



IL SENSO DI INTERAGIRE CON LA MACCHINA IN LINGUAGGIO NATURALE

FEDERICO MARTELLI – ROBERTO NAVIGLI

Permettere a una macchina di interagire con l'essere umano in linguaggio naturale è una delle più grandi sfide che il settore dell'Intelligenza Artificiale si trova ad affrontare. Il linguaggio è una delle abilità più complesse di cui l'essere umano dispone e, di conseguenza, una delle funzionalità più difficili da riprodurre artificialmente. L'articolo si propone di rispondere, in maniera non esaustiva, al quesito: come può una macchina interagire efficacemente con un essere umano in linguaggio naturale?

Non

è un caso che Alan Turing nel 1950, agli albori di una nuova branca del sapere che prenderà presto il nome di Intelligenza Artificiale (AI), scelse il linguaggio naturale come canale preferenziale mediante il quale stabilire se una macchina possa essere definita intelligente o meno. In effetti, il linguaggio naturale, ovvero la facoltà di comunicare mediante segni convenzionali, è qualcosa che caratterizza l'entità intelligente per antonomasia: l'essere umano. Dunque, se è vero che l'obiettivo della AI è quello di creare macchine intelligenti, allora il linguaggio naturale, quale espressione massima dell'intelletto, deve inevitabilmente rivestire un ruolo di primaria importanza. Inoltre, la necessità di prestare particolare attenzione al linguaggio naturale deriva anche dalla crescente presenza delle macchine nella nostra vita. Infatti, nell'era dei grandi sviluppi tecnologici e dell'inarrestabile automazione, l'essere umano è sempre più esposto al contatto con la macchina che diviene ormai parte

imprescindibile del suo lavoro e della sua quotidianità. Dunque, consentire un'interazione con la macchina basata sul linguaggio naturale è uno dei più ambiziosi obiettivi della elaborazione del linguaggio naturale (*Natural Language Processing*, Nlp), un settore della AI che da anni studia approcci sempre più avanzati di comunicazione tra utente e macchina. A tale scopo, è indispensabile che questa sia in grado di cogliere l'intento comunicativo dell'interlocutore, alla base del quale vi è il significato del messaggio che si vuole trasmettere. Ne deriva la fondamentale importanza di permettere alla macchina di identificare il significato di una specifica interazione, come quella tra utente e dispositivi altamente tecnologici quali smartphone, assistenti intelligenti o altri sistemi di domotica. Come vedremo, consentire l'identificazione del significato di un messaggio costituisce un punto di svolta verso la realizzazione di entità sempre più capaci di interloquire in maniera via via più complessa e articolata. Ma che cos'è, esattamente, il «significato»? Come può una macchina cogliere il significato delle parole e degli enunciati in linguaggio naturale? Nonostante, almeno intuitivamente, sappiamo a cosa ci riferiamo quando parliamo di «significato», darne una definizione esaustiva è tutt'altro che impresa semplice. Ancora più importante è il fatto che da tale definizione dipenda l'implementazione di tecniche computazionali per la comprensione del linguaggio e, pertanto, vale la pena spendere qualche parola a riguardo.

La Srl, branca della linguistica che si occupa di studiare la lingua dal punto di vista del significato, fornisce molteplici definizioni di tale concetto, le quali variano a seconda di come quest'ultimo viene percepito. Secondo l'*approccio cognitivista* il significato coincide con il concetto a cui associamo una determinata espressione linguistica. L'*approccio referenzialista*, invece, afferma che il significato deriva dalla relazione che intercorre tra referente, ovvero elemento extralinguistico, ed espressione linguistica. Infine, l'*approccio strutturalista* sostiene che il significato viene a determinarsi nel momento in cui la lingua plasma il pensiero che non è strutturato. Negli ultimi anni, la semantica ha via via rivestito maggiore importanza nell'ambito del Nlp e in altri settori affini come quello del recupero delle informazioni (*Information Retrieval*, IR). Basti pensare alla recente introduzione di tecniche di *question answering* all'interno di tradizionali motori di ricerca, che sono ora in grado di rispondere automaticamente a determinati tipi di domande. Ad esempio, inserendo nel campo di ricerca la domanda: «Chi è l'at-

tuale Presidente degli Stati Uniti d'America?», una simile applicazione sarebbe in grado di restituire come risultato la risposta corretta, ovvero «Donald Trump», piuttosto che proporre all'utente tutte le pagine contenenti le singole parole che occorrono nell'interrogazione. Tuttavia, tali sistemi non sono in grado di estrarre ed elaborare informazioni in modo intelligente: sebbene in questo senso siano stati fatti notevoli passi avanti, è evidente che vi sia ancora ampio margine di miglioramento e che larga parte delle applicazioni che richiedono un'interazione in linguaggio naturale con l'utente non siano in grado di evolvere efficacemente dal livello meramente lessicale a quello semantico. I sistemi più diffusi di Nlp e IR tendono a trattare il testo come se fosse una mera sequenza di simboli, limitandosi a una elaborazione superficiale del suo significato. Ci soffermeremo ora sulle aree di ricerca del Nlp, che mirano a sopperire a queste mancanze investigando tecniche e metodi attraverso i quali è possibile consentire a una macchina di capire in profondità il significato di ciò che diciamo.

VERSO LA COMPrensIONE DEL SIGNIFICATO

Fantastichiamo di trovarci, in un futuro non così lontano, a dover interagire in linguaggio naturale con dei robot. Immaginiamo ora di assistere alla scena in cui un agente di polizia domanda a un suo collega robot: «Mi copri le spalle?» Come fa il robot a comprendere il significato della richiesta postagli? Un essere umano non avrebbe problemi a capire che l'agente non stia chiedendo che gli vengano coperte le spalle con un capo di abbigliamento, bensì che venga protetto dall'eventuale attacco di un nemico.

Al fine di consentire la comprensione del linguaggio, di risolverne le ambiguità e permettere alla macchina di cogliere l'intento comunicativo del parlante, il Nlp ha sviluppato delle sottobranchie interamente dedicate alla risoluzione di tali problematiche.

In questo elaborato saranno illustrate tre delle principali attività computazionali mirate alla comprensione del testo scritto (significando che il parlato non verrà qui trattato): *Word Sense Disambiguation* (Wsd), *Semantic Role Labeling* (Srl) e *Semantic Parsing*. La Wsd affronta il problema dell'ambiguità lessicale, ovvero il fatto che le parole in linguaggio naturale possano contare più significati. Ad esempio, si consideri la parola «piano». Essa può riferirsi a un concetto geometrico astratto, può essere utilizzata come sino-

nimo di lentamente, può fare riferimento al piano di un palazzo, può indicare uno strumento musicale e così via. In questo caso, si dice che la parola è polisemica, ossia che essa può assumere più significati ed è pertanto ambigua. Dunque, la Wsd consiste nell'individuazione del significato corretto di una parola in un dato contesto. Essa è considerata una delle attività computazionali più complesse del settore dell'elaborazione del linguaggio naturale a causa della difficoltà di cogliere, in un inventario ben definito, i significati che le parole possono assumere nei diversi contesti d'uso. Per questo motivo, le prestazioni dei sistemi di disambiguazione non superano la soglia del 75% di accuratezza. Tuttavia, anche se i sistemi di disambiguazione identificassero correttamente tutte le sfumature semantiche delle singole parole, ciò non garantirebbe la comprensione del significato di intere frasi. Per meglio comprendere questo concetto, si consideri il seguente esempio: «l'agente colpisce il terrorista». Disambiguare le tre parole (agente, colpisce e terrorista) non è sufficiente a comprendere la semantica della frase. Basti pensare alla proposizione «il terrorista colpisce l'agente» che, sebbene dotata della medesima struttura sintattica e contenente le stesse parole, convoglia il significato opposto rispetto alla prima frase. Tale problematica è affrontata dalla Srl, denominata anche *Shallow Semantic Parsing*, che si pone l'obiettivo di associare a parole o espressioni etichette che ne indichino il ruolo semantico all'interno della frase, quali agente, obiettivo e risultato. Ad esempio, nella frase «Il robot copre le spalle all'agente» si possono individuare i seguenti ruoli semantici: robot è l'agente, le spalle è la destinazione e l'agente è il tema. La Srl è un importante passo avanti verso l'elaborazione di una prima strutturazione semantica di un testo scritto in linguaggio naturale. Tuttavia, tale strutturazione interessa quasi esclusivamente verbi e sostantivi, e i rispettivi argomenti sintattico-semantiche. Per ottenere una rappresentazione del significato dell'intera frase, che includa ogni elemento semantico della medesima, è necessaria un'analisi ancora più complessa dei ruoli semantici e delle loro interazioni. Tale analisi è oggetto di studio del Semantic Parsing. Utilizzato per la prima volta da Yorick Wilks nella seconda metà del secolo scorso, il termine fa riferimento all'attività computazionale avente l'obiettivo di elaborare una rappresentazione semantica di una proposizione, includendo i ruoli semantici di tutti i suoi componenti. Al giorno d'oggi sono disponibili rappresentazioni semantiche sofisticate, che organizzano la conoscenza mediante strutture dati, quali grafi




e alberi. Una delle più diffuse è l'*Abstract Meaning Representation* (Amr) che, però, è fortemente legata ad alcuni costrutti della lingua inglese e dunque non è di immediata applicazione a testi scritti in altre lingue. Nel nostro laboratorio di ricerca è in corso di sviluppo una rappresentazione che superi tali limitazioni e sia totalmente indipendente dalla lingua, permettendo di integrare la conoscenza estrapolata da testi scritti in lingue diverse. Per raggiungere il loro scopo, le attività computazionali descritte fanno uso di diverse tecniche di trattamento automatico del linguaggio, che possono essere suddivise in tre approcci fondamentali: l'approccio supervisionato, l'approccio non supervisionato e quello basato sulla conoscenza. Il primo richiede la presenza di grandi quantità di dati annotati che, tuttavia, non sono disponibili nell'ambito della semantica computazionale, soprattutto per le lingue diverse dall'inglese. Al giorno d'oggi, gli approcci supervisionati sono quelli che permettono di ottenere i migliori risultati. L'approccio non supervisionato, al contrario, non richiede alcuna annotazione manuale e mira a individuare pattern ricorrenti in base ai quali inferire le informazioni necessarie quali, ad esempio, i significati di una parola. Non essendo basato su informazioni fornite da annotatori umani, le prestazioni dei metodi non supervisionati sono considerevolmente inferiori. Sia l'approccio supervisionato che quello non supervisionato si incentrano sull'utilizzo di complesse architetture neurali mirate al *Deep Learning* (DL). Infine, l'approccio imperniato sulla conoscenza richiede l'utilizzo di grandi risorse strutturate aventi lo scopo di immagazzinare e organizzare al meglio grandi quantità di dati linguistici. Risorse particolarmente utilizzate sono WordNet¹, lessico computazionale della lingua inglese elaborato da George Miller dal 1985 presso la Princeton University e curato oggi da Christiane Fellbaum, e BabelNet², la rete semantica multilingue più grande al mondo, sviluppata nel nostro laboratorio di ricerca. Uno dei vantaggi fondamentali offerti dall'approccio basato sulla conoscenza consiste nel permettere di elaborare testi indipendentemente dalla lingua in cui sono scritti ossia, come si dice in gergo, di scalare su più lingue. Inoltre, tali approcci affrontano il problema del cosiddetto collo di bottiglia della conoscenza sopperendo alla mancanza di grandi quantità di testi annotati con l'impiego di reti semantiche e tassonomie ad ampia copertura.

1. FELLBAUM 2012.

2. NAVIGLI – PONZETTO 2012.

DARE UN 'SENSO' ALL'INTELLIGENZA ARTIFICIALE

Il denominatore comune di tali approcci è l'implementazione di una rappresentazione computazionale delle parole, degli enunciati e del loro significato. Infatti, affinché una macchina possa capirci, è necessario trasformare il linguaggio naturale nell'unica forma di linguaggio a essa comprensibile: i bit. Ebbene, la maggior parte di tali tecniche è basata su architetture neurali che producono rappresentazioni computazionali che non possono essere interpretate da esseri umani. In questo caso le componenti della codifica computazionale di un determinato elemento linguistico non possono essere ricondotte all'elemento in questione a causa della natura latente di tale rappresentazione che viene appunto definita implicita. Ciò rende tali tecniche opache e difficili da correggere e ottimizzare. Per rendere più trasparente il processo decisionale di un complesso modello neurale, la comunità scientifica sta studiando tecniche mirate alla cosiddetta *explainability* delle architetture di DL. A tale scopo, le rappresentazioni implicite possono essere combinate con rappresentazioni semantiche esplicite derivanti dall'impiego di altre tecniche, tipicamente basate sulla conoscenza, che, invece, producono codifiche computazionali trasparenti, ossia facilmente riconducibili all'elemento linguistico in esame. L'idea sottostante è che la combinazione di rappresentazioni esplicite e implicite possa portare a una maggiore consapevolezza riguardo al funzionamento delle più moderne architetture neurali. In questo modo non solo sarà possibile interpretare e comprendere più a fondo tali processi, ma anche correggere con cognizione di causa le decisioni prese dal modello stesso che, a sua volta, sarà in grado di spiegare. Ad esempio, un autoveicolo autonomo, nell'interazione con un passeggero, potrà giustificare le proprie scelte («sto scegliendo questo percorso perché ho notato una manifestazione in piazza Venezia e mi hai chiesto di arrivare prima possibile») formulandole anche in lingue diverse a partire da un'unica rappresentazione semantica ispezionabile. Oltre che nelle conversazioni quotidiane («Mario, raccontami i tuoi problemi di cuore»), un veicolo autonomo o un pilota automatico saranno in grado di motivare decisioni anche ardue, quali urtare contro un albero piuttosto che finire in un burrone, o ammarare invece di schiantarsi contro una collina, magari argomentandole in un'aula di tribunale 

BIBLIOGRAFIA

- G. BERRUTO, *La semantica*, Zanichelli, Bologna 1976.
 F. CASADEI, *Lessico e semantica*, Carocci, Milano 2003.
 C. FELLBAUM, *WordNet*, «The Encyclopedia of Applied Linguistics» 2012.
 D. GILDEA – D. JURAFSKY, *Automatic labeling of semantic roles*, «Computational linguistics» XXVIII (2002) 3, pp. 245-288.
 R. NAVIGLI, *Word Sense Disambiguation: A Survey*, «Computing Surveys» XLI (2009) 2, Article 10.
 R. NAVIGLI, *Natural Language Understanding: Instructions for (Present and Future) Use*, IJCAI, 2018.
 R. NAVIGLI – S.P. PONZETTO, *BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network*, «Artificial Intelligence» CXCI (2012), pp. 217-250.

