

Analisi di testi con tecniche di Deep Learning

R. Urso

13 Giugno 2023

1 Introduzione

2 Sentiment Analysis

- Emotion Detection

3 BERT

- Meccanismo di attenzione

4 Riferimenti

Introduzione

- Il seguente progetto denominato "Analisi di testi con tecniche di Deep Learning" mira all'applicazione di un modello di Deep Learning per l'elaborazione del linguaggio naturale (**NLP - Natural Language Processing**).
- Obiettivo del progetto è quello di estrarre il *sentimento* da un testo presente su un social network quale Facebook o Twitter mediante l'utilizzo di un nuovo modello chiamato BERT (**Bidirectional Representations from Transformers**).
I BERT sono stati introdotti nel 2018 da Jacob Devlin et al. del gruppo Google AI Language.

Sentiment Analysis

Perchè la *Sentiment Analysis*?

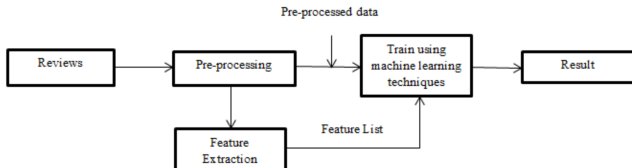
Oggigiorno, milioni di persone condividono le loro opinioni, i loro punti di vista ed i loro feedback con un semplice Tweet.

In un'ottica cliente-azienda, quest'ultima potrebbe essere interessata alla natura del tweet (se esprime un sentimento positivo, negativo o neutrale) circa un loro prodotto o servizio, in modo tale da poter migliorare *la customer satisfaction* e il prodotto stesso.

Che metodi sono utilizzati?

Sono utilizzati diversi algoritmi di Machine Learning, tra cui Naive Bayes, Maximum Entropy, Support Vector Machine oppure tecniche basate sul lessico, anche se spesso vengono utilizzate le Reti Neurali che restituiscono performance migliori.

- Il flow che viene seguito (Kharde et al., 2016) è il seguente:



- Il pre-processing dei tweet include i seguenti punti:
 - ➊ Rimozione degli URL, hashtag e tag
 - ➋ Correzione dei caratteri ripetuti
 - ➌ Rimozione delle emoticons, dei simboli e di numeri
 - ➍ Rimozione delle stop words
- Nella fase della feature extraction sono considerate:
 - ➊ Parole e la loro frequenza
 - ➋ Parti del discorso, come avverbi e aggettivi
 - ➌ Frasi di opinione
 - ➍ Posizione delle parole
 - ➎ Negazione

Emotion Detection

- L'**Emotion Detection** (ED) è una branca della Sentiment Analysis che cerca di estrarre emozioni da dati vocali, immagini o testi (Chen et al, 2021).
I modelli di emozione definiscono come un'emozione si differenzia dalle altre.
- Sono stati proposti diversi schemi di rappresentazione (discreta e dimensionale) riguardo questi modelli. Ritroviamo:
 - 1 Il modello di Paul Ekman (1999)
 - 2 Il modello di Robert Plutchik (1980)
 - 3 Il modello di Russell (1980)
- Se consideriamo l'estrazione di emozioni da testi scritti, la formulazione è la seguente:

$$r : A * T \rightarrow E$$

dove:

A è un autore che scrive un testo specifico,

T è il testo scritto da cui estrarre emozioni,

r è la relazione tra autore e testo

BERT

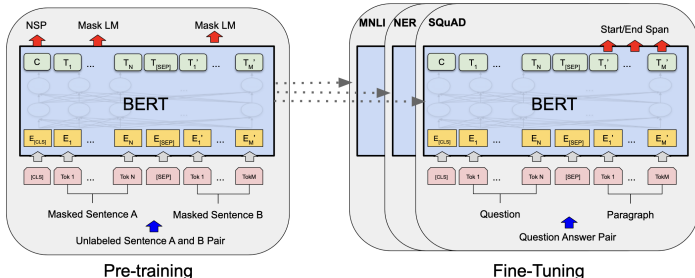
- Tale modello, a differenza dei recenti modelli di rappresentazione linguistica, è stato progettato per formare rappresentazioni profonde a partire da testo non etichettato (Devlin et. al., 2018).
- Esistono due strategie per applicare le rappresentazioni linguistiche:
 - 1 *feature-based*
 - 2 *fine-tuning*

Nel caso di approccio *fine.tuning* (come il **Generative Pre-trained Transformer - OpenAI GPT**, proposto da Radford et al. nel 2018), l'architettura è unidirezionale (da sinistra a destra) in cui ogni token può tendere solo ai token precedenti (Vaswan et al., 2017).

- BERT supera questo vincolo utilizzando un *modello di linguaggio mascherato (MLM)*, in cui, in maniera casuale, vengono mascherati alcuni token con l'obiettivo di riprodurli nella maniera più fedele possibile.

Fasi del BERT

- Sono previste due fasi:
 - 1 Fase di pre-training, in cui il modello viene addestrato su dati non etichettati.
 - 2 Fine-tuning, fase in cui i parametri sono *messi a punto*.



- Nello schema proposto, le coppie di frasi vengono riunite in un'unica sequenza, separate da un token speciale [SEP]. dove distingueremo i token relativi alla frase A e alla frase B.

- Tuttavia, si crea una discrepanza tra la fase di pre-training e il fine-tuning, in quanto il token [MASK] non viene utilizzato nel fine-tuning.

Come avviene il modello di linguaggio mascherato?

- Si maschera a caso una percentuale dei token in input con lo scopo di predirli.

Il problema si risolve non ponendo sempre il token [MASK].

Ciò che avviene è:

- 1 Rimpiazzo 80% delle volte con token [MASK]
 - 2 Rimpiazzo 10% delle volte con una parola random
 - 3 Nel restante 10%, il token resta invariato
- La fase della **Next Sentence Prediction - NSP** si basa sulla comprensione della relazione tra le due frasi, in modo tale da predire la frase successiva a quella considerata.
 - Considerando la fase di fine-tuning, essa risulta computazionalmente meno costosa rispetto a quella di pre-training.

Meccanismo di attenzione

Particolarità dei BERT

I BERT, più in generale i *Transformer Models*, si basa su meccanismi di attenzione, che evita la ricorrenza e consente una parallelizzazione all'interno degli esempi di addestramento.

- Più precisamente:

"L'autoattenzione è un meccanismo di attenzione che mette in relazione diverse posizioni di una singola sequenza per calcolare una rappresentazione della sequenza stessa."

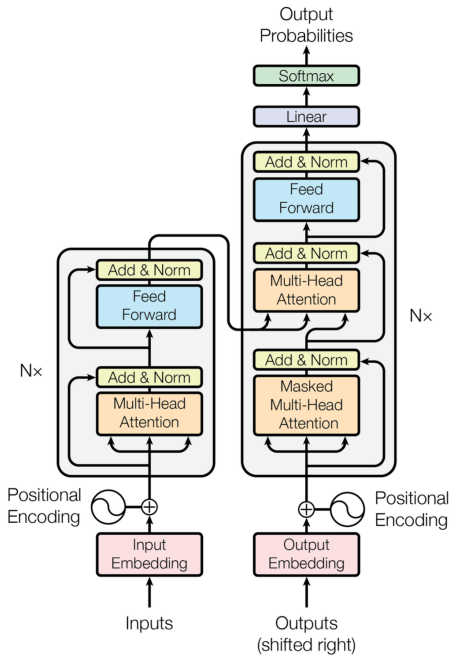
(Waswan et al., 2017).

Architettura

Il modello ha una struttura encoder-decoder, in cui l'encoder mappa una sequenza (x_1, x_2, \dots, x_n) in una sequenza di rappresentazioni continue (z_1, z_2, \dots, z_n) .

Il decoder genera una sequenza di simboli in uscita (y_1, y_2, \dots, y_n) .

- L'encoder è composto da 6 strati identici, dove ogni strato è composto da due sottolivelli:
 - ① Un meccanismo di attenzione multi-head
 - ② Una rete feed-forward
- Il decoder è anch'esso composto da 6 strati identici come l'encoder, ma presenta un terzo sottolivello che esegue l'attenzione multi-head sull'output dell'encoder.



Riferimenti

- Kharde, V.A., Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. arXiv preprint arXiv:1601.06971.
- Dang, N. C., Moreno-García, M. N., De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3), 483.
- Borele, P., Borikar, D. A. (2016). An approach to sentiment analysis using artificial neural network with comparative analysis of different techniques. IOSR Journal of Computer Engineering (IOSR-JCE), 18(2), 64-69.
- Acheampong, F. A., Nunoo-Mensah, H., Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review, 1-41.

- Sahayak, V., Shete, V., Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser L., Polosukhin I. (2017). Attention Is All You Need. Advances in neural information processing systems, 30.
- Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.