

Homework2: Best Match Models

| B10632026 吳苡瑄

這次做 Best Match 的算是比較簡單，只要照著定義去把算是做出來就差不多了，一開始我也是就先把 BM25 做出來，然後慢慢調參數，大概是 $b = 0.9$ 、 $k1 = 1.5$ 可以到過 baseline 好一點，然後 $k3$ 不管調什麼都沒什麼差別所以我就一直都固定在 1 沒怎麼動它。

到了一直卡在 baseline 附近後，就開始做一些奇怪的調整，像是把 BM11 裡面的前面那一項加進去看看，效果與沒有特別好，後來學長跟我說有個奇妙的事情，就是在 sigma 裡面加上開 1.2 次方不知道為什麼效果會稍微變好，很莫名其妙但是還真的，所以後來就對所有東西都做了開方啊，次方之類的事情，還有跟學長他們比出了一個最厲害的奇怪算式，可以拿到 0.723 之類的分數，底下附上怪算式本人。

```
up = (k1 + 1) * tfij * (k3 + 1) * tfiq * pow(log_thing, 1.3)
down = (k1*((1 - b) + (b * len(doc_content) / avg_doclen)) + tfij) * (k3 + tfiq)
sim += giq + (up / down) ** 1.19
```

最後為了衝高一點分數，用了一個雖然作業一就知道但沒用的奧步，就是把所有 query 裡面有出現的字都抓出來，document 就只考慮這些字就好的話，分數就會好很多，所以最後我就用了這個方法 + 奇怪算式 + ($b = 0.87$ 、 $k1 = 3$ 、 $k2 = 0.1$ 、 $k3 = 1$) 到 0.74 多了，結果分數那麼高不是靠調參數，是靠兜那些式子和一些小奧步，感謝電機系的學長和 NLP 實驗室的 Hank 學長幫忙，可喜可賀，雖然很方式很怪但是也還在 BM 範圍裡 (用 BM11、BM25) 湊的，應該還是有遵守男子漢的約定啦！