

RETI NEURALI PER L'APPRENDIMENTO DEI TRATTI DELLA PERSONALITÀ DAL LINGUAGGIO NATURALE

Relatore: Prof. Stella Fabio Antonio

Co-relatore: Dott. Marelli Marco

Relazione della prova finale di:

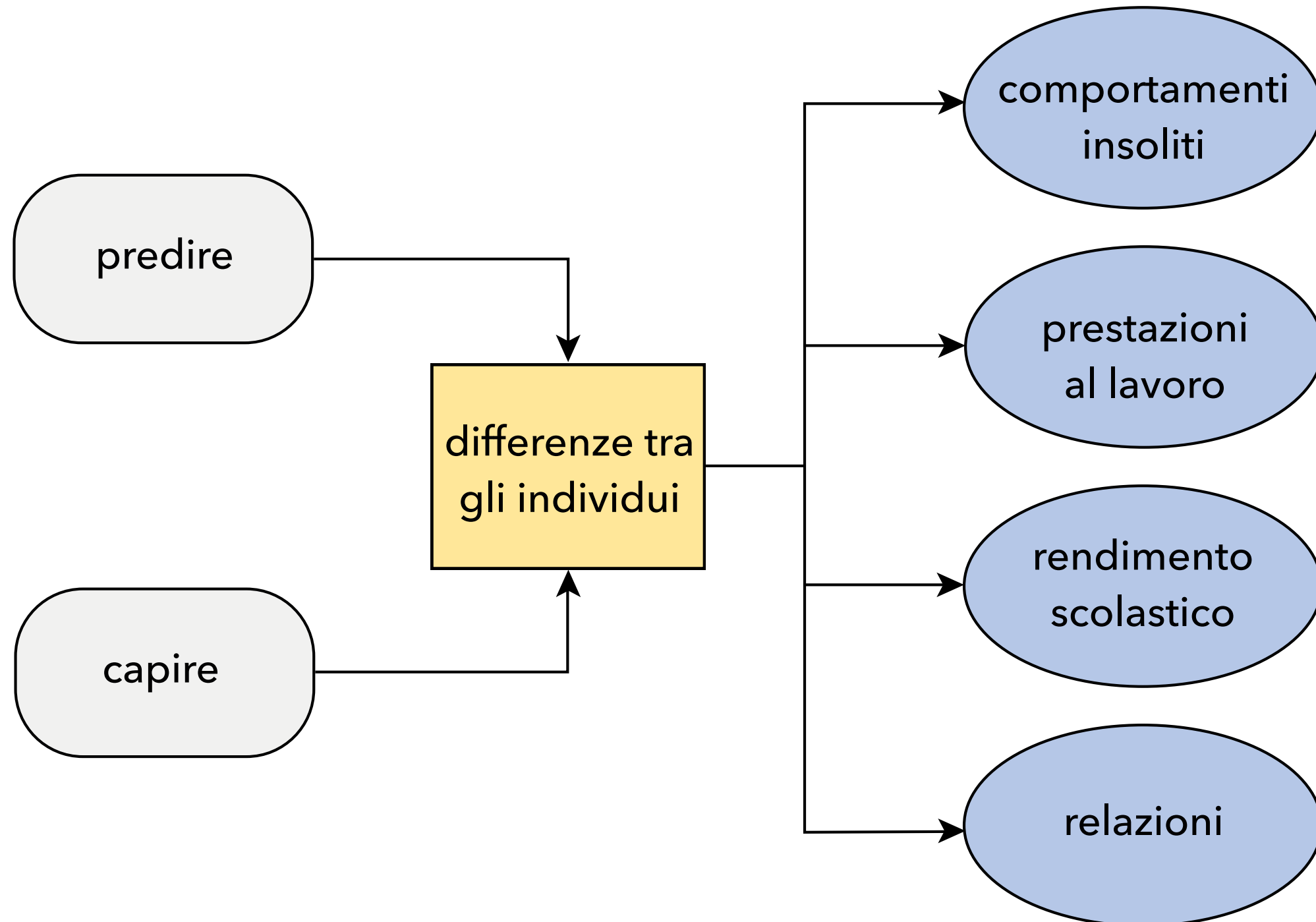
Giorgia Adorni

Matricola 806787

Anno Accademico 2017-2018


Introduzione al problema

Perché la personalità?




Introduzione al problema

- Mappatura diretta tra una review e il vettore di personalità OCEAN [1]
- Ogni tratto di personalità viene calcolato come la media del valore degli aggettivi presenti nel testo



Alan Turing

4 ★ ★ ★ ★ ☆



Ada Lovelace

3.5 ★ ★ ★ ☆ ☆

” VERY GOOD:

This is my first Indian meal in London and the quality of food was authentic and great to taste! They are super fast, efficient and courteous. I recommend this restaurant.

[1] Paul T. Costa e Robert R. McCrae. «The revised neo personality inventory (neo-pi-r)». In: *The SAGE handbook of personality theory and assessment* 2.2 (2008).

Adjective	O	C	E	A	N
...
Courteous	0,000	0,230	-0,070	0,530	0,020
Cowardly	0,037	-0,154	-0,335	0,176	0,268
Creative	0,492	0,019	0,135	0,073	-0,115
Critical	0,170	0,060	-0,010	-0,320	0,310
Cruel	0,000	-0,040	0,000	-0,400	-0,060
Cute	-0,030	0,110	0,450	0,020	0,180
Cynical	0,165	-0,088	-0,096	-0,401	0,078
Dangerous	-0,050	-0,110	0,220	-0,430	0,110
	0,140	-0,150	0,030	-0,000	0,070
	0,150	0,430	0,180	-0,030	-0,220
	-0,050	-0,040	-0,170	-0,170	0,410
	-0,094	-0,156	0,292	-0,017	0,538
	-0,183	0,388	0,240	-0,153	-0,154
	0,220	0,140	0,330	-0,070	-0,130
	-0,040	-0,160	-0,240	-0,450	-0,040
	-0,037	-0,180	0,031	-0,316	0,058
	-0,190	0,050	-0,280	0,240	-0,040
	-0,110	0,060	-0,040	-0,100	0,010
Domineering	0,051	0,170	0,450	-0,333	0,028
Emotional	-0,320	-0,140	0,390	0,190	-0,010
Formal	0,010	0,390	-0,130	0,140	-0,140
Effective	0,040	0,200	0,460	-0,150	-0,090
Efficient	-0,023	0,463	-0,249	0,010	0,062
Egocentric	0,120	-0,120	0,040	-0,340	0,130
...

Introduzione al problema



Preparazione e preprocessing

- **Suddivisione di ciascuna review in frasi**
- **Segmentazione di parole**
- **Stemming:** ogni parola viene ridotta alla sua forma radice (es. “argued”, “arguing” sono mappati ad “argu”)
- **Eliminazione delle stop-word** ovvero le parole considerate senza uno specifico significato semantico.
- **Costruzione di un dizionario** del corpus di training
- **Codifica** di ogni parola del dizionario con un valore intero univoco

the girl who took our order was friendly

334

105

45

-1

girl

took

order

friendly

Approcci e metodologie

Esperimento 1

- **Input Features:** i dati testuali vengono rappresentati come una *bag-of-words* [2].

food is well prepared

1	1	1	0	0	1	0
is	food	well	go	box	prepared	net

- **Architettura della rete:** viene utilizzata una rete *feed-forward* [3] densa o *fully-connected* [4].

[2] Hanna M Wallach. «Topic modeling: beyond bag-of-words». In: *Proceedings of the 23rd ICML*. ACM. 2006.

[3] Daniel Svozil, Vladimir Kvasnicka e Jiri Pospichal. «Introduction to multi-layer feed-forward neural networks». In: *Chemometrics and intelligent laboratory systems* 39.1 (1997).

[4] Tara N Sainath et al. «Convolutional, long short-term memory, fully connected deep neural networks». In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2015 IEEE IC on. IEEE. 2015.

Approcci e metodologie

- Come funzione di loss viene utilizzato l'errore quadratico medio, in inglese *mean squared error* (MSE).

Prestazioni esperimento 1

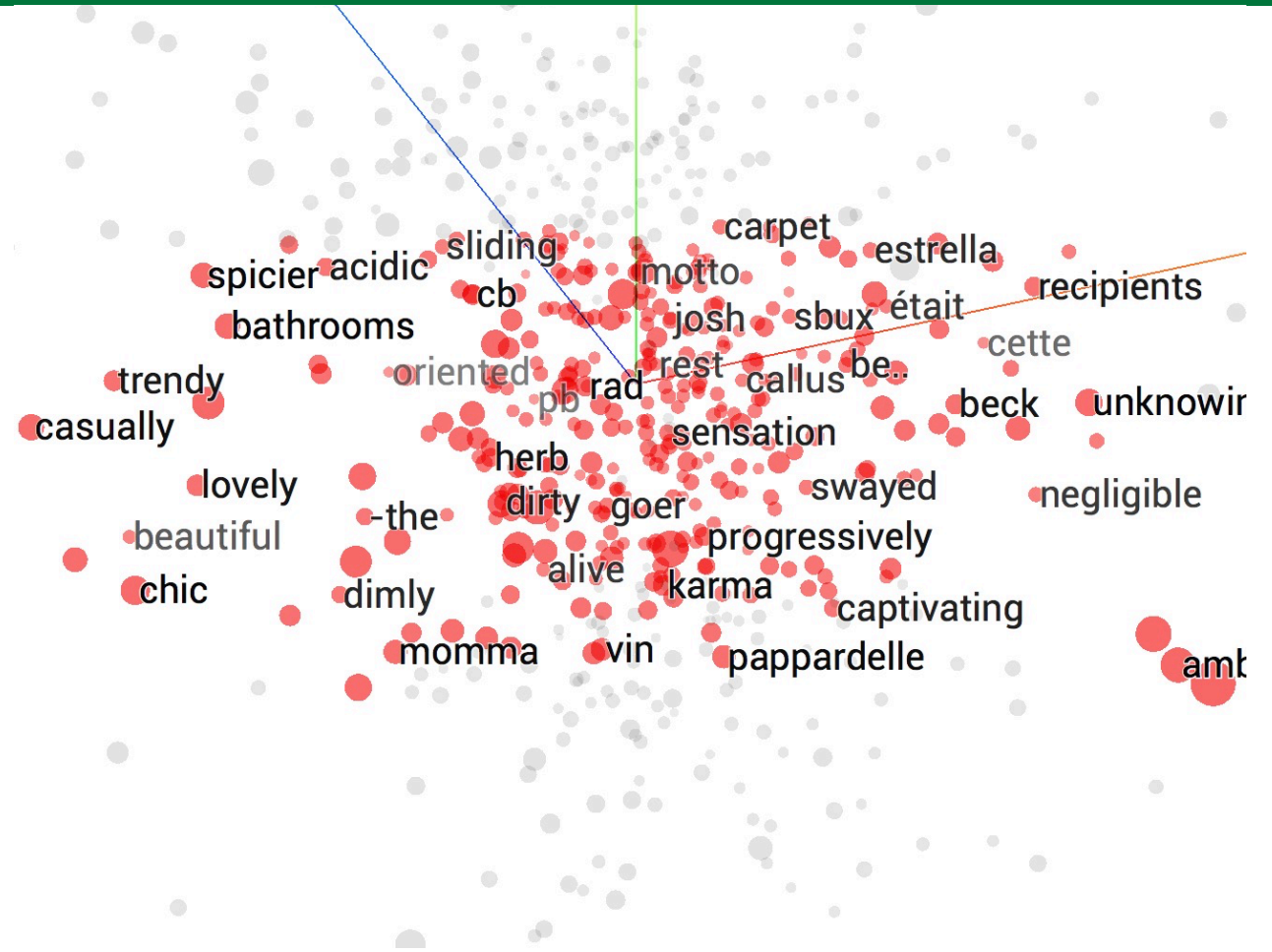
	Train loss	Test loss	Tempo di training
Modello 1	0,061	0,062	235 min
Modello 2	0,090	0,061	250 min
Modello 3	0,068	0,062	265 min

Modelli		Root Mean Squared Error				
		O	C	E	A	N
Modello 1	Modello	0,148	0,227	0,224	0,251	0,351
	Modello 0	0,145	0,224	0,213	0,218	0,318
Modello 2	Modello	0,147	0,226	0,225	0,251	0,341
	Modello 0	0,141	0,227	0,213	0,208	0,305
Modello 3	Modello	0,147	0,226	0,225	0,262	0,348
	Modello 0	0,233	0,307	0,262	0,373	0,546

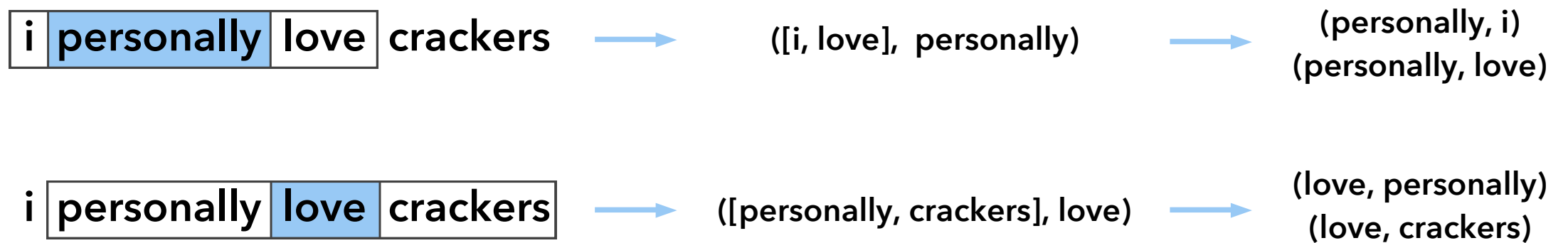
Approcci e metodologie

Esperimento 2

- **Input Features:** viene costruito un embedding di tutte parole tramite il modello *skip*-gram, una versione dell'algoritmo word2vec di Mikolov [5].



i personally love crackers



Approcci e metodologie

Esperimento 2

- **Architettura della rete:** viene utilizzata una rete *convoluzionale* [6]. Come funzione di loss viene utilizzato il *mean squared error* (MSE).

Prestazioni

Modelli		Train loss	Test loss	Tempo di training
Embedding 1	Modello 4	0,061	0,058	200 min
	Modello 5	0,052	0,060	310 min
	Modello 6	0,042	0,060	540 min
Embedding 2	Modello 7	0,038	0,057	225 min
	Modello 8	0,058	0,117	250 min

[6] Yoon Kim. «Convolutional neural networks for sentence classification». In: *arXiv preprint arXiv:1408.5882* (2014).

Esperimento 3

- **Input Features:** si ricorre al modello *skip-gram*, ma definendo l'embedding solo per le coppie degli aggettivi contenuti nel dizionario OCEAN e i loro contesti.
- **Architettura della rete:** viene utilizzata una rete *convoluzionale*. Come funzione di loss viene utilizzato il *mean squared error* (MSE).

Prestazioni

	Train loss	Test loss	Tempo di training
Modello 9	0,043	0,060	18 h
Modello 10	0,050	0,059	17 h

Approcci e metodologie

Esperimento 4

- **Input Features:** si riutilizzano i metodi precedenti più efficienti.
- **Output:** il problema viene trasformato in un compito di classificazione binaria multi-label.
- **Architettura della rete:** viene utilizzata una rete *convoluzionale*.

Prestazioni

Modelli		Train/Test Accuracy [%]				
		O	C	E	A	N
Embedding 2	Modello 11	61/61	60/59	63/60	58/56	57/54
	Modello 12	62/59	61/50	64/45	61/56	61/56
	Modello 13	63/65	63/60	63/61	63/58	62/59
Embedding 3	Modello 14	61/61	61/57	62/52	60/57	60/56
	Modello 15	63/60	62/48	64/60	62/49	62/48

Conclusioni e sviluppi futuri

Conclusioni

- Il problema è estremamente complesso, ad oggi non è stato affrontato in modo approfondito
- La scelta di rappresentazione del testo è fondamentale: che domanda poniamo al modello?
- La classificazione sembra mostrare risultati lievemente migliori rispetto alla regressione

Sviluppi

- Rappresentazioni alternative del testo ricorrendo ad altri strumenti di NLP, annotazioni, part-of-speech [7]
- Utilizzo di altre architetture di apprendimento: reti ricorrenti [8]

[7] Roger W Brown. «Linguistic determinism and the part of speech.» In: *The Journal of Abnormal and Social Psychology* 55.1 (1957)

[8] Tomas Mikolov et al. «Recurrent neural network based language model. »In: *Eleventh Annual Conference of the International Speech Communication Association*. ISO 690 (2010).