

Abstract

Questo progetto di tesi esplora il ruolo dell'analisi sintattica e delle sue proprietà di composizionalità nella comprensione dei fenomeni semantici, con un focus specifico sulla sentiment analysis. L'obiettivo principale è confrontare l'efficacia dei metodi kernel con i large language models (LLM) nella classificazione del sentiment a livello di frasi e delle loro componenti individuali. Il dataset utilizzato, Sentiment Penn TreeBank, etichetta sia le frasi intere che le loro singole componenti con il rispettivo sentiment, fornendo un ambiente ideale per questa analisi.

Inizialmente, sono stati testati metodi kernel all'avanguardia, combinando funzioni kernel per alberi, per sfruttare la struttura sintattica del testo. Successivamente, i risultati ottenuti sono stati confrontati con quelli derivanti dall'uso di LLM, che offrono avanzate rappresentazioni semantiche di token e frasi. In particolare, sono stati selezionati modelli LLM allenati su task di parsing per valutare le loro capacità sul dataset citato.

L'approccio metodologico ha incluso una rigorosa sperimentazione e ottimizzazione degli iperparametri per ciascun modello, seguita da un'analisi comparativa dei risultati. Il lavoro si è concluso con la documentazione dettagliata dei modelli sviluppati e dei risultati ottenuti, assicurando la replicabilità degli esperimenti. Sebbene le conclusioni definitive siano ancora in fase di elaborazione, i risultati preliminari suggeriscono che la combinazione di metodi kernel e LLM possa offrire miglioramenti significativi nella performance della classificazione del sentiment.