

使用元學習技術於資訊檢索任務之初步研究

A Preliminary Study on Using Meta-learning Technique for Information Retrieval

林崇恩 Chong-En Lin

國立臺灣科技大學系資訊工程系

Department of Computer Science and Information Engineering

National Taiwan University of Science and Technology

M10815018@mail.ntust.edu.tw

陳冠宇 Kuan-Yu Chen

國立臺灣科技大學系資訊工程系

Department of Computer Science and Information Engineering

National Taiwan University of Science and Technology

kychen@mail.ntust.edu.tw

摘要

資訊檢索(Information Retrieval)是從海量的文件中，依據使用者提供的查詢(Query)關鍵字，從而回傳和關鍵字相關的文件，這個任務是自然語言處理(Natural Language Processing, NLP)中相當重要的研究方向。在資訊檢索中，使用少量訓練資料進行模型訓練，往往得到的成績不是很理想。所以本研究提出運用元學習(Meta Learning)中與模型無關的元學習演算法(Model-Agnostic Meta-Learning, MAML)於資訊檢索任務，透過與模型無關的元學習演算法從不同主題的文件集學習模型初始化參數，實現在新主題資料集中快速適應文件的能力。在實驗中，我們的方法使用深度類神經網路(Deep Neural Networks)架構的檢索模型（深層結構語義模型 Deep Structured Semantic Model, DSSM）搭配與模型無關的元學習演算法進行訓練，並在 TREC Robust04 資料集上進行評估，實驗結果初步表明使用與模型無關的元學習演算法訓練方式相較於傳統訓練方式有明顯的分數提升。就我們所知，本研究是首先使用元學習技術應用在資訊檢索領域，開啟了後續研究的新方向。

關鍵詞：資訊檢索、元學習、與模型無關的元學習、深層結構語義模型

Abstract

Information retrieval aims at searching documents to satisfy a user's query from a large collection of documents. The research subject is a fundamental challenge in the context of natural language processing. We all agree upon that obtaining a neural network-based retrieval model usually needs a large amount of training examples. To modulate the problem, in this study, we make a step forward to leverage the meta learning strategy for training a neural network-based retrieval model. More formally, the model-agnostic meta-learning algorithm (MAML) is employed to learn a set of model parameters from a set of training queries. After that, the model is easily to be adapted for each test query. By doing so, query dependent retrieval model can be simply obtained and achieves better results than conventional retrieval models. In the experiments, a simple neural network-based retrieval model (i.e., deep structured semantic model, DSSM) is adopted, and the resulting model is evaluated on the TREC Robust04 dataset. Preliminary results show that the use of model-agnostic meta-learning algorithm training method has a significant improvement compared with traditional training methods. As far as we know, this research is the first one, which introduces the meta-learning technique to information retrieval task.

Keywords: Information Retrieval, Meta-learning, Model-agnostic Meta-learning, Deep Structured Semantic Models

一、緒論

伴隨著互聯網蓬勃的發展和網路文件資訊的增長，各種不同瀏覽器的搜尋引擎功能，已經是使用者挑選瀏覽器的必要條件之一，而在網路的世界中，存在的文件檔案就有成千上萬筆資料，如何能有效率搜尋到和使用者相關的文件，這就是檢索系統中資訊檢索 (Information Retrieval) 的重要研究課題。資訊檢索的技術大致可以分為三大類：多媒體資訊檢索、資料庫檢索與文件資訊檢索，本研究主要探討和文件檔案相關的文件資訊檢索技術。和文件檔案相關的文件資訊檢索技術以在瀏覽器中的搜尋引擎為主，運作方式為使用者在搜尋引擎中的查詢介面提供查詢 (Query) 的關鍵文字，然後搜尋引擎會協助使用者快速檢索到和關鍵文字相關 (Relevance) 的網路文件資訊。

近年來在機器學習領域中有一個新興的研究方向：元學習 (Meta Learning)，這項技術近幾年在機器學習的各大領域引起了廣泛的討論。元學習的概念是讓機器學習到「如何學習 (Learning to Learn)」的能力，與一般的機器學習最大的差別在於元學習會先學習到一個先驗知識，這會讓機器在面對新的任務時，能更快更好得適應新的數據。在資源匱乏的訓練資料環境設置中，元學習的技術已經在很多領域上取得了不錯的成績，例如

影像視覺方面[1]、機器翻譯方面[2]和語音處理方面[3-6]。就我們所知，到目前為止，尚無研究探討使用元學習技術於資訊檢索任務之中，為了探究此一研究方向，本研究提出使用元學習技術於資訊檢索的任務，期望可以獲得任務成效的提升。

二、相關研究

(一)傳統資訊檢索模型

1、向量空間模型(Vector Space Model, VSM)

在向量空間模型中，我們會將每個查詢和文件以向量表示，其中每一個分量對應於字典中的每個單字，在空間中靠近的向量在遣詞用字上較為相近，所以我們會依照查詢文件向量與不同檢索文件向量之間的餘弦相似度(Cosine Similarity)進行文件排序。更明確地，向量中的每一個分量，通常使用詞頻－倒文件頻(Term Frequency-Inverse Document Frequency, TF-IDF)的方式計算權重：

$$TF - IDF_{i,j} = tf_{i,j} \times idf_i \quad (1)$$

$$tf_{i,j} = 1 + \log(count(w_i, d_j)) \quad (2)$$

$$idf_i = \log \left[\frac{1 + N}{1 + df(w_i)} \right] + 1 \quad (3)$$

其中， $TF - IDF_{i,j}$ 代表單詞 w_i 在文件 d_j 向量中的權重，這個權重由局部權重($tf_{i,j}$)和全局權重(idf_i)的乘積所組成。局部權重是以單詞 w_i 在文件 d_j 中出現的頻率 $count(w_i, d_j)$ 經過簡單的計算而得；全局權重則是以單詞 w_i 出現在整個文件集中的文章數 $df(w_i)$ 和整個文件集的文章數 N ，透過簡單的計算而得。因此，如果單詞 w_i 在文件 d_j 中出現的次數很高，並且它僅在少數的文件中出現，則此一單詞是相當重要的資訊；反之，若單詞 w_i 在很多篇文件中都有出現，或是在文件 d_j 中出現的次數很低，甚至是 0，則表示此一單詞對文件 d_j 而言並不重要。

2、最佳匹配模型 25(Best Match 25, BM25)

最佳匹配模型 25(Best Match 25, BM25)是資訊檢索領域用來計算查詢文件與檢索文件相似度分數的經典模型。最佳匹配模型是在對機率模型的變化進行一系列實驗後創建的，包括最佳匹配模型 1、包括最佳匹配模型 11、包括最佳匹配模型 15 以及最佳匹配模型 25 等模型。最佳匹配模型 25 是一個融合語句的倒文件頻率資訊、詞頻資訊以及文件長

度標準化的排序計算公式：

$$BM25(d_j, q) = \sum_{w_i \in \{d_j \cap q\}} tf_{i,j}^1 \times tf_{i,q}^2 \times idf_i \quad (4)$$

$$tf_{i,j}^1 = \frac{(K_1 + 1) \times count(w_i, d_j)}{K_1 \times \left[(1 - b) + b \times \frac{len(d_j)}{avg_{doclen}} \right] + count(w_i, d_j)} \quad (5)$$

$$tf_{i,q}^2 = \frac{(K_3 + 1) \times count(w_i, q)}{K_3 + count(w_i, q)} \quad (6)$$

$$idf_i = \log \left(\frac{N - df(w_i) + 0.5}{df(w_i) + 0.5} \right) \quad (7)$$

其中， d_j 為檢索文件， q 為查詢文件， N 為所有檢索文件的數目， $count(w_i, d_j)$ 和 $count(w_i, q)$ 分別為單詞 w_i 在文件 d_j 和查詢 q 中出現的次數， $df(w_i)$ 為 w_i 出現在整個文件集中的文章數， $len(d_j)$ 為文件 d_j 的長度（字數）， avg_{doclen} 為所有文件的平均長度（字數）， K_1 、 K_3 和 b 為可調整的模型參數。由最佳匹配模型 25 的計算方式可知，字詞出現在查詢文件 q 的頻率資訊會藉由權重函數 $tf_{i,q}^2$ 進行適當的調整，其中參數 K_3 會考慮到字詞是否出現於查詢文件當中，並且更進一步地將字詞在查詢文件中出現的頻率資訊做加權運算；權重函數 $tf_{i,j}^1$ 則用於計算當前文件與查詢共同出現的字詞在檢索文件中的重要性，並且對文件的語句長度進行標準化，最後用可調參數 K_1 負責進行適當加權調整；倒文件頻率資訊 idf_i 是用來決定一個字詞普遍重要性的度量，也就是會加強比較稀少出現字詞的權重，而削弱比較常出現字詞的重要性。

(三)元學習相關研究

通常在進行元學習或小樣本學習時，會有幾個重要的專有名稱，分別是： N 類別 K 樣本 (N -way K -shot)、支撐集(Support Set)、查詢集(Query Set)、元學習訓練集($D_{meta-train}$)以及元學習測試集($D_{meta-test}$)。 N 類別 K 樣本是在進行小樣本學習中常見的實驗設置，是指模型訓練一次任務時所會用到的資料數量，例如在分類任務中，表示每個訓練任務包含 N 個類別，並且每個類別分別有 K 筆訓練資料，因此，每個訓練任務共會使用 $N \times K$ 筆資料進行訓練。元學習訓練集則是由多組訓練任務資料組合而成，每組訓練任務皆包含支撐集和查詢集。這些資料通常是利用具有豐富資源的資料來進行模型預訓練。元學

習測試集同樣是由多組支撐集和查詢集組合而成。這些資料通常是用在新的任務上，模型會先經過元學習訓練集的資料訓練後得到一個預訓練的模型或者是一組模型參數，繼續使用元學習測試集中的支撐集進行模型微調，最後使用元學習測試集中的查詢集進行模型測試。

元學習方法大致分為三類：基於度量(Metric-based)、基於模型(Model-based)和基於參數優化(Optimization-based)。基於度量的元學習方法是學習度量空間，然後將其比較低資源的測試樣本和豐富的訓練樣本之間距離分佈進行建模，使得同類樣本靠近，異類樣本遠離，此類別的代表方法包括孿生網路(Siamese Neural Networks)[7]、原型網路(Prototype Network)[8]、匹配網路(Matching Networks)[9]和關係網路(Relation Network)[10]。基於模型的元學習方法是使用一個新的元學習器模型，通過一些訓練示例學習參數，讓原始模型使用其參數後有更好結果，此類別的代表方法包括元學習長短期記憶(Meta Learning Long Short-term Memory)[11]。基於優化參數的元學習方法通常設計成有利於快速適應新任務的演算法，此類別的代表方法為可擴展的元學習演算法(Reptile)[12]。

孿生網路是屬於基於度量的元學習方法。孿生網路具體的運作原理是透過監督式(Supervised)的方式來訓練具有兩個輸入的類神經網路，訓練時通過不同的成對輸入樣本進行訓練，兩個輸入樣本分別對應到元學習訓練(Meta-train)的支撐集(Support Set)和查詢集(Query Set)，在最後一層對經過訓練所得到的兩個嵌入向量進行相似度計算，判斷它們之間是否屬於同一個類別，並且產生對應的機率分布。原型網路是屬於基於度量的元學習方法，它的運作原理和孿生網路很類似，兩者之間的差別在於孿生網路每次輸入到類神經網路的資料只能兩筆，而原型網路輸入到類神經網路則可以輸入兩筆以上資料。原型網路在訓練時，每個任務中的每個類別裡，樣本經過學習後得到相對應的度量空間，每個類別的嵌入會在內部類別樣本中進行均值計算，然後會得到各自代表相對應類別的一個原型來表示這個類別的嵌入，最後就是拿測試樣本經過學習後得到的度量空間，將測試的度量空間對每個類別的原型表達進行相似度計算，最後進行交叉熵計算損失來更新網路。匹配網路是屬於基於度量的元學習方法，它的運作原理和原型網路很相似，兩者之間的差別在於原型網路輸入到類神經網路的資料都是分開處理，而匹配網路則是輸入到類神經網路的資料是一起處理，因為匹配網路覺得輸入的資料彼此間互相關係。另外一個差別是匹配網路在計算出分數以後，會經過多層次(Multiple Hop)的處理，

最後再經過軟性最大化函數(Softmax Function)得到每個輸入相對應的類別機率。關係網路是屬於基於度量的元學習方法，它的運作原理和原型網路很類似，最主要的差別是關係網路先把訓練資料和測試資料抽取其各自的嵌入，接著將測試資料的嵌入接在每個訓練資料的嵌入後面，最後經過另一個新的類神經網路訓練算出其測試資料和其對應訓練資料的相似度為多少。

元學習長短期記憶(Meta Learning Long Short-term Memory)是屬於基於模型的元學習方法，它的構想是把梯度下降法用長短期記憶(Long Short-term Memory, LSTM)來取代，透過長短期記憶進行模型參數更新。這個方法目的是藉由吸取所有任務的基礎知識，以及透過少量的資料訓練，來提升模型泛化的能力和提供對模型有更好的初始參數。可擴展的元學習演算法(Reptile)是屬於基於梯度優化的元學習演算法，它的運作原理是學習類神經網路初始化參數的方法，讓類神經網路在新的目標任務上，僅使用少量新任務訓練就可以得到好的結果。可擴展的元學習演算法的運作方式是在每個任務中執行大於一次的隨機梯度下降法(Stochastic Gradient Descent, SGD)或適應性矩估計演算法(Adam)，每個任務結束後再更新參數。

(四)與模型無關的元學習演算法(Model-Agnostic Meta-Learning, MAML)

與模型無關的元學習演算法[13]是屬於基於優化參數的元學習方法，這種類型的方法通常應用在類神經網路架構的模型上，目的是希望藉由使用與模型無關的元學習演算法直接優化梯度，讓模型進行參數優化，最終學習到一組表現很好的初始參數，並且將訓練過後的模型參數當作新任務的初始參數，接續在新任務上對模型進行微調(Fine-tuning)動作，也就是將新的任務進行傳統訓練，最後得到比直接使用傳統預訓練方式有更好的結果。通常元學習的方法使用背景是在資源匱乏(Low-resource)的任務上進行訓練，然而在低資源的任務上進行模型訓練，最常遇見的問題可能會發生模型過度擬合(Overfitting)的現象，更詳細地說模型在面對訓練集有很好的成績，但面對測試集時模型就會表現得很差。所以近年來有許多新興的技術在研究如何在資源匱乏的任務上得到好成績，而元學習的優點剛好是它能在新的任務中，僅使用較低的資料進行訓練，並且經由少次的梯度下降法(Gradient Descent)進行參數更新，模型就會取得良好的泛化效果。演算法一是與模型無關的元學習演算法虛擬碼。

演算法中的 $p(\mathcal{T})$ 代表的是元學習訓練集(即 $D_{meta-train}$) 中任務的分布， α 、 β 指的

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks

Require: α, β : step size hyperparameters

```
1: Randomly initialize  $\theta$ 
2: while not done do
3:   Sample batch of tasks  $\mathcal{T}_i \sim p(\mathcal{T})$ 
4:   for all  $\mathcal{T}_i$  do
5:     Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  with respect to  $K$  examples
6:     Compute adapted parameters with gradient descent:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$ 
7:   end for
8:   Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$ 
9: end while
```

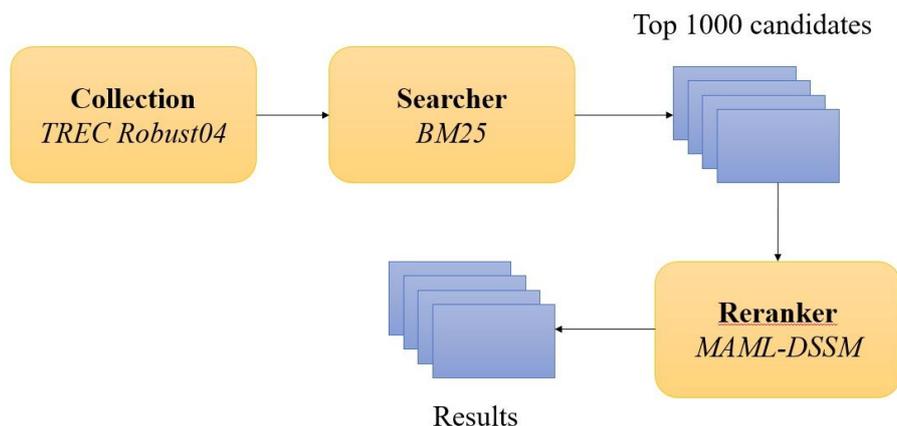
演算法一：與模型無關的元學習演算法虛擬碼。

是進行梯度下降法的學習率超參數。與模型無關的元學習演算法會先隨機初始化一組參數當作模型的初始參數，接著在每一輪迭代的過程中，先從 $p(\mathcal{T})$ 中取樣一堆任務，然後對每個任務中的支撐集計算參數梯度，並且更新當前任務中模型參數，重要的是每個任務都是從參數為 θ 的模型開始更新，做完全部任務後，會將每個任務的查詢集拿來和剛剛更新過參數的相對應模型來計算損失，然後使用得到的梯度來更新參數為 θ 的模型，最後迭代過程做完後會得到一組強而有力的參數，我們通常會把這組參數當作新任務的模型初始參數，然後進行模型微調動作。

三、使用元學習演算法於資訊檢索任務

在本研究中，我們提出的方法是把不同的查詢想像成不同的類別，而在資訊檢索中，每個查詢文件剛好會對應到幾篇和自身相關的檢索文件，這些對應到的文件可以視為這個類別下的資料，這樣的行為和機器學習中的分類問題很相似，所以我們就想把資訊檢索中的分類行為結合元學習中的與模型無關的元學習演算法；換句話說，我們使用與模型無關的元學習演算法來訓練資訊檢索模型，讓模型的參數在進行檢索任務時可以獲得更好的成效。

圖一為我們所提出方法的系統架構圖，整個流程為先使用最佳匹配模型 25 來為每個查詢進行檢索，檢索完後再利用本研究所提出的方法對前一千篇文件進行重新排序，重新排序過後的檢索結果即作為最終的輸出。



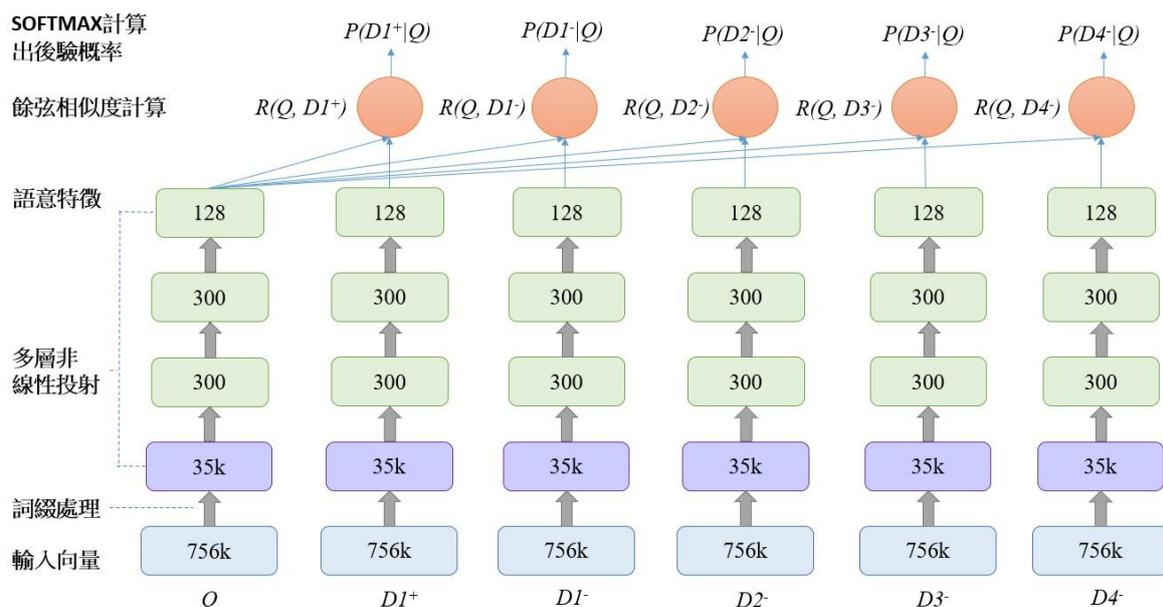
圖一、MAML-DSSM 系統架構圖。

(一)深層結構語義模型(Deep Structured Semantic Models, DSSM)

深層結構語義模型(Deep Structured Semantic Models, DSSM)[14]是基於深度神經網路的著名資訊檢索模型，它將查詢和文件投影至低維度的向量表示，再通過餘弦相似度計算查詢和檢索文件的相似度，最終訓練出可預測兩個句子之間語義相似度的低維度語義向量表達模型。更明確地，深層結構語義模型利用詞綴處理(Word Hashing)後，將查詢或文件轉換成詞頻－倒文件頻的向量表示，然後將這個向量經過三層全連接神經網路(Fully-connected Neural Network)進行降維，就可以獲得低維度的語義特徵向量。接下來，在訓練時，我們將一篇查詢和一篇正相關檢索文件以及四篇負相關檢索文件經過模型輸出後，分別獲得低維度的向量表示法，接著透過餘弦相似度計算，可以獲得這五篇文件對於這個查詢的相關分數，然後將這五個數值通過軟性最大函數(Softmax Function)，最後輸出一組機率分佈，而模型的訓練目標，即是希望正相關的檢索文件可以獲得非常接近 1 的分數，其餘四篇附相關文件則是獲得相當接近 0 的分數。圖二為深層結構語義模型的模型架構圖。

(二)結合與模型無關的元學習演算法與深層結構語義模型

本研究提出使用與模型無關的元學習演算法於訓練深層結構語義模型應用在資訊檢索的任務之中，此後我們簡稱為此一方法為 MAML-DSSM。在 MAML-DSSM 中，我們可以將訓練資料中不同的查詢想像成不同的類別，這樣的構想就和用無關的元學習演算法做分類問題的方法類似，所以每個類別的每一筆資料是由一篇查詢、一篇和查詢相關的



圖二、深層結構語義模型之模型架構圖。

檢索文件以及四篇和查詢不相關的文件組合而成。在我們的設定中，每一次訓練迭代過程會訓練 32 個任務，每個任務分別會獲取 10 篇查詢，也就是會提取 10 個類別，每個類別分別會拿 5 筆資料進行訓練和 5 筆資料進行測試，所以每訓練一個任務會一次訓練到 50 筆資料（即支撐集），然後這些資料在與模型無關的元學習演算法中的內部迴圈會進行 10 次隨機梯度下降法來更新模型參數，當完成內部迴圈後會使用另外 50 筆資料（即查詢集）計算損失。有一點要注意的是在每一次迭代過程中，做每個任務的模型初始參數都是同一組數值，所以在本實驗中，每一次迭代會藉由 32 個任務得到不同的損失值，當做完內部迴圈運算後，會用加總過的內部迴圈損失值經由適應性矩估計演算法來更新外部迴圈的模型參數，最後做完迭代次數後，我們將模型參數存下來以利在新任務上進行評估。在測試過程中，本研究的測試模型和訓練模型是相同架構，所以將經過與模型無關的元學習演算法得到的模型參數直接應用在測試模型上，查詢文件和檢索文件分別藉由測試模型得到 128 維語意特徵向量，然後去計算兩者間的餘弦相似度，最後就可以得到彼此的相似度分數。

四、實驗

(一) TREC Robust04 資料集

TREC Robust04 是由金融時報(Financial Times)、聯邦公報 94(Federal Register 94)、洛杉

磯時報(LA Times)和外國廣播資訊處(FBIS)的文件組合成一個大型資料集。該資料集包含 250 個主題文件：文件編號 301 到 450 為 TREC6 到 TREC8 資料集的主題文件，文件編號 601 到 650 為在 2003 年熱門發展的新主題文件，文件編號 651 到 700 為 2004 年新發展的主題文件，這些主題文件會取用文件內的主題(Title)語句當成實驗中所配置的查詢關鍵字。另外檢索文件則是由上述來源所提及的資料集彙整而成，共有 528,155 篇文章，並且使用文件中的文字區段(TEXT)當作實驗配置所對應的檢索文件。其中，TREC 特別設計此系列資料集，目的是用來評估資訊檢索系統面對困難查詢關鍵字的處理能力。

(二)實驗設置

我們對 TREC Robust04 資料集進行實驗，在 528,155 篇檢索文件中，先使用最佳匹配模型 25 進行初步檢索，然後對初步檢索後排名前 1,000 篇的候選文件進行重新排序。在本實驗中，我們採用 5-fold 交叉驗證，其中 4-fold 用於訓練，剩下的 1-fold 用於評估結果。並且使用平均精確度(Mean Average Precision, MAP)來進行效能評估。在資料前處理方面，我們會先將查詢文件和檢索文件做 Porter 詞幹提取處理，詞幹處理是去除詞綴得到單詞最一般寫法的過程，目的是讓整個詞庫不會很龐大。在參數設置方面，最佳匹配模型 25 的超參數我們設定 K_1 為0.8、 b 為0.75和 K_3 為1000；深層結構語義模型的批次大小設定為 4，採用適應性矩估計演算法優化器，學習率設定為 0.00001，訓練 100 個世代(Epoch)。在本研究提出的方法 MAML-DSSM 中，我們對適應性矩估計演算法使用單個梯度步長訓練模型，步長大小設定為0.00001，元批次大小為 32 個任務，並使用隨機梯度下降優化器進行十個梯度步長($\alpha = 0.001$)來評估內部模型，迭代次數設定為 1000。對於 N 類別 K 樣本的設置，我們採用每個任務會有 10 個類別，每個類別使用 5 筆資料，所以每個梯度都是由 50 筆資料來計算，並且每個類別也會有 5 筆資料來評估更新後的元梯度。以上實驗參數為多次實驗後調整所得到。

(三)實驗結果

首先在第一組實驗中，我們使用三組基礎方法來進行比較，分別是向量空間模型(VSM)和最佳匹配模型 25(BM25)，以及使用類神經網路架構的深層結構語義模型(DSSM)，實驗結果如表一所示，我們可以發現深層結構語義模型可以獲得最好的成績。接著，我們亦比較這些基礎系統與本研究所提出之 MAML-DSSM，實驗結果同樣如表一所示。經過實驗證實，我們提出的方法相對於基礎方法來說有明顯的改進，並且在評估指標中，

Model		MAP
(a)	VSM	0.1297
(b)	BM25	0.2409
(c)	DSSM	0.2502
(d)	MAML-DSSM	0.2584

表一、基礎方法與本研究提出方法之實驗結果。

Model		MAP
(a)	MAML-DSSM	0.2584
(b)	(a)+ Finetune	0.2595

表二、本研究提出方法加上微調之實驗結果。

Model	MAP	
	1-shot	5-shot
MAML-DSSM	0.2601	0.2584

表三、不同設定在 MAML-DSSM 中的實驗結果。

相對於深層結構語義模型有 3.1%的效能提升。由此可知，在資訊檢索中使用元學習方法確實能讓模型訓練的更好。接著，我們探討對模型進行微調的成效，所以我們使用經過最佳匹配模型 25 檢索後的前 10 篇文件對 MAML-DSSM 再進行微調，實驗結果如表二所示，進行微調動作有助於提升資訊檢索的表現。最後，為了分析訓練樣本多寡所帶來的差異，我們分別在 10 類別 1 樣本(10-way 1-shot)和 10 類別 5 樣本(10-way 5-shot)的設置下進行訓練，其實驗結果如表三所示。我們可以發現通常使用較多的樣本進行訓練，所帶來的結果會比較好，但是在本實驗中的結果卻是少樣本的效能比較好，這一部分值得進行更深一步探討。

五、結論

本研究提出使用元學習方法應用在資訊檢索領域上。經過實驗結果顯示，我們所提出之結合與模型無關的元學習演算法與深層結構語義模型在資訊檢索任務上，可以獲得一定的任務成效提升。值得一提的是，在本研究中，我們只使用深層結構語義模型架構的模型來進行訓練，但是基於與模型無關的元學習演算法與模型無關的特性，這種方法可以應用在更多不同類型的網路架構，所以在未來的研究中，我們會將此方法應用在資訊檢索中不同的模型上面，並使用不同的資料集來評估其有效性。

致謝

This work is supported by the Ministry of Science and Technology (MOST) in Taiwan under grant MOST 109-2636-E-011-007 (Young Scholar Fellowship Program), and by the Project K367B83100 (ITRI) under the sponsorship of the Ministry of Economic Affairs, Taiwan.

參考文獻

- [1] A. A. Rusu *et al.*, "Meta-Learning with Latent Embedding Optimization," in *International Conference on Learning Representations*, 2018.
- [2] J. Gu, Y. Wang, Y. Chen, V. O. Li, and K. Cho, "Meta-Learning for Low-Resource Neural Machine Translation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 3622-3631.
- [3] J.-Y. Hsu, Y.-J. Chen, and H.-y. Lee, "Meta learning for end-to-end low-resource speech recognition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 7844-7848: IEEE.
- [4] G. I. Winata *et al.*, "Learning fast adaptation on cross-accented speech recognition," 2020.
- [5] F. Mi, M. Huang, J. Zhang, and B. Faltings, "Meta-learning for low-resource natural language generation in task-oriented dialogue systems," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 3151-3157: AAAI Press.
- [6] O. Klejch, J. Fainberg, and P. J. a. p. a. Bell, "Learning to adapt: a meta-learning approach for speaker adaptation," 2018.
- [7] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, 2015, vol. 2: Lille.
- [8] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Advances in neural information processing systems*, 2017, pp. 4077-4087.
- [9] O. Vinyals, C. Blundell, T. Lillicrap, and D. Wierstra, "Matching networks for one shot learning," in *Advances in neural information processing systems*, 2016, pp. 3630-3638.
- [10] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1199-1208.
- [11] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," 2016.
- [12] A. Nichol, J. Achiam, and J. J. a. p. a. Schulman, "On first-order meta-learning algorithms," 2018.
- [13] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of

- deep networks," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 2017, pp. 1126-1135.
- [14] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 2333-2338.