Strutture basate su regole e strutture basate su approssimazioni

Antonio Lieto

University of Salerno, IT Cognition Interaction and Intelligent Technologies Laboratory (CIIT Lab) ICAR-CNR, Palermo, IT

> Web: <u>https://www.antoniolieto.net</u> Lab: <u>https://www.ciitlab.org</u>





Cognition Interaction and Intelligent Technologies Lab

HOME MEMBRI RICERCA

PUBBLICAZIONI

CIIT LAB

TIROCINIO E PROGETTI DI TESI

CONTATTI

NEWS

Il Laboratorio CIIT (Cognition, Interaction and Intelligent Technologies Lab) st attività di ricerca su tematiche relative a metodologie, tecniche e tecnologie dell'Intelligenza Artificiale, dell'Interazione Umano-Macchina e delle Scienze Cogn Computazionali.

Il Laboratorio riunisce studiosi di differenti ambiti disciplinari appartenenti all'Ate salernitano e ad Università europee, americane ed asiatiche e si avvale della consolidata collaborazione con: il laboratorio di "Sistemi Cognitivi per la Robotica dell'ICAR- CNR di Palermo, il "Center for Logic, Language and Cognition" dell'Università degli Studi di Torino, l'Urban/Eco Research Center, il Content Cen Computing Group del Dipartimento di Informatica dell'Università di Torino, l'ISIS del Dipartimento di Informatica dell'Università di Salerno, il PRISCA Lab (DIETI, Università Degli Studi di Napoli Federico II).

9 A-Z Contatta CIIT Lab Per info e approfondimenti RICERCA TIROCINI E PROGETTI DI PUBBLICAZIONI TESI

https://www.ciitlab.org



Attività di ricerca

MODELLI COMPUTAZIONALI DELLA NARRAZIONE E STORYTELLING INTERATTIVO: Analisi e sviluppo di sistemi e modelli computazionali della narrazione in grado di abilitare, in agenti intelligenti (incarnati, es. robot, o meno), la capacità di interpretare, analizzare e generare storie e diversi livelli di astrazione.

MODELLI COMPUTAZIONALI PER LA COMPRENSIONE E GENERAZIONE DI METAFORE E ANALOGIE: Analisi e sviluppo di sistemi di Intelligenza Artificiale in grado riconoscere (nei testi) e generare automaticamente nuove analogie e metafore tramite l'utilizzo di sistemi neuro-simbolici, LLM e sistemi di IA simbolica.

SISTEMI SEMANTICI E DI KNOWLEDGE MANAGEMENT: Analisi e sviluppo di sistemi semantici di rappresentazione della conoscenza e ragionamento di automatico basati su tecniche di logica computazionale e intelligenza artificiale simbolica, intelligenza artificiale neuro-simbolica e modelli neurali.

ARCHITETTURE COGNITIVE: Analisi e sviluppo di architetture cognitive (software biologicamente o cognitivamente ispirati) in grado di permettere ad un agente artificiale (inclusi robot) di effettuare più compiti intelligenti.

MODELLI COMPUTAZIONALI DELLA COGNIZIONE: Analisi e sviluppo di modelli cognitivi per compiti specifici (es. inferenze, classificazione etc.) o multi-tasking.

TECNOLOGIE PERSUASIVE: Analisi e sviluppo di attività di ricerca su interfacce e tecnologie persuasive.

MENTE-MACCHINA: Analisi e studio degli aspetti epistemologici riguardanti il ruolo esplicativo di sistemi artificiali per l'analisi del rapporto mente-macchina.

TECNOLOGIE DEL LINGUAGGIO: Analisi e sviluppo di attività di ricerca su tecnologie del linguaggio con applicazioni in diversi ambiti: dal settore health a quello della robotica sociale.

Questo intervento

Spunti e collegamenti tra alcuni aspetti legati alla teoria **semantica** della categorizzazione e il modo in cui i **Large Language Models** (LLMs) gestiscono aspetti **sintattici** e **semantici** del linguaggio.

Confronto basato sulla dicotomia:

strutture (rigide/"universali") basate su regole

strutture ("graduali"/"contestuali") basate su approssimazioni

Tipicalità



Approcci computazionali sul tema (overview parziale)



Es. Rag. Senso Comune - Categorizzazione

X {haCoda, Scodinzola, Abbaia}



Cosa è X ????

Es. Rag. Senso Comune - Categorizzazione

Un elemento X è categorizzato come CANE perché: X {haCoda, Scodinzola, Abbaia}





Nessuno di questi tratti è definitorio di cane

Concetti/Rappresentazioni tipiche

Teoria classica dei concetti



TRIANGOLO= Poligono regolare con 3 lati e 3 angoli

Struttura basata su regola

Wittgenstein: Somiglianze di famiglia



Nessun volto condivide tutte le caratteristiche di un altro volto, ma ciascuno possiede alcuni tratti di altri.

PROTOTYPE THEORY

• Teoria dei concetti dominante in psicologia dagli anni'70 sviluppata da Eleanor Rosch.

 → categorie di senso comune non definibili sulla base di condizioni necessarie e sufficienti

Typicality ratings

TYPICALITY RATINGS FOR INSTANCE-CATEGORY PAIRS

Bird	s	Mammals				
Instance	Rating	Instance	Rating			
Robin	3.00	Deer	2.83			
Sparrow	3.00	Horse	2.76			
Bluejay	2.92	Goat	2.75			
Parakeet	2.83	Cat	2.67			
Pigeon	2.83	Dog	2.67			
Eagle	2.75	Lion	2.67			
Cardinal	2.67	Cow	2.58			
Hawk	2.67	Bear	2.58			
Parrot	2.58	Rabbit	2.58			
Chicken	2.00	Sheep	2.58			
Duck	2.00	Mouse	2.25			
Goose	2.00	Pig	2.17			

Smith, Shoben, & Rips (1974, p. 218)

E'una "sedia"?



E' un "cane"?



E' un "gatto"?



'Fuzzy' or 'graded' categorization

 Le condizioni necessarie e sufficienti dovrebbero valere per tutti gli individui di una categoria, ma questo non accade:

- Strutture concettuali graduali

Graded Structure

 Item tipici simili al prototipo in base alla distanza da esso



Exemplars and Exemplar-based Reasoning



Teorie in Conflitto?

 Prototypes, Exemplars and other conceptual representations (for the same concept) can co-exists and be activated in different contexts (Malt 1989).

Heterogeneous Proxytypes Hypothesis

Prototypes, Exemplars, Theories and other conceptual **representations** can co-exists and be activated in different contexts



Different representational structures have different **accessing procedures** (reasoning) to their content.

(Lieto, BICA 2014) https://doi.org/10.1016/j.procs.2014.11.078

DUAL PECCS: DUAL- Prototype and Exemplars Conceptual Categorization System

Lieto, Radicioni, Rho (IJCAI 2015, JETAI 2017)

2 Cognitive Assumptions



22

1) Multiple representations for the same concept

2) On such diverse, but connected, representation are executed different types of reasoning (**System 1/ System 2**) to integrate.

Type 1 Processes	Type 2 Processes
Automatic	Controllable
Parallel, Fast	Sequential, Slow
Pragmatic/contextualized	Logical/Abstract

Heterogeneous Proxytypes in DUAL-PECCS



Lieto et al. (2015), A Common-Sense Conceptual Categorization System Integrating Heterogeneous Proxytypes and the Dual Process of Reasoning, IJCAI, AAAI Press.

Lieto, A., Radicioni, D., Rho, V, (2017). Dual PECCS: a cognitive system for conceptual representation and categorization, JETAI, 29 (2), 433-452, Taylor and Francis..

23

Co-referring representational Structures via Wordnet



Lieto, A., Mensa, E., Radicioni, D., 2016. A resource-driven approach for anchoring linguistic resources to conceptual spaces. In Conference of the Italian Association for Artificial Intelligence (pp. 435-449). Springer, Cham.

Overview



Evaluation Accuracy Metrics (only with Prototypes and Exemplars)

- Two evaluation metrics have been devised:
 - Concept Categorization Accuracy: estimating how often the correct concept has been retrieved;
 - Proxyfication Accuracy: how often the correct concept has been retrieved AND the expected representation has been retrieved, as well.

test	CC-ACC	P-ACC
with no IE	89.3% (100/112)	79.0% (79/100)
with IE	77.7% (87/112)	71.3% (62/87)

Lieto, A., Radicioni, D., Rho, V, (2017). Dual PECCS: a cognitive system for conceptual representation and categorization, JETAI, 29 (2), 433-452, Taylor and Francis..

Accuratezza

Model	CC-Acc	P-Acc
Dual-PECCS (no IE)	89.3%	79.0%
ChatGPT-40	82.1%	81.5%
ChatGPT-4	80.3%	75.5%
Dual-PECCS	77.7%	71.3%
(with IE)		
ChatGPT-3.5	70.5%	76.0%
Microsoft Copilot	65.0%	74.0%
DALL·E 3	53.3%	85.8%

Table 9: Overall accuracy rates

lavoro con Isabella Cossidente

In che modo tali teorie semantiche hanno a che fare il modo in cui i Large Language Models (LLMs) gestiscono aspetti sintattici e semantici del linguaggio?

Regola accordo

- a. The big black dogs run fast.
- b. The big black dogs that chase the cats run fast.
- c. *The big black dogs runs fast.
- d. *The big black dogs that chase the cats runs fast.
- a. Colourless green ideas sleep furiously.
- b. Colourless green ideas that cook the door sleep furiously.
- c. *Colourless green ideas sleeps furiously.
- d. *Colourless green ideas that cook the door sleeps furiously.

Il punto: se imparata, questa regola sintattica, vale indipendentemente dalla semantica delle frasi

LLM imparano la regola?

Structure description	Example
Simple agreement	The boy laughs / laugh*
In a sentential complement	The boy knows the girls play/plays*
Across a prepositional phrase	The plate near the glasses breaks/break*
Across a subject relative clause	The cat that chases the mice runs/run*
In a short verb phrase coordination	The boy smiles and laughs/laugh*
Across an object relative clause	The mouse that the <u>cats</u> chase runs/run*
Within an object relative clause	The mouse that the cats chase/chases* runs
Across an object relative clause (no that)	The mouse the <u>cats</u> chase runs/run*
Within an object relative clause (no that)	The mouse the cats chase/chases* runs

Marvin & Linzen, 2018

Targeted syntactic evaluation of language models. EMNLP 2018.

Sembra di si

	BERT	BERT	LSTM	Humans	# Pairs
	Base	Large	(M&L)	(M&L)	(# M&L Pairs)
SUBJECT-VERB AGREEMENT:		_			
Simple	1.00	1.00	0.94	0.96	120 (140)
In a sentential complement	0.83	0.86	0.99	0.93	1440 (1680)
Short VP coordination	0.89	0.86	0.90	0.82	720 (840)
Long VP coordination	0.98	0.97	0.61	0.82	400 (400)
Across a prepositional phrase	0.85	0.85	0.57	0.85	19440 (22400)
Across a subject relative clause	0.84	0.85	0.56	0.88	9600 (11200)
Across an object relative clause	0.89	0.85	0.50	0.85	19680 (22400)
Across an object relative (no that)	0.86	0.81	0.52	0.82	19680 (22400)
In an object relative clause	0.95	0.99	0.84	0.78	15960 (22400)
In an object relative (no that)	0.79	0.82	0.71	0.79	15960 (22400)

Goldberg, 2019

Assessing BERT's syntactic abilities. *arXiv preprint arXiv:1901.05287*.

ma...

Does BERT *really* agree ? Fine-grained Analysis of Lexical Dependence on a Syntactic Task

Karim Lasri^{a,β} Alessandro Lenci^β Thierry Poibeau^a ^aLattice (École Normale Supérieure-PSL, CNRS, U. Sorbonne Nouvelle) ^βUniversity of Pisa karim.lasri@ens.psl.eu

alessandro.lenci@unipi.it thierry.poibeau@ens.psl.eu

Abstract

Although transformer-based Neural Language Models demonstrate impressive performance on a variety of tasks, their generalization abilities are not well understood. They have been shown to perform strongly on subject-verb number agreement in a wide array of settings, suggesting that they learned to track syntactic dependencies during their training even without explicit supervision. In this paper, we examine the extent to which BERT is able to perform lexically-independent subject-verb number agreement (NA) on targeted syntactic templates. To do so, we disrupt the lexical patterns found in naturally occurring stimuli for each targeted structure in a novel fine-grained analysis of BERT's behavior. Our results on nonce sentences suggest that the model generalizes well for simple templates, but fails to perform lexically-independent syntactic generalization when as little as one attractor is present.

the noun belongs to an embedded relative clause. These NA properties have made it one of the preferred test beds to investigate the ability of neural language models (NLMs) to learn abstract, hierarchical syntactic structures (Linzen et al., 2016; Gulordava et al., 2018; Marvin and Linzen, 2018; Goldberg, 2019; Bacon and Regier, 2019; Lakretz et al., 2019). Although recurrent and transformerbased NLMs have been shown to possess syntactic abilities on the task, their nature is not fully understood (Baroni, 2019).

Can NLMs really perform lexically-independent number agreement, regardless of the syntactic structure? To answer this question, we test BERT (Devlin et al., 2019) against the NA task while controlling both the syntactic constructions and the meaningfulness of the stimuli presented to the model.

Our experiments provide two main findings. Contrarily to previous observations that BERT per-

ma...

- LLM non imparano davvero la regola...
- Ad es. per frasi tipo "colourless green ideas..." non riescono a generalizzare questa struttura sintattica

Does BERT *really* agree ? Fine-grained Analysis of Lexical Dependence on a Syntactic Task

Karim Lasri^{∂,β} **Alessandro Lenci**^β **Thierry Poibeau**[∂] [∂]Lattice (École Normale Supérieure-PSL, CNRS, U. Sorbonne Nouvelle) ^βUniversity of Pisa

karim.lasri@ens.psl.eu alessandro.lenci@unipi.it thierry.poibeau@ens.psl.eu

Abstract

Although transformer-based Neural Language Models demonstrate impressive performance on a variety of tasks, their generalization abilities are not well understood. They have been shown to perform strongly on subject-verb number agreement in a wide array of settings the noun belongs to an embedded relative clause. These NA properties have made it one of the preferred test beds to investigate the ability of neural language models (NLMs) to learn abstract, hierarchical syntactic structures (Linzen et al., 2016; Gulordava et al., 2018; Marvin and Linzen, 2018; Goldbarg, 2010; Pacen and Pacier, 2010; Lebratz

2022

Errori separazione "sintagmatica"



Critiche

- Questo risultato mostra che quando ci sono frasi sintatticamente ben formate (ma semanticamente no), questi sistemi non generano l'accordo
- Apprendimento non generale => struttura approssimata della sua generalizzazione
- => critiche (es. Berent & Marcus, 2019)

Berent, I., & Marcus, G. (2019). No integration without structured representations: Response to Pater. *Language*, *95*(1), e75-e86.

Tuttavia...

MIT Press Direct 🗸		Search				Neurob	iology of Lar	iguage ~	Q	Adva Searc
Neurobiology of Language					Issues	Online Early	About 🗸	Submit 🗸	,	
Volume 1, Issue 1 2020	March 01 2020	s the Core	• Driver of	the Langua	age-se	elective Ne	twork a			
	Francis Mollica 🖾 , Ma Peng Qian, Evelina Feo Check for upda > Author and Article Neurobiology of Langu https://doi.org/10.1162/	tthew Siegelman lorenko ⊠ tes Information age (2020) 1 (1): 1 nol_a_00005	n, Evgeniia Diach 104–134. Article history nissions	œ 오 오 오 오 오 오 오 오 오 오 오 오 오 오 오 오 오 오 오	O Vie	Zachary Mineroff	, Richard Futrel	II, Hope Kean	,	

risonanza magnetica con frasi che presentano inversioni crescenti nell'ordine delle parole (=> crescente incoerenza sintattica ma con parole che presentano una coesione lessicale...ci sono parole che possono essere combinate tra loro)



Frasi sintatticamente scorrette valutate con struttura graduale di plausibilità semantica



Ma per il cervello non conta...



Frasi perfettamente grammaticali o meno trattate ugualmente

Unica differenza per liste di parole o con sequenze non coese

- Semantica autonoma dalla "sintassi"
- La composizione semantica non passa necessariamente da quella sintattica
- Prospettiva invertita rispetto all'impianto chomskiano...
- "We think of semantics as an independent computational system that obeys its own rules for how words are bound together during language comprehension. Of course, many of these rules have correlates in syntax, but nevertheless we conceive of semantic composition as a process that can take place independently from syntactic structure building" (Mollica et al. 2020).

 Aspetti legati alla capacità di generalizzare e apprendere strutture approssimate (es. basate su esempi e su prototipi con gradualità) gioca un ruolo nei processi di comprensione linguistica

https://www.youtube.com/clip/ UgkxBW1Z9fU8bRShFk_M0CTv8d3_N_LB0JuR



THE NOBEL PRIZE IN PHYSICS 2024

Interview with **Geoffrey E. Hinton**

First Reactions | Geoffrey Hinton, Nobel Prize in Physics 2024 | Telephone interview

Δ 2.1K

ም









HE

Tuttavia

- Questi sistemi ancora non sono in grado di generalizzare out of the distribution (mentre questa è una capacità che noi abbiamo)
- Alcune critiche chomskiane (es. povertà dello stimolo) rimangono (es. dati Navigli su Minerva 7B):
- dataset di 1,5 trilioni di parole
- instruction tuning: 580mila esempi

Per concludere

Nessun modello esistente è in grado di dare conto della varietà dei fenomeno linguistici (alcuni modelli catturano meglio certe proprietà ma non tutte)

Bisogno di integrare questo tipo di approcci

"Allo stato attuale delle cose, la molteplicità dei punti di vista, dei modelli e degli approcci sperimentali sembra essere il "marchio di fabbrica" delle scienze dell'artificiale. Questo fatto può lasciarci sgomenti o entusiasti ma, di certo, non indifferenti" (Cordeschi, 2002, The Discovery of the Artificial, Springer).