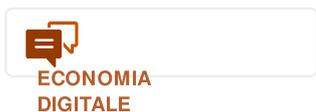


I computer apprendono il maltese

Alcuni ricercatori hanno addestrato un modello linguistico sui dati testuali in maltese e gli hanno insegnato a individuare il sentimento e i nomi di tag in un testo.



© graphicwithart/Shutterstock.com

Le tecnologie del linguaggio basate sull'intelligenza artificiale stanno aprendo nuove strade per la comunicazione digitale in tutte le lingue europee. Tuttavia, mancano gli strumenti e le risorse di tecnologia linguistica per il maltese. Sin dal suo avvio nel 2021, il progetto LT-BRIDGE, finanziato dall'UE, si adopera per colmare tali lacune.

Nell'elaborazione del linguaggio naturale, i modelli linguistici vengono addestrati per l'associazione di parole con altre in un

determinato contesto mediante l'uso di approcci di rete neurale. I ricercatori dell'Università di Malta (UM), coordinatrice del progetto LT-BRIDGE, hanno addestrato un modello simile, BERTu, sui dati testuali in maltese.

Riempire gli spazi vuoti

Ma cosa sono esattamente i modelli linguistici? Kurt Micallef, dottorando presso l'Università di Malta, li descrive in un recente [articolo](#) pubblicato sulla pagina web del «Times of Malta»: «I modelli linguistici sono una comprensione astratta di una lingua. Li si può considerare un'“intuizione” di cosa sia una lingua. Ad esempio, se si dovesse riempire lo spazio vuoto nella frase “Jien _____ il-gazzetta” (Io _____ il giornale), potrebbe venire in mente “qrajt” (ho letto) oppure “xtrajt” (ho comprato), ma con minore probabilità si suggerirebbe “kilt” (ho mangiato) o “karozza” (auto).»

Un metodo per allenare tali modelli linguistici consiste nel ricorrere alla modellizzazione del linguaggio mascherato. Le parole in un testo vengono mascherate o coperte in modo casuale e il modello deve prevedere la parola mancante. «Pertanto, considerato l'esempio di cui sopra, il modello dovrebbe prevedere, idealmente, “qrajt”», spiega Micallef. Ciò viene ripetuto per molte frasi

affinché il modello linguistico possa apprendere il maltese. La rete neurale viene aggiornata con ogni frase mediante l'uso di algoritmi di apprendimento automatico e viene attribuita una probabilità a parole potenzialmente inseribili all'interno della frase.

Altri compiti

Due altri compiti su cui è stato addestrato BERTu sono l'analisi del sentimento e il riconoscimento delle entità nominate. «L'analisi del sentimento è il processo d'identificazione del sentimento di un determinato testo», osserva il ricercatore in [un altro articolo del «Times of Malta»](#). «La forma più semplice consiste nel decidere se una parte di testo trasmette un sentimento positivo o negativo rispetto a un determinato argomento o concetto. Ad esempio, considerati gli annunci sul bilancio Malta, il presente commento sostiene o è contrario alle comunicazioni diffuse? Questo tipo di compito è definito problema di classificazione, poiché per il testo che otteniamo come input, generiamo un'etichetta di classificazione (in questo esempio, positiva o negativa).»

Micallef illustra ulteriormente il secondo compito: «Il riconoscimento delle entità nominate è un compito d'etichettatura in cui generiamo un'etichetta per ogni parola contenuta nel testo di input. Dato un testo di input, il compito consiste nel decidere quali etichette si riferiscano a entità nominate e di che tipo di entità si tratti. Rispetto all'analisi del sentimento, il livello di questo compito è piuttosto basso e verrebbe tipicamente utilizzato per integrare altri sistemi linguistici. Ad esempio, potremmo adoperare i dati di classificazione per individuare i nomi di persona e renderli anonimi, in modo da rispettare le normative riguardanti la protezione dei dati.»

Il team di ricerca ha messo a punto il modello BERTu pre-addestrato su tali compiti aggiungendo un livello supplementare al di sopra del modello per ogni compito per poi eseguire algoritmi di apprendimento automatico su insiemi di dati al fine di scoprire i parametri del livello supplementare. È stato dimostrato che i risultati ottenuti da BERTu siano migliori rispetto a quelli di altri modelli linguistici, a volte di oltre il 20 %. Questo modello consente attualmente l'esplorazione di compiti di comprensione linguistica più complessi in maltese. Il Progetto LT-BRIDGE («Bridging the technology gap: Integrating Malta into European Research and Innovation efforts for AI-based language technologies») terminerà a dicembre 2023.

Per maggiori informazioni, consultare:
[pagina web del progetto LT-BRIDGE](#)

Parole chiave

[LT-BRIDGE](#)

[intelligenza artificiale](#)

[linguaggio](#)

[maltese](#)

[modello linguistico](#)

[BERTu](#)

[testo](#)

[dati testuali](#)

[parola](#)

Progetti correlati



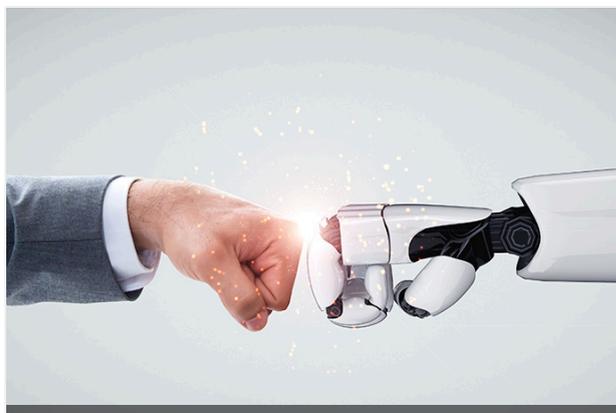
“Bridging the technology gap: Integrating Malta into European Research and Innovation efforts for AI-based language technologies”

LT-BRIDGE

7 Aprile 2025

PROGETTO

Articoli correlati



NOTIZIE

PROGRESSI SCIENTIFICI

Spianare la strada a un'IA affidabile



2 Novembre 2022



RISULTATI IN BREVE

L'applicazione personalizzata basata su IA che insegna le lingue straniere a livello di conversazione



20 Marzo 2020



In che modo le macchine interpretano il linguaggio umano



18 Marzo 2022



Una piattaforma di intelligenza artificiale vocale e conversazionale all'avanguardia incentiva la produttività aziendale



18 Marzo 2019

Ultimo aggiornamento: 2 Dicembre 2022

Permalink: <https://cordis.europa.eu/article/id/442665-computers-learn-maltese/it>

European Union, 2025