

深入淺出 SELF-ATTENTION — 讓深度學習和自然語言處理領域轟動的技術！



在自然語言處理（NLP）的歷史進程中，一個名為 Self-attention 的技術正顛覆傳統，引發行業內的轟動。這項技術是由 Transformer 架構所引入，這種架構完全摒棄了傳統的循環神經網路（RNN）和卷積神經網路（CNN），為處理語言任務帶來了一種更高效的方法。

自注意力機制

SELF-ATTENTION MECHANISM

Transformer 通過自注意力機制
聚焦於文本中最關鍵的部分
有效地避免了傳統 RNN 的
長期依賴問題，同時提
高了平行處理能力
Parallel Processing
Capability



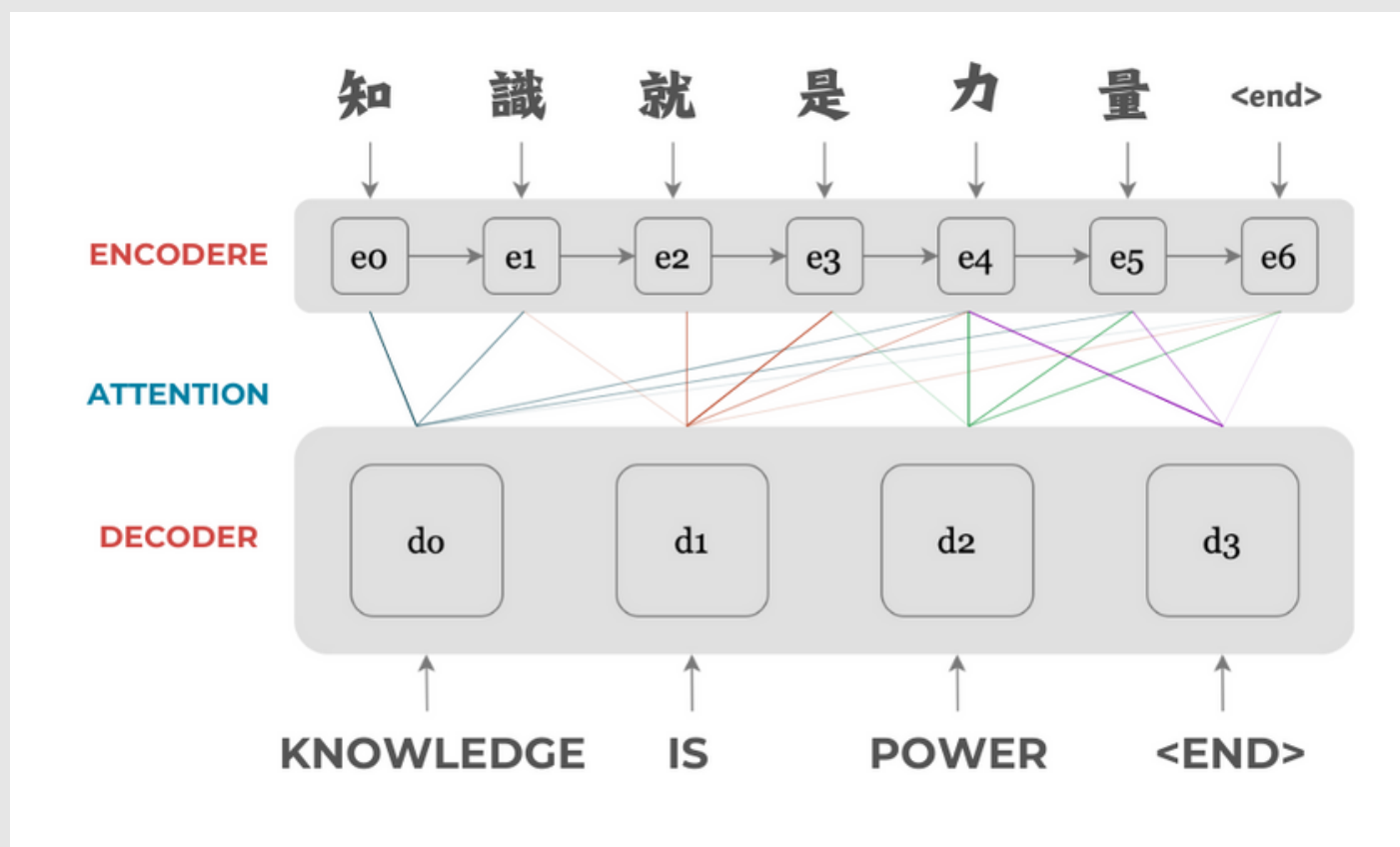
核心技術揭秘：

Transformer透過其核心組件——自注意力機制（self-attention mechanism），能夠專注於文本中最關鍵的部分。這不僅有效解決了RNN在處理長序列時的依賴問題，還大幅提升了平行處理的能力，開創了數據處理的新紀元。

詞向量化與其意義：

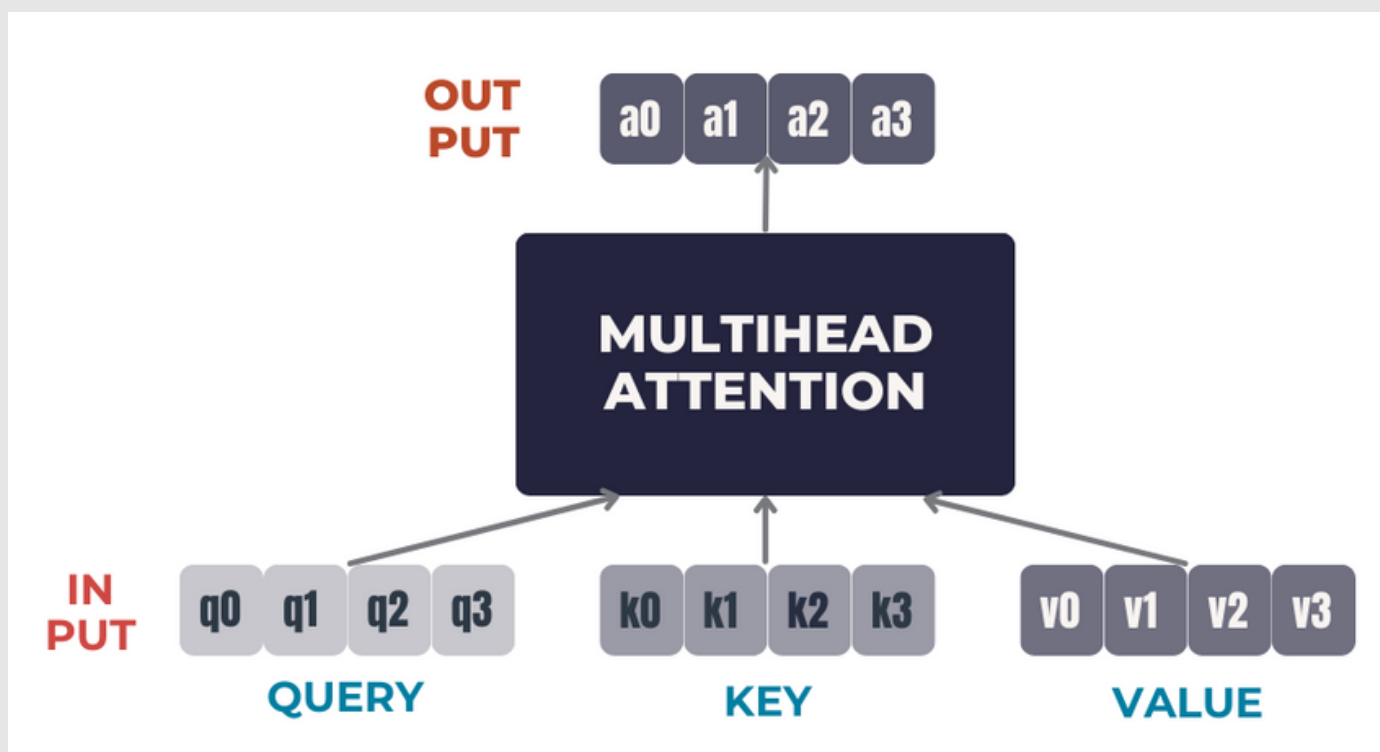
在NLP的核心操作中，每個詞都被轉換成數值向量，這些向量不僅代表著詞語本身，更蘊含了豐富的語義信息，使得模型能夠透過這些數學表示來捕捉詞語間的細膩語境關係。

EX：「陽明交通大學 人工智慧檢測中心」中的每個詞（「陽明」、「交通」、「大學」、「人工智慧」、「檢測」、「中心」）都被轉化成一個向量。這些向量是詞的數學表示，包含了詞的語義信息。「大學」這個詞的向量會與「學校」、「教育」等詞的向量在多維空間中相似。



QUERY, KEY, VALUE的角色扮演：

在Self-attention的過程中，每個詞的向量化表示將進一步被轉換成Query、Key、Value三種形式，這三者的相互作用就像是打開知識寶庫的鑰匙，找到並放大了最有「價值」的信息。



Query (Q)：看作是一個「問題」，它是對特定信息的查詢。
在計算注意力時，每一個詞的Query向量都會用來與其他所有詞的Key向量進行比較，以確定它們之間的相似性或關聯度。

Key (K)：被視為一種「索引」，它代表著可以被Query查詢的信息
當Query與Key進行點積運算時，這個得分決定了元素間的關聯程度。

Value (V)：包含了與Key相關的實際信息
當Key與Query匹配時，對應的Value將會被用來計算最終的輸出。
如果一個Key與Query的匹配程度越高，那麼這個Key對應的Value將會在最終輸出中佔有更大的權重

多頭注意力的引入：

為了捕捉數據的多維度特性，我們不僅使用一套Query、Key、Value，而是採用了多套配置，即所謂的多頭注意力機制，它讓模型能從多個角度同時捕捉信息，提供更全面的理解。

技術的重要性：

Self-attention機制的重要性在於其能夠直接計算序列中任意兩點之間的關聯性。這使得模型在捕捉長距離依賴關係上特別有效，解決了傳統RNN和LSTM模型在這方面的困難。

Self-attention技術不僅是一項理論上的創新，它的實際應用已經在語音識別、機器翻譯和內容摘要生成等領域展示了強大的效能，並且將繼續推動AI技術邁向更高的智能處理語言的未來。