trasduzioni strutturali attraverso il paradigma neurale. In una disciplina fortemente orientata alle applicazioni come il Machine Learning, il costo computazionale e la classe degli input trattabili rappresentano infatti dei vincoli molto forti. Il paradigma del Reservoir Computing offre soluzioni efficaci per rispondere a simili necessità, ma presenta tuttavia degli elementi critici. In particolare la presenza di una porzione di rete fissata a priori, e mantenuta inalterata durante l'apprendimento, si configura come una limitazione. La determinazione della topologia di rete appropriata e l'opportunità di modificare le dinamiche dell'intera rete sulla base del problema affrontato, sono infatti limiti la cui risoluzione rimane ad oggi un problema aperto [28, 52, 38]. L'approccio costruttivo offre tuttavia soluzioni in tal senso, proponendo una strategia efficiente per determinare la topologia della rete in maniera automatica e per sfruttare le informazioni raccolte nel corso del processo incrementale di costruzione della rete.

Quanto verrà discusso nel seguito ha dunque l'obiettivo di introdurre nuovi modelli di Reti Neurali Ricorsive appartenenti all'ambito del Reservoir Computing, che permettano l'adozione di una strategia costruttiva nell'apprendimento di trasduzioni strutturali. In particolare, i modelli descritti nel seguito rappresentano un'estensione delle GraphESN, caratterizzandosi dunque come modelli efficienti per il trattamento di domini strutturati generici. L'introduzione della strategia costruttiva, innovativa nell'ambito del Reservoir Computing, ha impatto sui modelli proposti dal punto di vista computazionale

ed algoritmico, riducendo il costo computazionale ed evitando il problema di dover fissare a priori la topologia della rete. Applicato al trattamento dei domini strutturati, inoltre, l'approccio costruttivo offre l'opportunità di sfruttare localmente informazioni globali e supervisionate. Tale caratteristica viene usata, nei modelli proposti, per influenzare le dinamiche di quella porzione di rete che nel Reservoir Computing non è soggetta ad allenamento. Ne risulta un meccanismo stabile per introdurre, anche in assenza di una fase di adattamento specifica, dell'informazione supervisionata che possa influenzare le dinamiche della rete in maniera consistente con il problema affrontato.

Le capacità dei modelli introdotti sono testate in maniera rigorosa su problemi di Chemioinformatica relativi a dataset reali noti in letteratura. Parte dei risultati riportati nel seguito è inoltre oggetto dell'articolo scientifico *Constructive Reservoir Computation with Output Feedbacks for Structured Domains*, che sarà discusso nel corso dello European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning¹.

L'esposizione degli argomenti nel testo è articolata come segue:

Nel capitolo 1 si introducono i temi, le tecniche ed i modelli alla base del lavoro svolto. Nel corso di questa rassegna dello stato dell'arte vengono in particolare presentati i principi dell'approccio costruttivo e del Reservoir Computing, descrivendo anche modalità ed obiettivi nel trattamento dei domini strutturati.

Nel capitolo 2 vengono introdotti i modelli originali, oggetto del lavoro di tesi, ne vengono presentate le caratteristiche e discussi gli aspetti

¹<http://www.dice.ucl.ac.be/esann/>

specifici.

Nel capitolo 3 sono descritti i risultati sperimentali ottenuti testando le capacità dei modelli su una serie di problemi del mondo reale appartenenti all'ambito della Chemioinformatica.


Nel capitolo 4 la tesi si conclude con una valutazione del lavoro svolto nonché l'individuazione di quegli aspetti che potrebbero essere oggetto di un ulteriore, specifico, lavoro di indagine e sperimentazione.



Bibliografia

- [1] Pierre Baldi e Gianluca Pollastri. The principled design of large-scale recursive neural network architectures–DAG-RNNs and the protein structure prediction problem. *J. Mach. Learn. Res.*, 4:575–602, December 2003.
- [2] Yoshua Bengio, Paolo Frasconi, e Patrice Simard. The problem of learning long-term dependencies in recurrent networks. In *Neural Networks, 1993., IEEE International Conference on*, pp. 1183–1188 vol.3, 1993.
- [3] Carlo Bertinetto, Celia Duce, Alessio Micheli, Roberto Solaro, Antonina Starita, e Maria Rosaria Tiné. Evaluation of hierarchical structured representations for QSPR studies of small molecules and polymers by recursive neural networks. *Journal of Molecular Graphics and Modelling*, 27(7):797–802, 2009.
- [4] Anna Maria Bianucci, Alessio Micheli, Alessandro Sperduti, e Antonina Starita. Application of cascade correlation networks for structures to chemistry. *Appl. Intell.*, 12(1-2):115–145, 2000.

-
- [5] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.
 - [6] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314, 1989.
 - [7] Kenji Doya. Bifurcations of recurrent neural networks in gradient descent learning. *IEEE Transactions on Neural Networks*, 1:75–80, 1993.
 - [8] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
 - [9] Scott E. Fahlman e Christian Lebiere. The Cascade-Correlation learning architecture. In *Advances in Neural Information Processing Systems 2*, pp. 524–532. Morgan Kaufmann, 1990.
 - [10] Chrisantha Fernando e Sampsa Sojakka. Pattern recognition in a bucket. In *Advances in Artificial Life*, Lecture Notes In Computer Science, pp. 588–597. Springer Berlin / Heidelberg, 2003.
 - [11] Paolo Frasconi, Marco Gori, e Alessandro Sperduti. A general framework for adaptive processing of data structures. *IEEE Transactions on Neural Networks*, 9:768–786, 1998.
 - [12] Holger Fröhlich, Jörg K. Wegner, Florian Sieker, e Andreas Zell. Optimal assignment kernels for attributed molecular graphs. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pp. 225–232, New York, NY, USA, 2005. ACM.

- [13] Claudio Gallicchio e Alessio Micheli. Graph Echo State Networks. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN) 2010*, pp. 1–8, July 2010.
- [14] Claudio Gallicchio e Alessio Micheli. TreeESN: a preliminary experimental analysis. In *Proceedings of the ESANN 2010*, pp. 333–338, d-side, 2010.
- [15] Claudio Gallicchio e Alessio Micheli. Architectural and markovian factors of Echo State Networks. *Neural Networks*, 24(5):440–456, 2011.
- 
- [16] Barbara Hammer, Alessio Micheli, e Alessandro Sperduti. Universal approximation capability of cascade correlation for structures. *Neural Computation*, 17(5):1109–1159, 2005.
- [17] Barbara Hammer e Peter Tiño. Recurrent neural networks with small weights implement definite memory machines. *Neural Computation*, 15:1897–1929, 2003.
- [18] Trevor Hastie, Robert Tibshirani, e Jerome H. Friedman. *The Elements of Statistical Learning*. Springer, corrected edizione, July 2003.
- [19] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Macmillan, New York, 1994.
- [20] Christoph Helma, Ross D. King, Stefan Kramer, e Ashwin Srinivasan. The predictive toxicology challenge 2000-2001. *Bioinformatics*, 17(1):107–108, 2001.

-
- [21] Guang-Bin Huang, Qin-Yu Zhu, e Chee-Kheong Siew. Extreme Learning Machine: theory and applications. *Neurocomputing*, 70(1-3):489–501, December 2006.
- [22] Herbert Jaeger. The “echo state” approach to analyzing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for information Technology, 2001. <http://www.faculty.iu-bremen.de/hjaeger/pubs/EchoStatesTechRep.pdf>.
- [23] Herbert Jaeger. Short term memory in Echo State Networks. Technical Report GMD Report 152, German National Research Center for information Technology, 2001. <http://www.faculty.iu-bremen.de/hjaeger/pubs/STMEchoStatesTechRep.pdf>.
- [24] Herbert Jaeger. Reservoir riddles: Suggestions for Echo State Network research (extended abstract of invited talk). In *Proceedings of the International Joint Conference on Neural Networks 2005*, pp. 1460–1462, August 2005.
- [25] Herbert Jaeger e Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, April 2004.
- [26] Herbert Jaeger, Wolfgang Maass, e Jose Principe. Editorial: Special issue on Echo State Networks and Liquid State Machines. *Neural Networks*, 20(3):287–289, 2007.
- [27] Enno Littmann e Helge Ritter. Learning and generalization in cascade network architectures. *Neural Comput.*, 8:1521–1539, October 1996.

- [28] Mantas Lukoševičius. Echo State Networks with trained feedbacks. Technical Report No. 4, Jacobs University Bremen, 2007. http://wwwback.jacobs-university.de/imperia/md/content/groups/research/techreports/tfbesn_iubtechreport.pdf.
- [29] Mantas Lukoševičius e Herbert Jaeger. Reservoir Computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, August 2009.
- [30] Wolfgang Maass, Thomas Natschläger, e Henry Markram. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput*, 14(11):2531–60, 2002.
- [31] Mario Martelli. *Introduction to Discrete Dynamical Systems and Chaos*. Wiley, 1999.
- [32] Alessio Micheli. Neural Network for Graphs: a contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3):498–511, 2009.
- [33] Alessio Micheli, Diego Sona, e Alessandro Sperduti. Contextual processing of structured data by recursive cascade correlation. *IEEE Transactions on Neural Networks*, 15:1396–1410, 2003.
- [34] Alessio Micheli, Alessandro Sperduti, Antonina Starita, e Anna Maria Bianucci. Analysis of the internal representations developed by neural networks for structures applied to quantitative structure-activity relationship studies of benzodiazepines. *Journal of Chemical Information and Computer Sciences*, 41(1):202–218, 2001.

-
- [35] Tom Mitchell. *Machine Learning*. McGraw Hill, 1997.
- [36] Lutz Prechelt. Investigation of the CasCor family of learning algorithms. *Neural Networks*, 10:885–896, 1996.
- [37] Danil Prokhorov. Echo State Networks: appeal and challenges. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pp. 1463–1466 vol. 3, 2005.
- [38] Felix R. Reinhart e Jochen J. Steil.
- [39] Dana Ron, Yoram Singer, e Naftali Tishby. The power of amnesia: Learning probabilistic automata with variable memory length. In *Machine Learning*, pp. 117–149, 1996.
- [40] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, e Gabriele Monfardini. The Graph Neural Network Model. *IEEE Transactions on Neural Networks*, 20(1):61–80, January 2009.
- [41] Benjamin Schrauwen, David Verstraeten, e Jan Van Campenhout. An overview of Reservoir Computing: theory, applications and implementations. In *Proceedings of the 15th European Symposium on Artificial Neural Networks*, pp. 471–482, April 2007.
- [42] Frank J. Śmieja. Neural network constructive algorithms: trading generalization for learning efficiency? *Circuits Syst. Signal Process.*, 12:331–374, February 1993.

-
- [43] Qingsong Song e Zuren Feng. Effects of connectivity structure of complex Echo State Network on its prediction performance for nonlinear time series. *Neurocomputing*, 73(10-12):2177 – 2185, 2010.
- [44] Alessandro Sperduti e Antonina Starita. Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8:714–735, 1997.
- [45] Ashwin Srinivasan, Stephen H. Muggleton, Ross D. King, e Michael J. E. Sternberg. Mutagenesis: ILP experiments in a non-determinate biological domain. In *Proceedings of the 4th International Workshop on Inductive Logic Programming*, volume 237, pp. 217–232, 1994.
- [46] Jochen J. Steil. Backpropagation-Decorrelation: online recurrent learning with $O(N)$ complexity. In *Proc. IJCNN*, volume 1, pp. 843–848, July 2004.
- [47] Peter Tiño, Michal Čerňanský, e Ľubica Beňušková. Markovian architectural bias of recurrent neural networks. *IEEE Transactions on Neural Networks*, 15(1):6–15, 2004.
- [48] Ah Chung Tsoi e Andrew Back. Discrete time recurrent neural network architectures: A unifying review. *Neurocomputing*, 15(3-4):183–223, 1997.
- [49] David Verstraeten, Benjamin Schrauwen, Michiel D’Haene, e Dirk Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391–403, April 2007.

-
- [50] Paul J. Werbos. Backpropagation Through Time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, August 2002.
- [51] Ronald J. Williams e David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Comput.*, 1(2):270–280, 1989.
- [52] Francis Wyffels, Benjamin Schrauwen, e Dirk Stroobandt. Stable output feedback in reservoir computing using ridge regression In *Proceedings of the 18th International Conference on Artificial Neural Networks*. A cura di V. Kurkova, R. Neruda, e J. Koutnik, pp. 808–817, Prague, 2008. Springer.