Dimensione Cognitiva
5. Come le macchine comprendono il linguaggio
Il Meccanismo di Attenzione

Giovanni Della Lunga
giovanni.dellalunga@unibo.it

A lezione di Intelligenza Artificiale

Siena - Giugno 2025

Indice

Il Meccanismo di "Attenzione"

Attenzione Selettiva

- L'**attenzione selettiva** è un meccanismo cognitivo fondamentale che consente agli esseri umani di **concentrare le risorse mentali** su un'informazione rilevante e **ignorare gli stimoli irrilevanti** presenti nell'ambiente.
- È ciò che ci permette, per esempio, di ascoltare una conversazione specifica in una stanza affollata (fenomeno noto come *cocktail party effect*), oppure di leggere un libro ignorando i rumori esterni.

Attenzione Selettiva

Dal punto di vista neurocognitivo:

- L'attenzione selettiva **filtra** l'informazione a livello sensoriale e/o percettivo.
- È **limitata**: non possiamo processare coscientemente tutto ciò che ci circonda, perciò il cervello seleziona ciò che ritiene più rilevante.
- È **dinamica**: può essere guidata da stimoli esterni (attenzione esogena) o da obiettivi interni (attenzione endogena).
- È spesso **modulata da contesto, esperienza e aspettative**.

Attenzione Selettiva

Il meccanismo di attenzione nei **modelli di deep learning**, come i transformer introdotti nel celebre paper *"Attention Is All You Need"*, si ispira (astrattamente) all'idea di attenzione umana. In particolare:

- Entrambi i sistemi **danno più "peso" alle informazioni rilevanti** rispetto a quelle irrilevanti.
- Entrambi **distribuiscono le risorse di elaborazione** in maniera non uniforme.
- L'attenzione nei transformer **filtra e valorizza dinamicamente** certi input rispetto ad altri in base al contesto, proprio come fa il cervello quando decide a cosa prestare attenzione.

Attention is all you need

Embeddings Contestuali

- Nel capitolo precedente abbiamo visto come, grazie a *Word2Vec*, sia stato possibile costruire dei vettori, gli *embeddings*, che permettono di far capire al computer nozioni complesse, come le parole e il loro significato.
- Purtroppo, **questa tecnica non è infallibile, anzi: se un embedding ben costruito ha la capacità di separare in maniera netta concetti diversi, è difficile che funzioni bene sempre.**

Embeddings Contestuali

- Immaginiamo di voler costruire gli *embeddings* delle parole “arancia” all’interno dello schema visto nel capitolo precedente.
- La parola “arancia” sarà più vicina a “limone” o a “Joker” (personaggio del film)?
- Anche se non c’è una risposta precisa, è molto probabile che, durante l’addestramento, la parola “arancia” sia stata principalmente associata al frutto, piuttosto che ad un film (ad esempio *Arancia meccanica*)
- In parole povere, è più probabile trovare frasi generiche che parlino dell’arancia come frutto che discorsi sulla pellicola cinematografica
- Quindi il suo *embedding* potrebbe essere simile a quello della parola “limone”.

Embeddings Contestuali

Ma in una frase come: **“Il film Joker mi ricorda molto Arancia Meccanica”** vorrei che la macchina intendesse questa parola più come un film piuttosto che come un frutto. Ed è anche quello che si aspettavano Ashish Vaswani e i suoi amici di Google nel 2017.

Embeddings Contestuali

- Nel mappare le parole, Word2Vec considera solo il loro contesto più prossimo, senza una comprensione più profonda del significato complessivo della frase.
- Ogni istanza di una specifica parola ha lo stesso vettore, indipendentemente dalla frase in cui appare, e questo può rappresentare un problema.
- Tra le parole che cambiano significato a seconda della frase possiamo citare "arancia" (frutto o film), ma anche "cane" (animale o parte di un'arma), "rosa" (persona, colore o fiore), "mela" (frutto o città, la Grande Mela) e così via.

Embeddings Contestuali

- Ashish e i suoi collaboratori capiscono che **serve quindi un trucco che modifichi il vettore della parola, l'embedding, in base alla frase in cui essa si trova**
- In particolare, puntano a spostare il vettore "arancia" verso "Joker" (film) nel caso di **"Il film Joker mi ricorda molto Arancia Meccanica"**
- oppure verso "limone" (frutto) nel caso di **"Arancia e limone contengono molta vitamica C"**

Attenzione Selettiva

- Ma come si fa? Con l'attenzione selettiva, che nel mondo dell'intelligenza artificiale si chiama *self-attention*.
- L'intelligenza artificiale come la conosciamo oggi deve gran parte dei suoi successi alla ricerca di questo gruppo di Google, e al loro famosissimo articolo chiamato *"Attention Is All You Need"*, focalizzato sulla traduzione automatica.

Attenzione Selettiva

Solo cinque o sei anni fa la traduzione automatica era ancora un grande problema. Se ci pensate bene, anche una frase semplice come

"Ti piace questo libro"

merita qualche accorgimento non banale nel caso volessimo tradurla automaticamente dall'italiano al francese.
Mi piace questo libro → J'**aime** ce livre
Ti piace questo libro → tu **aimes** ce livre
Ci piace questo libro → nous **aimons** ce livre
. . . Mi piace questo mare → J'aime **cette** mer

Attenzione Selettiva

- Per tradurre correttamente la parola "piace" in francese, il modello di traduzione ha bisogno di capire che si riferisce a "ti" che lo precede.
- Questo perché in francese, il verbo "piacere" cambia coniugazione a seconda del soggetto.
- Quindi per questa traduzione serve solo il pronome personale complemento!
- Il resto della frase è apparentemente inutile... attenzione selettiva!

Attenzione Selettiva

- Lo stesso discorso vale per l'aggettivo dimostrativo "questo"
- per una corretta traduzione, il modello ha bisogno di sapere che si riferisce alla parola "libro", perché in francese questo si traduce diversamente a seconda che il sostantivo a cui si riferisce sia maschile o femminile.
- In gergo si dice che "piace" presta molta attenzione a "ti", e che "questo" presta molta attenzione a "libro".

Attenzione Selettiva

Prendiamo la frase:

L'ANIMALE NON HA ATTRAVERSATO LA STRADA PERCHÉ **ERA** *TROPPO STANCO*

La parola "era" si riferisce all'animale, giusto? Immaginate di disegnaee una freccia che da "era" va verso "animale". Avere appena rappresentato il legame di attenzione tra le due parole.

Attenzione Selettiva

Ma cosa succederebbe se cambiassimo la parola "stanco" con la parola "trafficata"?

L'ANIMALE NON HA ATTRAVERSATO LA STRADA PERCHÉ **ERA** *TROPPO TRAFFICATA*

Così facendo "era" non si riferirà più ad "animale" ma a "strada"

Dobbiamo cambiare la nostra freccia!

Attenzione Selettiva

- La *self-attention* capisce tutto questo e lo rappresenta con una tabella piena di numeri chiamati scores.
- Ma come si calcola in pratica?
- Lo abbiamo già visto... con la similarità tra gli embeddings!
- Riprendiamo in esame le due frasi precedenti:
  “Il film Joker mi ricorda molto Arancia Meccanica”
  “Arancia e limone contengono molta vitamica C”
- Mappiamo tutte le parole con un *embedding* di tre dimensioni, "fruttosità", "filmosità", e una terza dimensione fittizia a caso.

Attenzione Selettiva

Gli *embeddings* delle parole saranno:

JOKER: [0, 1, 0] (solo filmosità)
ARANCIA: [0.5, 0.5, 0] (a metà tra fruttosità e filmosità)
LIMONE: [1, 0, 0] (sostituito con i valori di mela, quindi solo fruttosità)
UNA, ED, UN, E: [0, 0, 1] (asse fittizio)
MECCANICA: [0.1, 0.9, 0] (prevalentemente filmosità)

Attenzione Selettiva

Se moltiplico tra di loro le parole della frase ottengo una tabella con le varie similarità, ossia i prodotti scalari delle varie parole.

Attenzione Selettiva

- Questa rappresentazione mostra come le parole siano correlate tra loro nelle due frasi, in base alle dimensioni di "fruttosità", "filmosità" e l'asse fittizio.
- Ovviamente è solo un esempio, ma possiamo considerarlo come un rudimentale meccanismo dell'attenzione.
- Quindi se nella prima frase 'arancia' dipende un po' da se stessa e un po' da "limone", nel secondo esempio "arancia" viene trascinata verso l'asse della "filmosità" da "Joker".

Attenzione Selettiva

- Queste tabelle possono essere viste come delle matrici.
- Le matrici in matematica sono usate per modificare i vettori.
- Se moltiplicate un vettore (*embedding*) per una matrice (la tabella dell'attenzione) ottieni un nuovo vettore (*embedding*), ruotato e scalato rispetto all'originale.

Attenzione Selettiva

Quindi l'attenzione svolge proprio questo ruolo, ossia modifica tutti gli embeddings iniziali della nostra frase a seconda del contesto della frase stessa.

Un semplice Esempio Numerico