



<https://ithelp.ithome.com.tw/questions> 技術文章 (<https://ithelp.ithome.com.tw/articles?tab=tech>)

<https://ithelp.ithome.com.tw/articles?tab=event> 2019 iT 邦幫忙鐵人賽

2019 iT 邦幫忙鐵人賽

(<https://ithelp.ithome.com.tw/ironmar>)

<https://ithelp.ithome.com.tw/tags> 聊天室 (<https://ithelp.ithome.com.tw/talks>)

(<https://ithelp.ithome.com.tw/ironman>)

<https://ithelp.ithome.com.tw/ironman?sc=nav>

sc=iThelpR)

累計文章數

9283 篇

參賽組數

523 組

完賽人數

264 人

團體組數

21 組

登入/註冊 (<https://ithelp.ithome.com.tw/users/login>)

▲ AI & Data (<https://ithelp.ithome.com.tw/ironman/ai-and-data>)

0

機器學習模型圖書館：從傳統模型到深度學習

(<https://ithelp.ithome.com.tw/users/20103529/ironman/1732>)系

☆ (<https://ithelp.ithome.com.tw/users/login>)

28 Transformer

2019鐵人賽

(<https://ithelp.ithome.com.tw/tags/articles/2019%E9%90%B5%E4%BA%BA%E8%B3BD>)

deep learning

(<https://ithelp.ithome.com.tw/tags/articles/deep%20learning>)

深度學習

(<https://ithelp.ithome.com.tw/tags/articles/%E6%B7%B1%E5%BA%A6%E5%AD%B8%E7%B5%B2>)

self-attention (<https://ithelp.ithome.com.tw/tags/articles/self-attention>)

attention model

(<https://ithelp.ithome.com.tw/tags/articles/attention%20model>)



杜岳華 (<https://ithelp.ithome.com.tw/users/20103529/ironman>)

2018-10-28 23:13:14

737 瀏覽

繼 Attention model 之後，由於 recurrent 架構的特性一直無法善用 GPU 的資源做加速。

這時 Google Brain 團隊就看到別人在用 self-attention 機制，也是基於 recurrent 架構，解決了不少問題，用在閱讀理解、摘要、文字蘊涵 (textual entailment) 及語句表示的問題上。

鐵人賽熱門主題

輕量高效.NET Core開源Blog引

擎:Miniblog.Core

(<https://ithelp.ithome.com.tw/users/2010>

5988/ironman/1699?sc=iThelpR)

三十天.NET ♥ Azure漸進式開發專案

(<https://ithelp.ithome.com.tw/users/2010>

5988/ironman/1622?sc=iThelpR)

程式設計師大冒險

(<https://ithelp.ithome.com.tw/users/2009>

1910/ironman/1874?sc=iThelpR)

Go劍復國-30天導入Golang

(<https://ithelp.ithome.com.tw/users/2009>

2379/ironman/2062?sc=iThelpR)

可不可以不要寫糙 code

(<https://ithelp.ithome.com.tw/users/2010>

7637/ironman/1927?sc=iThelpR)

Angular 深入淺出三十天

(<https://ithelp.ithome.com.tw/users/2009>

0728/ironman/1600?sc=iThelpR)

[Andriod] Andriod Studio 從入門到進入狀

況

(<https://ithelp.ithome.com.tw/users/2010>

5694/ironman/1642?sc=iThelpR)

DevOps學習之旅

(<https://ithelp.ithome.com.tw/users/2007>

2651/ironman/1779?sc=iThelpR)

英雄集結：深度學習的魔法使們

(<https://ithelp.ithome.com.tw/users/2011>

2540/ironman/2064?sc=iThelpR)

python 入門到分析股市

(<https://ithelp.ithome.com.tw/users/2011>

1390/ironman/1791?sc=iThelpR)

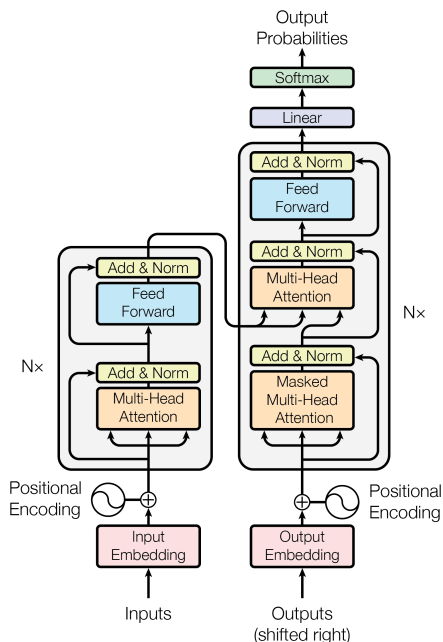
Google Brain 團隊就提出了史上第一個不需要依賴任何 recurrent 架構的 self-attention 機制模型，Transformer。

(跟變形金剛一樣的名字耶！帥吧！)

Transformer

這個架構上延續了 encoder-decoder 的架構，encoder 會將輸入的序列 (x_1, x_2, \dots, x_n) 轉換成 $z = (z_1, z_2, \dots, z_n)$ ，而 decoder 會將 z 轉換成 (y_1, y_2, \dots, y_m) ，一次轉換一個。在每一步當中，模型都是 auto-regressive 的，也就是說，前一次產生的結果會被當成下一次的輸入進行運算。

整個 Transformer 的架構就是在 encoder 及 decoder 上使用了 stacked self-attention 以及全連接的網路。我們來看圖，在圖的左半邊就是 encoder，右半邊就是 decoder：



Encoder

Encoder 由 N 個完全一樣的層堆疊 (stack) 起來 ($N=6$)。每層包含兩個子層，第一個是一個 multi-head self-attention 的機制，第二個是簡單的全連接層網路。每個子層外都包了 residual connection 以及 layer normalization，看起來就像 $LayerNorm(x + Sublayer(x))$ 。

Residual connection 主要可以將前層的資訊繞過一層，直接與下一層做運算。Layer normalization 有穩定學習效果的作用。

前往鐵人賽

([https://ithelp.ithome.com.tw/ironman?](https://ithelp.ithome.com.tw/ironman?sc=iThelpR)
sc=iThelpR)

熱門問題

請問VM和VPS的差異

(<https://ithelp.ithome.com.tw/questions/10193833>)

最近公司在評估次世代防火牆，但是不知道該從何比較...希望有前輩給點建議~

(<https://ithelp.ithome.com.tw/questions/10193806>)

沒百度帳號，想下載百度檔案

(<https://ithelp.ithome.com.tw/questions/10193866>)

數學問題(圓內周圍如何佈滿8個圓)

急!!!!!!

(<https://ithelp.ithome.com.tw/questions/10193823>)

超討厭光口跟電口的稱呼。

(<https://ithelp.ithome.com.tw/questions/10193809>)

想轉職程式設計師，該從哪種程式語言開始？

(<https://ithelp.ithome.com.tw/questions/10193867>)

點陣式噪音疑問

(<https://ithelp.ithome.com.tw/questions/10193794>)

SQL計算同一橫列的null數

(<https://ithelp.ithome.com.tw/questions/10193841>)

SQL的日期計算~自動推算每5天一個日期顯示

(<https://ithelp.ithome.com.tw/questions/10193797>)

疑似DNS被惡意攻擊

(<https://ithelp.ithome.com.tw/questions/10193863>)



(https://adcl.pchome.com.tw/clickAd2.html?q=47~ithome_forum~1262326)

· 不斷進取，追求卓越，成為更好的IT人

(https://adcl.pchome.com.tw/clickAd2.html?q=2~ithome_forum~963316)

Decoder

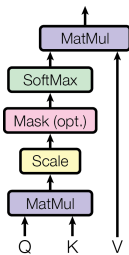
Decoder 大致上與 encoder 相同，差別是在用了兩個 multi-head self-attention 的機制，所以總共有3個子層。

Attention

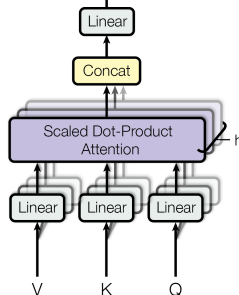
在 attention 的機制上，我們在上一篇講過了。這邊我們要進一步探討這個模型用到的 scaled dot product attention。在這邊就是分成 query、key 跟 value 三者，首先要先將 query 跟所有 key 做過一次的內積，並且除以 $\sqrt{d_k}$ ，然後過一次 softmax 函數。計算到這邊就是權重的部份，最後權重再跟 value 去計算結果。其中 d_k 是 key 向量的維度。公式在這邊：

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

Scaled Dot-Product Attention



Multi-Head Attention



圖的左邊是 scaled dot product attention。為什麼要除以 $\sqrt{d_k}$ 呢？文章中有提到，內積會讓整個結果變很大，會讓梯度變得極小，這會不利於訓練，所以需要除以 $\sqrt{d_k}$ 。

在圖的右邊，是 multi-head self-attention，核心就是平行計算多個 scaled dot product attention 的結果，並把這些結果都串接起來。有了這樣的機制就可以不只注意一個地方，可以有多个關注點。

在 self-attention 的機制，意味著所有的 query、key 跟 value 都來自於自己。不像之前的 attention 橫跨 encoder 跟 decoder，所以資訊會從雙方而來。

在 Transformer 模型當中，有一個是 encoder-decoder attention layer，然後 encoder 跟 decoder 各有一個 self-attention layer，就是各自的第一個子層。

熱門tag

看更多 ▸ (<https://ithelp.ithome.com.tw/tags>)

鐵人賽

(<https://ithelp.ithome.com.tw/tags/questions/%E9%90%>)

2019鐵人賽

(<https://ithelp.ithome.com.tw/tags/questions/2019%E9%90%>)

2018鐵人賽

(<https://ithelp.ithome.com.tw/tags/questions/2018%E9%90%>)

2017鐵人賽

(<https://ithelp.ithome.com.tw/tags/questions/2017%E9%90%>)

windows

(<https://ithelp.ithome.com.tw/tags/questions/windows>)

windows server

(<https://ithelp.ithome.com.tw/tags/questions/windows>)

javascript

(<https://ithelp.ithome.com.tw/tags/questions/javascript>)

linux

(<https://ithelp.ithome.com.tw/tags/questions/linux>)

php

(<https://ithelp.ithome.com.tw/tags/questions/php>)

程式設計

(<https://ithelp.ithome.com.tw/tags/questions/%E7%A8>)

分享

(<https://ithelp.ithome.com.tw/tags/questions/%E5%88>)

資訊安全

(<https://ithelp.ithome.com.tw/tags/questions/%E8%B3>)

職場

(<https://ithelp.ithome.com.tw/tags/questions/%E8%81>)

專案管理

(<https://ithelp.ithome.com.tw/tags/questions/%E5%B0>)

工作

(<https://ithelp.ithome.com.tw/tags/questions/%E5%B7>)

伺服器

(<https://ithelp.ithome.com.tw/tags/questions/%E4%BC>)

網頁設計

(<https://ithelp.ithome.com.tw/tags/questions/%E7%B6>)

sql

(<https://ithelp.ithome.com.tw/tags/questions/sql>)

c#

(<https://ithelp.ithome.com.tw/tags/questions/c%23>)

mysql

(<https://ithelp.ithome.com.tw/tags/questions/mysql>)

如此構成了整個 Transformer 模型，如果各位想知道這個模型的應用跟效能的話，請移駕去看論文，論文寫的還蠻簡單易懂的。

當然這麼模型當中有不少巧思在裡頭，有需要說明的話就提問囉！

💬 留言 ❤️ 追蹤 (<https://ithelp.ithome.com.tw/users/login>)

🚩 檢舉

(<https://ithelp.ithome.com.tw/users/login>)

< 上一篇

27 Attention model

(<https://ithelp.ithome.com.tw/articles/10205838>)

下一篇 >

29 Autoregressive generative model

(<https://ithelp.ithome.com.tw/articles/10206464>)

系列文

機器學習模型圖書館：從傳統模型到深度學習

(<https://ithelp.ithome.com.tw/users/20103529/ironman/1732>)

共 31 篇

目錄 📁 訂閱系列文

(<https://ithelp.ithome.com.tw/users/login>)

27 27 75 人訂閱

Attention model (<https://ithelp.ithome.com.tw/articles/10205838>)

28 28 Transformer (<https://ithelp.ithome.com.tw/articles/10206148>)

29 29 Autoregressive generative model

(<https://ithelp.ithome.com.tw/articles/10206464>)

30 30 結語 (<https://ithelp.ithome.com.tw/articles/10206812>)

31 番外篇 AI 的核心 (<https://ithelp.ithome.com.tw/articles/10206910>)

🔗 完整目錄

(<https://ithelp.ithome.com.tw/users/20103529/ironman/1732>)

尚未有邦友留言

立即登入留言 (<https://ithelp.ithome.com.tw/users/login>)

熱門回答

想轉職程式設計師，該從哪種程式語言開始？

(<https://ithelp.ithome.com.tw/questions/10193867>)

點陣式噪音疑問

(<https://ithelp.ithome.com.tw/questions/10193794>)

最近公司在評估次世代防火牆，但是不知道該從何比較...希望有前輩給點建議~

(<https://ithelp.ithome.com.tw/questions/10193806>)

產品資料庫應如何設計

(<https://ithelp.ithome.com.tw/questions/10193828>)

SQL計算同一橫列的null數

(<https://ithelp.ithome.com.tw/questions/10193841>)

沒百度帳號，想下載百度檔案

(<https://ithelp.ithome.com.tw/questions/10193866>)

請問VM和VPS的差異

(<https://ithelp.ithome.com.tw/questions/10193833>)

javascript image擷取

(<https://ithelp.ithome.com.tw/questions/10193876>)

超討厭光口跟電口的稱呼。

(<https://ithelp.ithome.com.tw/questions/10193809>)

SQL的日期計算~自動推算每5天一個日期顯示

(<https://ithelp.ithome.com.tw/questions/10193797>)



(https://adcl.pchome.com.tw/clickAd2.html?q=9~ithome_forum~1263554)

熱門文章

AWS從零開始，你有可能碰到的東西

(<https://ithelp.ithome.com.tw/articles/10211894>)

有禮貌的公司郵件起手式(筆記)

(<https://ithelp.ithome.com.tw/articles/10211893>)

USB 隨身碟檔案還原必學，輕鬆救回刪除

的 USB 資料
(<https://ithelp.ithome.com.tw/articles/10211908>)
剛開始接觸...
(<https://ithelp.ithome.com.tw/articles/10211898>)
Web網站系統程式碼比重介紹
(<https://ithelp.ithome.com.tw/articles/10211889>)
DAY5.PHP-檢查輸入字串的格式
(<https://ithelp.ithome.com.tw/articles/10211905>)
DAY4.PHP-foreach迴圈與核取方塊
(<https://ithelp.ithome.com.tw/articles/10211895>)
DAY5.PHP-隨機顯示廣告影片
(<https://ithelp.ithome.com.tw/articles/10211903>)
台灣中小企業採用之商用資料庫品牌佔有率
(<https://ithelp.ithome.com.tw/articles/10211941>)
DAY6.PHP-將半形文字轉為全型
(<https://ithelp.ithome.com.tw/articles/10211918>)

一週點數排行

更多點數排行 ▶

(https://ithelp.ithome.com.tw/point_rank/total)

- 1 浩瀚星空 (yoching)
(<https://ithelp.ithome.com.tw/users/20101641>,
- 2 暉翰 (shps951002)
(<https://ithelp.ithome.com.tw/users/20105988>,
- 3 rogeryao (rogeryao)
(<https://ithelp.ithome.com.tw/users/20085021>,
- 4 dragonH (jack61811)
(<https://ithelp.ithome.com.tw/users/20117259>,
- 5 raytracy (raytracy)
(<https://ithelp.ithome.com.tw/users/20026603>,
- 6 小魚 (henrychang)
(<https://ithelp.ithome.com.tw/users/20105694>,
- 7 舜~ (chaosfractal)
(<https://ithelp.ithome.com.tw/users/20110337>,
- 8 海綿寶寶 (antijava)
(<https://ithelp.ithome.com.tw/users/20001787>,
- 9 froce (froce)
(<https://ithelp.ithome.com.tw/users/20104001>,
- 10 echochio (echochio)
(<https://ithelp.ithome.com.tw/users/20110611>,

(https://adcl.pchome.com.tw/clickAd2.html?q=41~ithome_forum~1262436)

iThome 服務 ▶ iThome online (<https://www.ithome.com.tw/>) iThome Learning (<https://learning.ithome.com.tw/>)

電週文化事業版權所有、轉載必究 | Copyright © iThome 刊登廣告 (<https://www.ithome.com.tw/aboutus/>)

授權服務 (<https://www.ithome.com.tw/aboutus/>) 服務信箱 (<mailto:ithelp@mail.ithome.com.tw>)

隱私權聲明與會員使用條款 (<https://www.ithome.com.tw/terms>) iT邦幫忙使用說明 (<https://ithelp.ithome.com.tw/guide>)