
Linguistica Computazionale – Lezione 15

Approccio sub-simbolico al NLP (e alla MT)

Mercoledì 12 Aprile 2007

Cristiano Chesi, chesi@media.unisi.it

Approcci subsimbolici al NLP

□ Indice

- Problemi che si prestano ad un approccio subsimbolico
 - in cosa consiste
 - quando e perché lo si preferisce ad un approccio simbolico
- Reti neurali
 - neurofisiologia
 - reti di neuroni artificiali
 - apprendimento del passato in inglese
- NLP con reti neurali: Simple Recurrent Networks
 - memoria a breve termine
 - apprendere incrementalmente il linguaggio
 - usare reti neurali per la MT
 - acquisire proprietà ricorsive con le SRN

Lecture, approfondimenti

□ Bibliografia essenziale

- Christiansen & Charter (1999) *Toward a Connectionist Model of Recursion in Human Linguistic Performance*. *Cognitive Science* 23(2):157-205
- Elman (1999) *Origins of language: a conspiracy theory*. In MacWhinney "The emergence of language".

□ Approfondimenti

- Elman, Bates, Karmiloff-Smith, Parisi, Plunket (1996) *Rethinking Innateness*. MIT Press (Cap. 2)
- Elman (1993) *Learning and development in neural networks: the importance of starting small*. *Cognition* 48:71-99
- Rohde & Plaut (2001) *Less is less in language acquisition*. In Quinlin "Connectionist modelling of cognitive development".
- Kandel, Schwartz, Jessell (1999) *Fondamenti delle neuroscienze e del comportamento*. Ambrosiana (Cap. 1-4)

Cosa significa "subsimbolico"

Problemi "subsimbolici"

- un approccio **subsimbolico** "usa" **rappresentazioni implicite**
 - **dynamic functional system** (Luria) ogni comportamento, o funzione cognitiva, è il frutto di un concerto di componenti localizzate in varie aree cerebrali
 - **Parallel Distributed Processing** (PDP, Rumelhart, McClelland & al.) la *conoscenza* (o *competenza*, da un punto di vista linguistico) non è una "memoria" ma una *struttura di processamento*
 - solo l'input e l'output vengono codificati come **entità "discrete"**;
 - nessuna combinazione delle componenti dell'input secondo operazioni primitive per formare **strutture esplicite**
 - la **complessità** del sistema (quindi delle sue rappresentazioni apparenti) è una **proprietà emergente** dalla semplice interazione delle sue parti

Quali problemi si prestano all'analisi subsimbolica

Problemi "subsimbolici"

- problemi difficilmente analizzabili in sottoproblemi più semplici
 - problemi molto **complessi** da descrivere (non lineari)
 - **rappresentazioni parziale** dello spazio del problema
 - algoritmi complessi che richiedono **approssimazioni**
 - difficoltà di elaborazione di **regole** e/o **euristiche**
 - alto grado di interazione tra le varie sottocomponenti (**multiple constraints**)

5

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Cos'è la grammatica

Problemi "subsimbolici"

- **Approcci simbolici**
la grammatica è un insieme statico di regole o principi (rappresentazione esplicita della competenza linguistica)
- **Approcci (subsimbolici) a reti neurali**
Grammatica è una struttura di processamento (quindi una forma di rappresentazione implicita) variabile dinamicamente in dipendenza dal contesto; le parole sono operatori che muovono il sistema verso certi stati piuttosto che altri

6

Linguistica Computazionale A.A. 2006-07 – C. Chesì

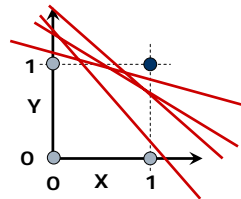
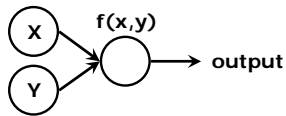
Un esempio

Problemi "subsimbolici"

- definire gli operatori logici
 - AND da un punto di vista simbolico:

X	AND	Y
1	1	1
1	0	0
0	0	1
0	0	0

- AND da un punto di vista subsimbolico:



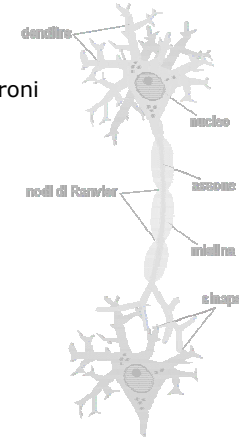
7

Linguistica Computazionale A.A. 2006-07 – C. Chesì

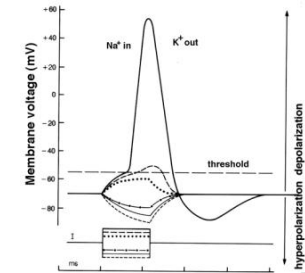
Componenti del sistema nervoso centrale

Neurofisiologia

- **neuroni**



- **potenziale sinaptico**



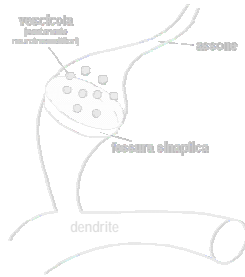
8

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Componenti del sistema nervoso centrale

Neurofisiologia

■ sinapsi



■ modalità di trasmissione sinaptica

- sinapsi **elettrica** (continuità citoplasmatica tra i neuroni, trasmissione via correnti ioniche, ritardo praticamente assente, comunicazione bidirezionale)
- sinapsi **chimica** (vescicole presinaptiche e zone attive + recettori postsinaptici, trasmettitori di natura chimica, ritardo di 0,3-5 ms o più, unidirezionale)

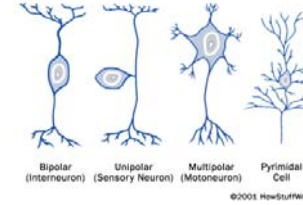
9

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Componenti del sistema nervoso centrale

Neurofisiologia

■ tipologia



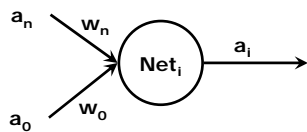
10

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti neurali artificiali

Reti Neurali

■ Nodi



$$Net_i = \sum_{j=0}^n w_{ij} a_j$$

$$a_i = f(net_i)$$

■ Connessioni

- Inibitorie
- eccitatorie

- Idea fondamentale: semplici unità di processamento collegate da connessioni pesate. Il risultato può essere un'interazione estremamente complessa (proprietà emergenti); brain-like solo a confronto dei computer seriali

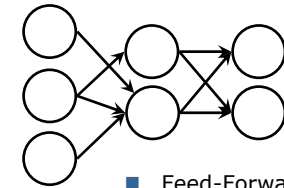
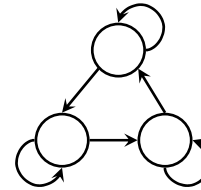
11

Linguistica Computazionale A.A. 2006-07 – C. Chesì

L'architettura delle reti neurali artificiali

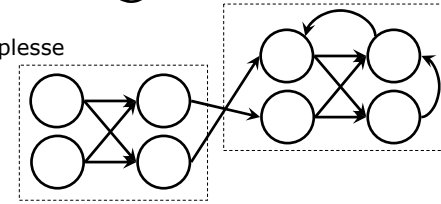
Reti Neurali

■ Completamente ricorrente



■ Feed-Forward

■ Complesse



12

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti "classiche"

Reti Neurali

- **Psicone** = operazione logica primitiva (McCulloch & Pitts, introducono l'inibizione)
- **Percettrone** (Roseblatt)
- **Pattern associators**
- **Mappe di Kohonen** (o reti associative)
- **Multilayer networks**

13

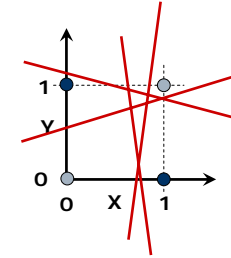
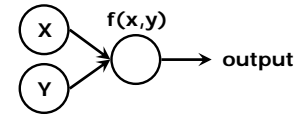
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Un problema "irrisolvibile"

Reti neurali

- un operatore logico un po' ostico (Minsky & Papert 69)
 - XOR da un punto di vista simbolico:

X	XOR	Y
1	0	1
1	1	0
0	1	1
0	0	0



14

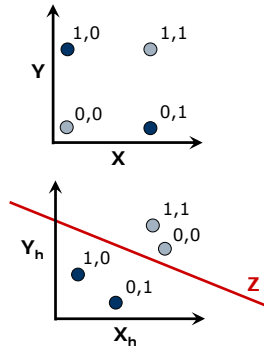
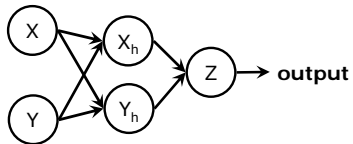
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Un problema "irrisolvibile"

Reti neurali

- un operatore logico un po' ostico (Minsky & Papert 69)
 - XOR da un punto di vista simbolico:

X	XOR	Y
1	0	1
1	1	0
0	1	1
0	0	0

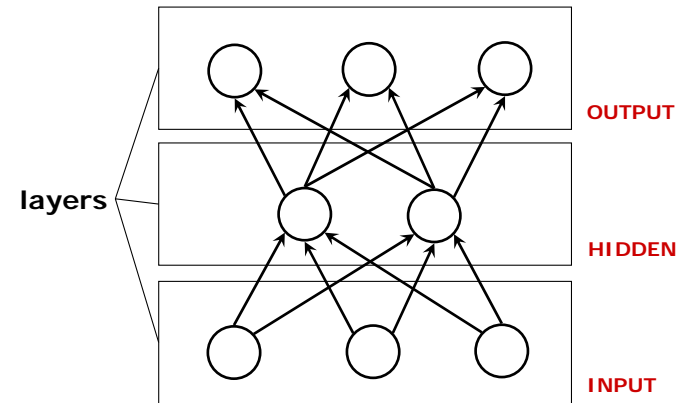


15

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Struttura delle reti multilayers

Reti Neurali



16

Linguistica Computazionale A.A. 2006-07 – C. Chesì

L'apprendimento nelle reti

Reti Neurali

- Apprendimento **supervisionato**
si informa la rete del fatto che ha sbagliato
- Apprendimento **non supervisionato**
l'apprendimento avviene implicitamente, prescindendo da qualsiasi informazione riguardo alla bontà dell'output prodotto
- **Hebbian learning** (Hebb 49)
 - "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (Hebb 1949:62)
 - $\Delta w_{ij} = \eta a_i a_j$
- **Perceptron Convergence Procedure** (PCP, Rosenblatt 59)
 - calcolare la differenza tra l'output **reale** e quello **atteso**.
 - obiettivo: minimizzare l'errore, **premiando** o **punendo** le connessioni di input (variando i pesi delle connessioni o la funzione di attivazione)

17

Linguistica Computazionale A.A. 2006-07 – C. Chesì

L'apprendimento nelle reti

Reti Neurali

■ Back Propagation (semplificata)

- nelle reti a molti livelli l'errore calcolato sulle unità di output viene ridistribuito proporzionalmente alle unità nascoste e quindi a quelle di input
- $\Delta w_{ij} = \eta \delta_{ip} o_{jp}$
dove
 - i = neurone di arrivo
 - j = neurone di partenza
 - p = pattern di attivazione
 - η = tasso di apprendimento (learning rate)
 - δ_{ip} = riflette l'errore di i rispetto al pattern p
 - o_{jp} = garantisce che l'errore sia proporzionale all'attivazione proveniente dal neurone j

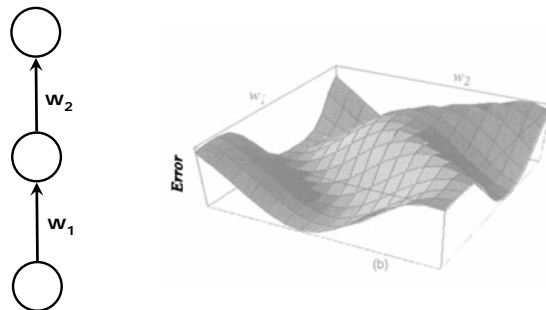
18

Linguistica Computazionale A.A. 2006-07 – C. Chesì

L'apprendimento nelle reti

Reti Neurali

- Apprendimento: alla ricerca del gradiente discendente



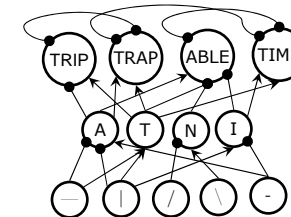
19

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Come si sceglie l'architettura di una rete?

Reti Neurali

- Le euristiche migliori sono le assunzioni teoriche (Rumelhart & McLelland 1982)



- Reti a **tre livelli** risolvono qualsiasi problema dal punto di vista teorico (Hornik, Stinchcombe & White 1989), ma non si sa a priori quanti neuroni ci vogliono o come devono essere interconnessioni

20

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Come si codifica l'input (e l'output)

Reti Neurali

- Raccolta dati di input
es. lessico e insieme di frasi che si vogliono testare
- Eventuali ipotesi di scomposizione dell'input in tratti
es. tratti fonologici; radici+suffissi
- Calcolo dei bit di informazione > numero unità di input
es. considerando una codifica binaria un insieme di 4 elementi (a, b, c, d) richiede, al minimo, 2 neuroni di input (a=00, b=01, c=10, d=11)
- Input ortogonali
input che non presentano nessuna "similarità" in termini di codifica (le generalizzazioni possibili su tali input sono solo legate all'ordine di presentazione, cioè alla distribuzione relativa del campione)
rappresentazione localista: 1 concetto = 1 nodo
ad esempio date 4 unità di input:
il (0001) cane (0010) corre (0100) . (1000)

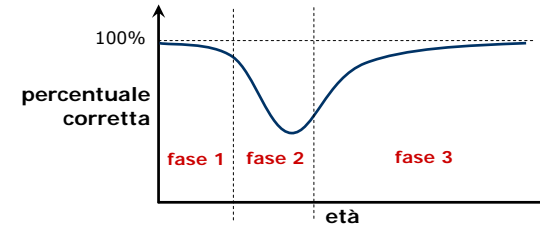
21

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Un esempio di rete che apprende un aspetto linguistico

Reti Neurali

- Apprendere il passato (Rumelhart & McClelland, 86)
 - Modello linguistico molto chiaro:
 - fase 1: pochi verbi alta frequenza (ipotesi i bambini hanno memorizzato poche forme complete)
 - fase 2: iperregolarizzazione anche sugli irregolari (cadere>cadò)
 - fase 3: uso verbi irregolari, coesistenza delle forme regolari e irregolari... fino alla scomparsa degli errori



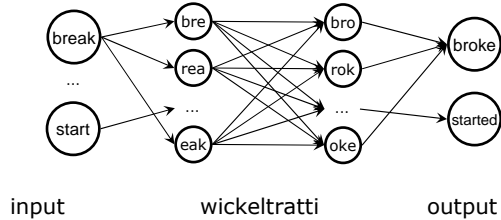
22

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Un esempio di rete che apprende un aspetto linguistico

Reti Neurali

- Apprendere il passato (Rumelhart & McClelland, 86)
 - non minimizzazione assoluta dell'errore, ma performance human-like
 - codifica dell'input e dell'output fonetica (Wickeltratti, da Wickelgren 69)
 - Struttura della rete (460 unità di input e di output codificati in wickeltratti)



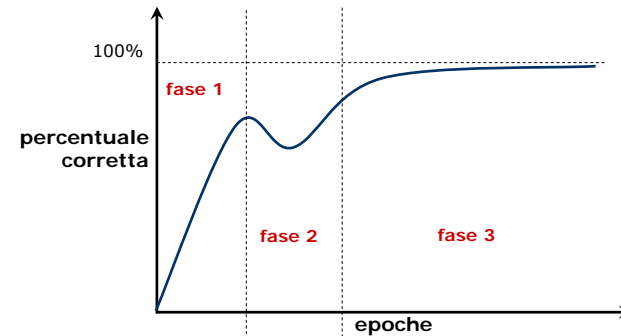
23

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Un esempio di rete che apprende un aspetto linguistico

Reti Neurali

- Apprendere il passato (Rumelhart & McClelland, 86)
 - risultati (dopo 420 verbi presentati 200 volte, cioè 84.000 esempi)



24

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Il tempo nelle reti neurali artificiali

Reti Neurali

- **Epoche**
unità temporale (discreta) durante la quale un input è ricevuto e processato dalla rete
- **Processing atemporale**
l'attivazione delle unità dipende soltanto dall'input corrente
- **Trucchetto per rappresentare simultaneamente il tempo**
dividere i nodi di input in gruppi; assegnare ad ogni gruppo un'unità temporale, quindi considerare l'elaborazione come frutto di una generalizzazione temporale anziché semplicemente sincrona
- **L'idea del context layer (Elman 1990)**
lo stato di attivazione dei neuroni nel livello nascosto viene copiato su un livello separato che poi viene ridato in pasto al livello nascosto all'epoca successiva

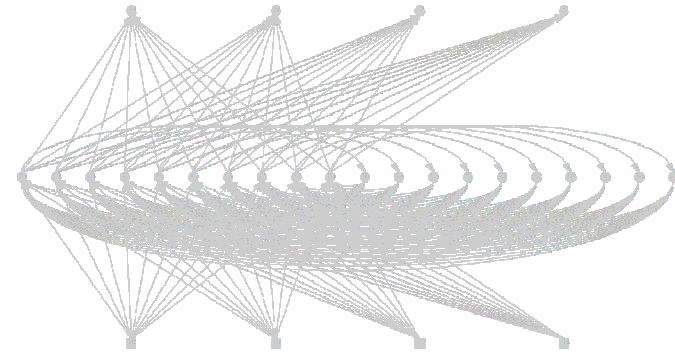
25

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(inizio ...)

Reti Neurali



powered by TLearn

26

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- **Prevedere la parola successiva**
es. La casa è sulla collina
Input = La
Output richiesto = casa

apprendimento "non supervisionato" (**auto-supervisionato**)
- **Plausibilità psico/neurologica (Cole & Robbins 92)**
una specie di priming
- **struttura input**
 - localista (1 nodo = 1 concetto)
 - rappresentazione distribuita (le stesse unità contribuiscono con la stessa attivazione a diversi concetti: nel pattern di attivazione risiede la differenza)

27

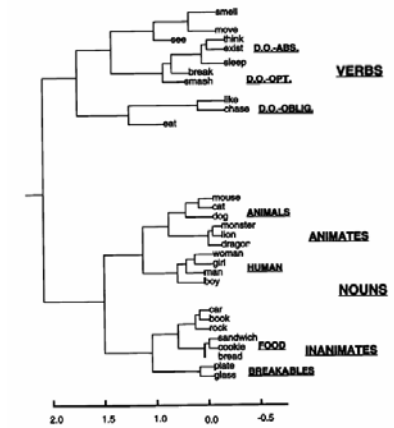
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- **Hierarchical clustering:**
rappresentazione (implicita) della conoscenza all'interno della rete
 - si cattura, sotto forma di vettore, l'attivazione delle unità nascoste per ogni input
 - si calcola la distanza euclidea tra tali vettori



28

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- The importance of starting small (Elman 93)
 - **obiettivo:** dimostrare che una SRN può gestire informazioni complesse e gerarchicamente organizzate
 - **ipotesi:** l'apprendimento è in funzione dello sviluppo
 - proprietà linguistiche che si cercavano di far apprendere alla rete:
 1. accordo soggetto verbo
 2. sottocategorizzazione verbale
 3. incassamento multiploes. *boys who chase dogs see girls*
dogs see boys who cats who mary feeds chase

(ognuna di queste proprietà causerebbe problemi alle macchine a stati finiti)

29

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- The importance of starting small (Elman 93)
 - fornendo subito tutti gli esempi alla rete, l'apprendimento non converge (es. *the boys who the girl *chase...*)
 - dividendo gli esempi per **complessità** e dando in pasto alla rete prima gli esempi più semplici la rete apprende:
 1. 10.000 frasi semplici (5 epoche di apprendimento)
 2. 7.500 frasi semplici + 2.500 frasi complesse (5 epoche)
 3. 50% frasi semplici + 50% frasi complesse (5 epoche)
 4. 25% frasi semplici + 75% frasi complesse (5 epoche)
 5. 10.000 frasi complesse (5 epoche)
 - **limitando** progressivamente la **memoria** la rete apprende:
 1. cancellazione unità contestuali ogni 3/4 parole (12 epoche)
 2. ogni 4/5 parole (5 epoche)
 3. ogni 5/6 parole (5 epoche)
 4. ogni 7 parole (5 epoche)
 5. nessuna cancellazione (5 epoche)

30

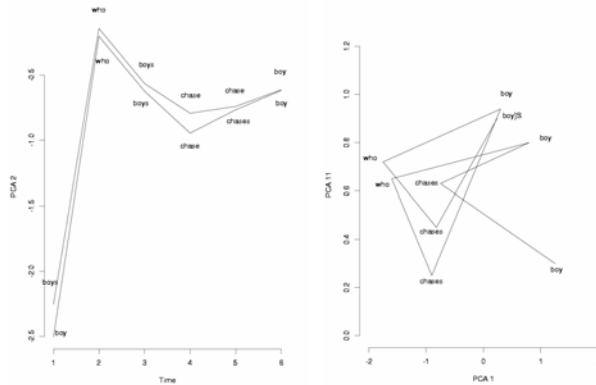
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- Principle Component Analysis (PCA, Gonzalez & Wintz 77)



31

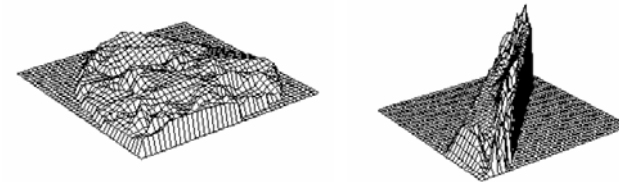
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

- Come mai "menomare" la rete produce buoni risultati?



- l'effetto dell'apprendimento sulle frasi semplici limita lo spazio del problema ad una regione più circoscritta (flessibilità all'inizio dell'apprendimento, maggiore rigidità alla fine)

32

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... continua ...)

Reti Neurali

■ Less is less (Rohde & Plaut 01)

dimostrano che lo "starting small" e' inutile e ridondante data la naturale tendenza delle SRN ad **apprendere prima le relazioni locali**

- Cleeremans & al. (89) dimostrano che le SRN apprendevano meglio le relazioni a distanza (es. accordo sogg-verbo) in grammatiche a stati finiti se l'elemento interveniente era "semanticamente" correlato al soggetto (es. il cane [che abbaia] è scappato)

vengono quindi create **5 classi** di grammatiche:

- A. nessun vincolo semantico tra frase principale e frase incassata
- B. 25% di vincoli semantico
- C. 50% di vincoli semantici
- D. 75% di vincoli semantici
- E. 100% di vincoli semantici

33

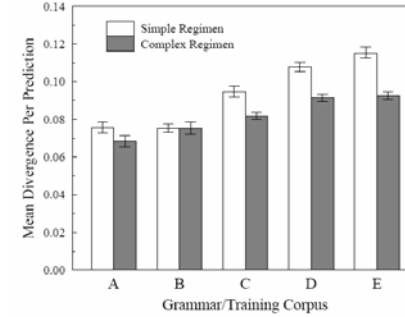
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Simple Recurrent Networks

(... fine)

Reti Neurali

■ Less is less (Rohde & Plaut 01)



media sui 16 migliori apprendimenti dei 20 test fatti

34

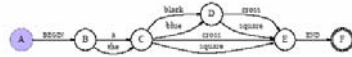
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti e traduzione automatica

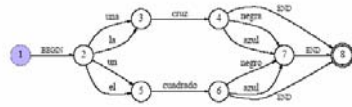
(inizio ...)

Reti Neurali

■ Finite State Automata (FSA) parsing



(a) Input FSA



(b) Output FSA

Prat, Casacuberta & Castro (2001) *Machine Translation with Grammar Association: Combining Neural Networks and Finite-State Models*. II Workshop on Natural Language Processing and Neural Networks

35

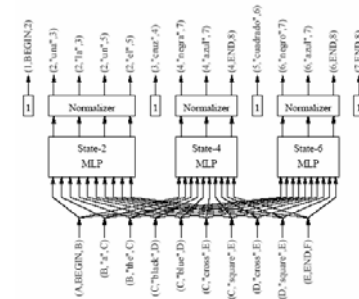
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti e traduzione automatica

(... continua ...)

Reti Neurali

■ Multi Layer Perceptron (MLP)



Prat, Casacuberta & Castro (2001) *Machine Translation with Grammar Association: Combining Neural Networks and Finite-State Models*. II Workshop on Natural Language Processing and Neural Networks

36

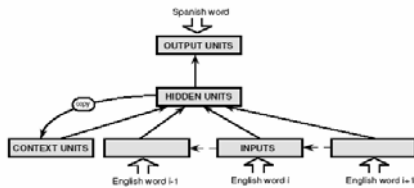
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti e traduzione automatica

(... continua ...)

Reti Neurali

■ (Delayed) SRN



Castano & Casacuberta (1997) *A connectionist approach to machine translation*. In *Procs. of the EuroSpeech'97*, volume 1, pag. 91-94

37

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti e traduzione automatica

(... fine)

Reti Neurali

■ In conclusione

- Le reti neurali possono quindi implementare un **approccio Example-Based**
- Esistono tentativi di usare sia **MLP** che **SRN** per la MT
- I risultati sembrano favorire le MLP piuttosto che le SRN (intuitivamente quando il transfer è meno diretto avere tutta la frase avvantaggia la rete!)

	Neural translator	Pruning	SAR	WAR
MLP	GA with just one MLP	None	85.1%	98.1%
	GA with one MLP per output state	None	90.3%	98.7%
	GA with one MLP per output state	Input layers	94.0%	99.2%
SRN	ReConTra	None	53.1%	93.3%

SAR (Sentence Accuracy Rate), WAR (Word Accuracy Rate)

38

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Apprendimento di proprietà ricorsive

Reti Neurali

- Modello connessionista della **performance** umana per il processing di strutture linguistiche ricorsive (Christiansen & Charter 1999)
- Ben diverso dal concetto di **competence**: insieme di regole ricorsive + limitazioni (sulla memoria)
- Un approccio connessionista che usa SRN include implicitamente queste nozioni in un unico modello

39

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Apprendimento di proprietà ricorsive

Reti Neurali

■ Tipologie di ricorsività linguistiche

- **Right-branching** (iterazione più che ricorsività)
Mario vede il gatto che ha mangiato il topo che ha rubato il formaggio....
- **Counting Recursion** ($a^n b^n$)
se è vero che se Mario vince il concorso ... allora si sposa, allora anch'io ci faccio un pensierino
- **Incassamento centrale** (ww^R)
il topo [che il gatto [che Mario ... vede] ha mangiato] ha rubato il formaggio
- **Dipendenze cross-seriali** (ww)
Mario, Giovanna, Giuseppe ... sono rispettivamente promosso, bocciata, promosso ...

40

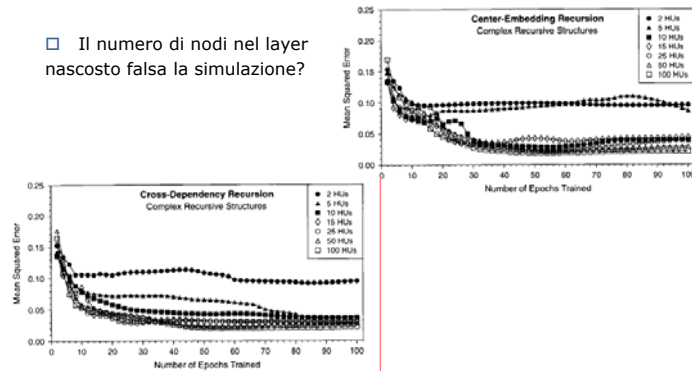
Linguistica Computazionale A.A. 2006-07 – C. Chesì

Apprendimento di proprietà ricorsive

Reti Neurali

- Risultati di apprendibilità con linguaggi ricorsivi (Christiansen & Charter 1999)

Il numero di nodi nel layer nascosto falsa la simulazione?



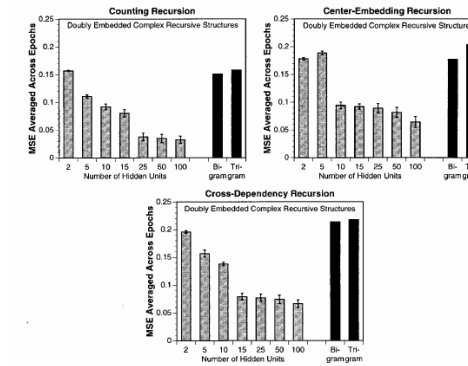
41

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Apprendimento di proprietà ricorsive

Reti Neurali

- Risultati di apprendibilità con linguaggi ricorsivi (Christiansen & Charter 1999)



42

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Reti neurali e innatezza

Reti Neurali

- Tipi diversi di innatezza (Elman 99)

- rappresentazionale**
specificare schemi di connessione e quindi di attivazione (ipotesi: gli stati del cervello provocano il comportamento > stati = pattern di attivazione)
- architetturale**
 - unit-based (tipologia dei neuroni, funzione di attivazione...)
 - vincoli sull'architettura locale (densità, circuiti locali...)
 - vincoli sull'architettura globale (connettività tra aree distanti...)
- timing** (cronotopia innata)
modificazione dei fattori strutturali della rete dovuta ad agenti esogeni (modifica plasticità neurale, possibilità di processare input diversi...)

43

Linguistica Computazionale A.A. 2006-07 – C. Chesì

Prossima lezione

(Domani, Giovedì 13 Aprile, ore 16:30-18:30, Aula G, Via Roma 47)

- Laboratorio

- **T-learn**

- costruzione di una rete neurale
- codifica dell'input
- training set e apprendimento
- analisi dell'output
- Principal Component Analysis e Cluster Analysis

44

Linguistica Computazionale A.A. 2006-07 – C. Chesì