Computational Linguistics and Chinese Language Processing
Vol. 22, No. 2, December 2017, pp. 1-16

語音文件檢索使用類神經網路技術

On the Use of Neural Network Modeling Techniques for Spoken Document Retrieval

羅天宏*、陳映文*、陳冠宇+、王新民#、陳柏琳*

Tien-Hong Lo, Ying-Wen Chen, Kuan-Yu Chen,

Hsin-Min Wang and Berlin Chen

摘要

近年來由於含有語音資訊的多媒體內涵不斷增長，語音文件檢索已成為一個相當熱門的議題並吸引許多學者與實務家的投入研究。除了發展強健的索引機制和有效的檢索模型外，如何正確地且有效率地對於查詢內容進行模型化對於增進語音文件檢索的表現也扮演著非常關鍵的角色。有鑒於此，在本論文，我們提出一個新穎的基於類神經網路之相關性感知模型來得到較佳的查詢表示方式，同時可以避免使用傳統較耗費時間的準相關回饋程序。再者，我們嘗試將查詢意向分類的概念融入我們所提出的模型架構中，以進一步獲取更精緻的查詢表示方式。在 TDT-2 語音文件及所進行的初步實驗顯示出本論文所提出方法的效用。

關鍵詞：語音文件檢索，查詢意向，類神經網路，準相關回饋

* 國立臺灣師範大學資訊工程學系
Department of Computer Science & Information Engineering, National Taiwan Normal University
E-mail: {teinhonglo, cliffchen, berlin}@ntnu.edu.tw
+ 國立臺灣科技大學資訊工程系
Department of Computer Science & Information Engineering, National Taiwan University of Science and Technology
E-mail: kychen @mail.ntust.edu.tw
中央研究院資訊科學研究所
Institute of Information Science, Academia Sinica
E-mail: whm@iis.sinica.edu.tw

Abstract

Due to ever-increasing amounts of publicly available multimedia associated with speech information, spoken document retrieval (SDR) has been an active area of research that captures significant interest from both academic and industrial communities. Beyond the continuing effort in the development of robust indexing and effective retrieval methods to quantify the relevance degree between a pair of query and spoken document, how to accurately and efficiently model the query content plays a vital role for improving SDR performance. In view of this, we present in this paper a novel neural relevance-aware model (NRM) to infer an enhanced query representation, extricating the conventional time-consuming pseudo-relevance feedback (PRF) process. In addition, we incorporate the notion of query intent classification into our proposed NRM modeling framework to obtain more sophisticated query representations. Preliminary experiments conducted on the TDT-2 collection confirm the utility of our methods in relation to a few state-of-the-art ones.

Keyword ： Spoken Document Retrieval, Query Intent, Neural Network, Pseudo-Relevance Feedback

1. 緒論 (INTRODUCTION)

伴隨著網際網路的發展與多媒體資訊的大量增長，影音的瀏覽與傳遞也逐漸成為我們的日常生活的一部分。在這環境下，如何利用語音的資訊，快速檢索符合資訊需求的內容，變成了一項新興的需求。因此，在過去的二十年(Chelba, Hazen & Saraclar, 2008) (Lee & Chen, 2005) (Huang, Ma, Li & Wu, 2011) (Chen, Chen, Chen & Chen, 2012)，語音文件檢索成為一個十分有魅力的研究主題。在語音文件檢索的任務上，過往有許多顯著成功的方法，如向量空間模型(Vector Space Model) (Salton, Wong & Yang, 1975)、Okapi BM25 model (Jones, Walker & Robertson, 2000)，以及主題模型(Topic Model) (Blei, Ng & Jordan, 2003)等。另一方面，將統計式語言模型(Statistical Language Model)應用在文字檢索(Information Retrieval)和語音文件檢索，在檢索任務上取得了嶄新的突破(Ponte & Croft, 1998) (Song & Croft, 1999) (Croft & Lafferty, 2003)，因此吸引了不少研究者的目光。在這樣的概念下，查詢對每個文件計算似然機率後作排名，我們稱這樣的排序方法為查詢似然測量(Query Likelihood Model Measure, QLM) (Manning, Raghavan & Schutze, 2008)。另一個知名的評估方式為 KL 散度測量(Kullback-Leibler Divergence Measure, KLM) (Zhai & Lafferty, 2001)，將查詢與文件皆表示為單元語法的語言模型(Unigram Language Model)，查詢與文件的相似程度即為兩個機率分佈的散度距離(Divergence Distance)。

最近，隨著深層類神經網路架構的流行，這類的方法也被大量應用在檢索的任務上。主要的研究方向為利用不同網路架構與訓練方法，以此來學習查詢與文件間的相似關係(Guo, Fan, Ai & Croft, 2016) (Mitra, Diaz & Craswell, 2017)。值得注意的是，大部分方法

的輸入資料皆是如詞頻的簡單統計資訊(Surface Statistics)，且期望類神經網路能夠區分出相關與不相關的文件。基於以上的想法，有兩個較為知名的方法，分別是深層結構語義模型(Deep Structured Semantic Model, DSSM) (Huang *et al*., 2013)和局部特徵向量模型(Locality Preserving Essence Vector, LPEV) (Chen, Liu, Chen & Wang, 2017)。DSSM 和 LPEV 兩種方法也皆致力於在非監督的環境下，將查詢與文件分別以低維度的向量作表示，使得向量不是單純的文字序列或簡單的統計關係(如鄰近的詞、共同出現的詞等更豐富的資訊)。除了以上兩種檢索模型，還有兩種較為知名的嵌入方法，分別是詞嵌入(word embedding-based methods)和段嵌入(paragraph embedding methods)的方法，前者的方法有略詞模型(Skip-gram model) (Mikolov, Sutskever, Chen, Corrado & Dean, 2013)和連續詞袋(Continuous bag-of-words) (Mikolov, Chen, Corrado & Dean, 2013)。後者的方法有分佈式內存模型(the distributed memory model) (Le & Mikolov, 2014)、分佈式詞袋模型(distributed bag-of-words model) (Le & Mikolov, 2014) (Chen, Lee, Wang, Chen & Chen, 2014)和本質向量模型(essence vector model) (Chen, Liu, Chen & Wang, 2016)等。

為了提升檢索效能，過往有許多方法嘗試猜測使用者意向進行查詢分類 (Shen *et al*., 2006)，給予對不同類別有興趣的使用者提供更為精確的結果。然而，如同傳統文字文件檢索，語音文件檢索也面臨了查詢過於簡短語意不清，且會隨著時間推移改變語句的意思，難以表達出使用者的資訊需求，因此查詢分類往往比文件分類困難。針對查詢語意不足的問題，查詢分類被定義為對使用者行為的建模，其一便是豐富語意表示的任務，稱為擴增查詢(Query Expansion)，擴增查詢主要可以分為兩個類別，第一種為引用外部的資源(如 Wikipedia 或 WordNet)分析語意，進一步擴展原始的查詢，其中的語意關係包括同義詞、反義詞、多義詞等；第二種為分析查詢的回饋，給予一個查詢，追蹤使用者點擊的文件，並以此做分類，以求第二次更為精確的結果，與非監督式的準相關回饋(Pseudo-Relevance Feedback) (Zhai & Lafferty, 2001) (Lavrenko & Croft, 2001) (Chen, Liu, Chen, Wang & Chen, 2016)的精神相似。第一個方法需要較複雜的自然語言處理技術，如語意表徵和推論。第二種方法則較為簡單，只需要取得前幾篇文章做分析，再適當地與原始查詢做結合即可。且因為分析的資料僅為回饋的文件，準相關回饋不需要額外的語料庫來學習。以下為幾個知名的準相關回饋(Manning *et al*., 2008)，相關性模型(Relevance Model, RM) (Lavrenko & Croft, 2001) 、簡易混合模型(Simple Mixture Model, SMM) (Zhai & Lafferty, 2001)、顯著詞模型(Significant Word Model, SWM) (Dehghani, Azarbonyad, Kamps, Hiemstra & Marx, 2016) (Chen, Chen, Wang & Chen, 2017)。儘管準相關回饋已經在檢索的領域上證實有效性，但仍需要做二次查詢，造成缺乏即時性的問題，因此實務上難以採用。

我們的動機便是解決上述提到的重要問題，有效性與即時性。因此這篇論文致力提出一個基於查詢分類的擴增查詢架構，其中利用到以類神經網路架構為基礎的 NRM，並探討了 NRM 利用查詢意向探索的可能性。

2. 相關文獻回顧 (RELATED WORK)

由於在現實中，查詢所用到的詞並不多，因此常用於估測查詢Q的詞出現機率的語言模型$P(w|Q)$的最大似然估測(Maximum Likelihood Estimator)便很難發揮所長。為了減緩這樣的侷限，有許多專家學者提出不同的方法，其中已經證實有效且泛用的方法便是準相關回饋，目的是用於估測出更為精確的查詢表示式。準相關回饋假設，查詢 Q 與檢索結果的前 N 篇文章相關，因此分析第一次查詢結果的文件$\mathbf{D}_F = \{D_1, \cdots, D_r, \cdots, D_{|\mathbf{D}_F|}\}$，來達到更為精確的查詢語言模型。我們探討的方法為以下三個，RM、SMM、SWM。

2.1 相關性模型(Relevance Model)

RM 假設每一個查詢Q皆有一個相關類別R_Q，且每個與查詢Q相關的文件皆由相關類別R_Q中產生。不幸的是，我們不會知道查詢Q的相關類別R_Q，因此作為替代，我們會將前 N 篇回饋的文章$\mathbf{D}_F$當作相關文章，並且利用準相關回饋近似真實的相關類別R_Q。對應查詢Q的相關性模型可利用以下公式計算：

$$P_{\mathrm{RM}}(w|Q) = \frac{\sum_{D_r \in \mathbf{D}_F} P(D_r) P(w|D_r) \prod_{w' \in Q} P(w'|D_r)}{\sum_{D'_r \in \mathbf{D}_F} P(D'_r) \prod_{w'' \in Q} P(w''|D'_r)}, \tag{1}$$

其中$P(D_r)$為文件的產生機率。由於我們沒有對於文件的先驗知識，因此決定機率的方式是用均勻分佈(Uniform Distribution)來實現。另一個語言模型$P(w|D_r)$則是利用最大似然估計的方式，利用文件的詞頻和該文件與查詢相似程度，計算出每個詞出現在文件D_r的機率，其餘符號以此類推。

2.2 簡易混合模型(Simple Mixture Model)

另一個估測查詢語言模型的觀點是 SMM。SMM 假設在回饋文件$\mathbf{D}_F$裡詞出現機率不是由單一模型估測，而是來自兩部份混合而成的語言模型，第一部份是專屬於該查詢Q的特殊詞的主題模型$P_{\mathrm{SMM}}(w|Q)$；第二部份是廣泛出現在各個文件中的背景語言模型$P_{\mathrm{BG}}(w)$。如此一來，便可透過分析回饋文件$\mathbf{D}_F$，最大化 SMM 的似然機率，並求得$P_{\mathrm{SMM}}(w|Q)$，以下為估測時使用的損失函數(loss function)：

$$L = \prod_{D \in \boldsymbol{D}_F} \prod_{w \in V} [\alpha \cdot P_{\mathrm{SMM}}(w|Q) + (1-\alpha) \cdot P_{\mathrm{BG}}(w)]^{c(w,D)}, \tag{2}$$

α 為預先定義好的參數，用於決定$P_{\mathrm{SMM}}(w|Q)$和$P_{\mathrm{BG}}(w)$兩者的比重關係。這樣的估計方式讓針對查詢Q的特殊詞得到較高的機率，進而獲得一個更有效的查詢模型。SMM 的假設提供一個有益的訊息，便是在回饋文件$\mathbf{D}_F$出現的詞，不僅與查詢Q本身相關，也與不存在於回饋文件$\mathbf{D}_F$的外部文件有關。舉例來說，背景詞儘管在特定查詢的回饋文件出現機率很高，但在其他文件的出現機率一樣很高，這樣詞的特徵便會被背景語言模型$P_{\mathrm{BG}}(w)$給捕捉。另一類詞只在特定查詢的回饋文件的出現機率高，那麼這樣的特徵便被特殊詞的主題模型$P_{\mathrm{SMM}}(w|Q)$捕捉。

2.3 顯著詞模型(Significant Word Model)

這樣的想法啟發自 Luhn's theory (Luhn, 1958)和 SMM，SWM 發現更精確估測查詢語言模型的方法。在 SWM 裡面，假設實際對於查詢有幫助的詞，必須不能是每個文件皆可看到的背景詞，且也不能過於集中出現在少數的回饋文件$\mathbf{D}_F$。因此 SWM 假設回饋文件的語言模型由下列三個模型混合而成，第一個模型為背景語言模型$P_{\mathrm{BG}}(w)$；第二個模型為特殊詞的語言模型$P_{\mathrm{S}}(w|Q)$，以及第三個學習到的顯著詞模型$P_{\mathrm{SW}}(w|Q)$，利用上述三個模型估測回饋文件$\mathbf{D}_F$的公式如下：

$$P(w|D) = \alpha \cdot P_{\mathrm{BG}}(w) + \beta \cdot P_{\mathrm{S}}(w|Q) + (1-\alpha-\beta) \cdot P_{\mathrm{SW}}(w|Q), \quad (3)$$

α 和 β 為可調整的參數，用來決定$P_{\mathrm{BG}}(w)$、$P_{\mathrm{S}}(w|Q)$，以及$P_{\mathrm{SW}}(w|Q)$三者對於語言模型$P(w|D)$的貢獻。$P_{\mathrm{BG}}(w)$為表示常見詞的模型，估測方式是在全部文件集合中，詞的出現次數；$P_{\mathrm{S}}(w|Q)$為表示太特殊的詞的模型，估測方式是在回饋文件$\mathbf{D}_F$中，僅集中出現在少數特定文件的特殊詞。$P_{\mathrm{SW}}(w|Q)$則是既不是常見詞，也不是太特定的詞，因此$P_{\mathrm{SW}}(w|Q)$是利用上述兩個模型，便可在回饋文件$\mathbf{D}_F$中估測出$P_{\mathrm{SW}}(w|Q)$的最大似然機率，並使用最大期望算法(Expectation-Maximum Algorithm) (Dempster, Laird & Rubin, 1977)調整參數。

3. 查詢意向與建模方法 (QUERY INTENT AND MODELING FRAMEWORK)

儘管先前的準相關回饋在資訊檢索和語音資訊檢索上，皆大幅提升原先不足的查詢效能，但卻無法應用在實際的檢索系統上。最大的原因是準相關回饋需要做第二次查詢，消耗過多的計算時間(Manning *et al.*, 2008)。因此針對耗時的問題，我們利用基於類神經網路技術的神經相關感知模型(NRM)的架構。在這樣的架構下，我們不僅能夠有效地重新建構一個更有效的查詢表示式，同時還能解決準相關回饋的耗時問題。

3.1 建立查詢語言模型的相關性 (MODELING RELEVANCE FOR QUERY)

在許多擴增查詢的方法定義了不同的查詢相關性。RM 用了系統性的方法去近似相關性模型；SMM 和 SWM 則是分別利用背景語言模型(Background Language Model)以及額外的特殊詞模型(Specific Word Model)來獲得回饋文件的相關性。在這裡的研究，我們是利用類神經網路的技術來學習上述的建模過程。

更詳細的說，給定一個查詢的集合$\mathbf{Q} = \{Q_1, \cdots, Q_t, \cdots, Q_T\}$，每一個在集合裡的查詢會分別對應到查詢與文件的相關資訊$\mathbf{R} = \{R_1, \cdots, R_t, \cdots, R_T\}$。為了解決查詢的長度不一的問題，我們首先會將每一個查詢用高維度的詞袋模型$P_{Q_t} \in \mathbb{R}^{|V|}$來表示，其中$P_{Q_t}$為出現在對應查詢$Q_t$裡的詞次數，$|V|$則是語料庫的詞典長度。首先將$P_{Q_t}$正規化，讓向量裡的值合計為一，接著再利用編碼器$f(\cdot)$將原始查詢降至低維度空間，如下所示：

$$f(P_{Q_t}) = v_{Q_t} \quad (4)$$

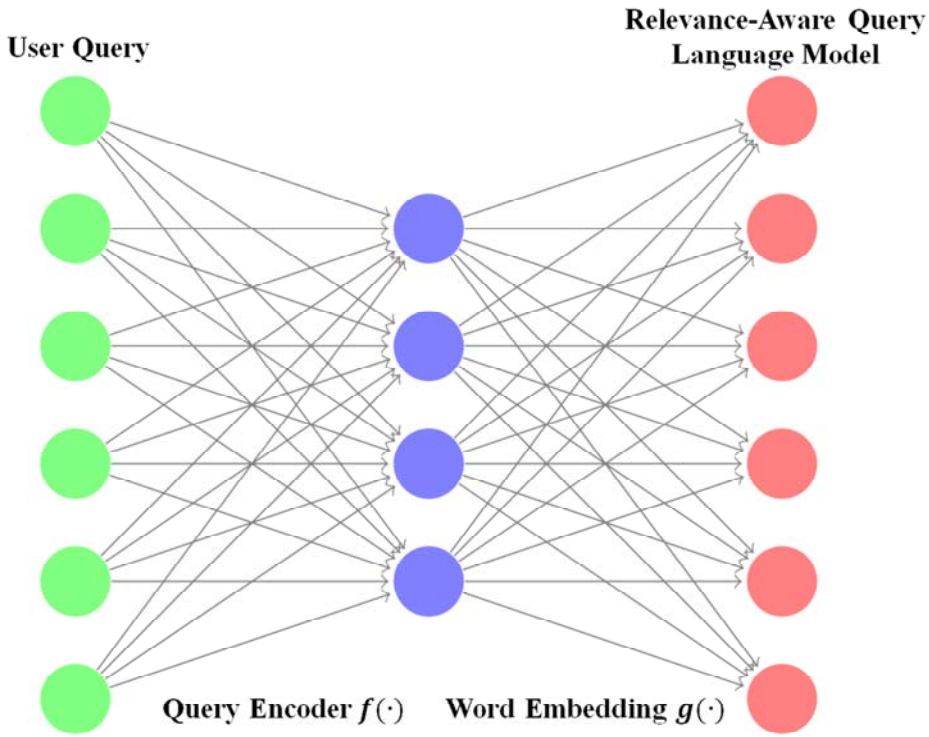

圖 1. 前饋式神經網路的 NRM 模型架構
[Figure 1. NRM framework with feed-forward neural network]

在這篇論文裡面，$f(\cdot)$為全連接的前饋神經網路(feed-forward fully-connected neural network)。以最終要增加檢索效果的查詢表示式為目標，直覺的想法便是先推論一組查詢與詞嵌入(word embedding)的表示方式，再重新構建一個新的查詢語言模型。為了做到這點，我們在$f(\cdot)$之上再堆疊一個解碼器$g(\cdot)$，$g(\cdot)$是一個全連接的前饋式神經網路，權重矩陣可表示成$\mathbf{W} \in \mathbb{R}^{k\times|V|}$。其中$k$為的查詢的大小，$|V|$為詞典的大小。神經相關感知模型(NRM)可表示為以下的式子：

$$P_{\text{NRM}}(w|Q_t) = g\left(f\left(P_{Q_t}\right)\right) = \frac{\exp(v_{Q_t}\cdot v_w)}{\sum_{w\prime\in V}\exp(v_{Q_t}\cdot v_{w\prime})} \tag{5}$$

其中v_w是在矩陣$\mathbf{W}$的第w行，也是詞w的詞嵌入表示。最後，為了捕捉到特定查詢的相關性分佈，我們設計了一個訓練目標，最佳解便是在訓練資料上搜尋到一組最大化似然機率的模型參數，調整公式如下：

$$L = \prod_{t=1}^{\mathrm{T}}\prod_{w\in V} P(w|R_t) log P_{\text{NRM}}(w|Q_t) \tag{6}$$

其中$P(w|R_t)$為表示語查詢Q_t對應的相關性分佈。總結這個章節的內容， NRM 的架構主要可分為兩個部分，分別是推斷低維度表示的編碼器$f(\cdot)$和表示相關性的詞嵌入與任意單詞的對應矩陣$\mathbf{W}$，並以最大化似然機率的準則調整模型參數。

3.2 使用者查詢意向 (QUERY INTENT)

為了瞭解使用者的查詢意向(Query Intent)，我們使用基於高維度詞袋模型P_{Q_t}做為分群的依據，使用的演算法為 K-means。首先，我們將P_{Q_t}正規化，讓向量裡的每個值加總為一。接著以查詢詞的出現多寡做為分群的依據，迭代至收斂則停止，這時視為找出使用者意圖(Query Intent)，並當作特定的查詢意向訓練類神經網路$P_{\mathrm{NRM}}(w|Q_t)$。在測試階段便不再分群，而是將測試查詢的詞袋模型與各集群的中心點做相似度計算，以最像的 NRM 預測，其中i為最相似的集群，max相似度的計算以該查詢Q_t和集群中心c_i的 KL 散度距離。可以用下列的式子表示：

$$P_{NRM}(w|Q_t) = \max_{i \in c} P_{NRM_i}(P_{Q_t}) \tag{7}$$

利用使用者查詢意向預測出相關性模型後，再和原始查詢與模型線性組合得到更有鑑別力的查詢模型。詳細組合的式子如下：

$$P_{Q_t'} = \alpha P_{Q_t} + (1-\alpha) P_{NRM}(w|Q_t) \tag{8}$$

3.3 實作細節 (IMPLEMENTATION DETAILS)

在語音資訊檢索和資訊檢索的領域裡面，根據使用者的檢索行為，最直覺的定義相關性的方式便是與使用者的資訊需求有關，系統必須從使用者的查詢和語料庫的文件中找出相關性。在訓練階段，為了萃取出查詢與相關文件的相關性分佈(圖 2 的藍色部份)，並用這相關性的分佈來訓練 NRM，我們設計兩個不一樣的情境。第一個情境是監督式的環

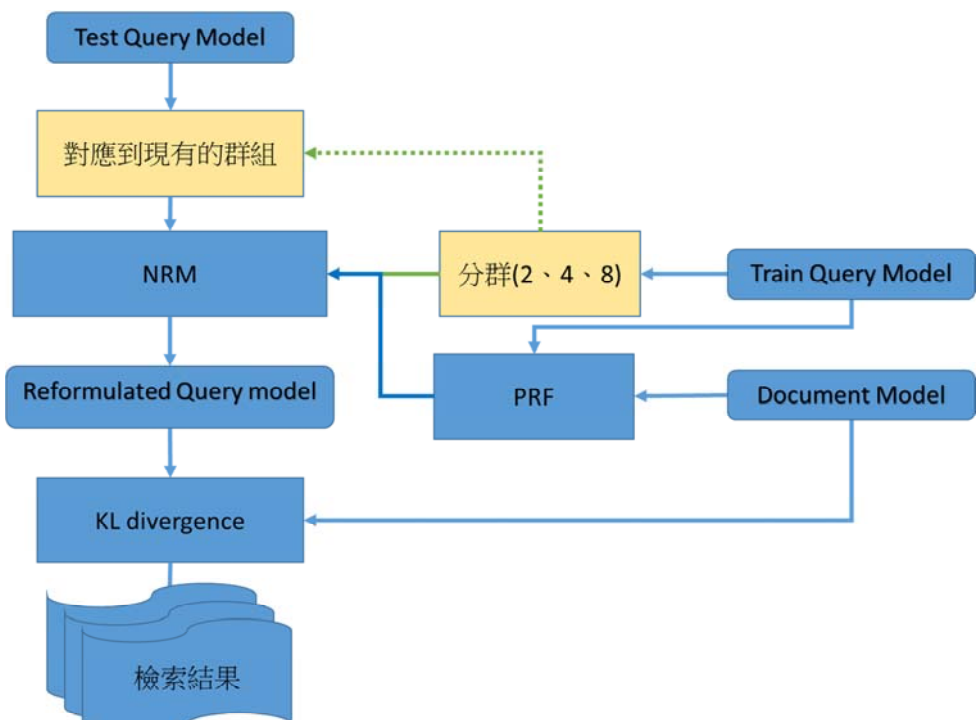

圖 2. 整體架構圖。藍色部份為檢索過程，黃色部分為查詢意向

[Figure 2. Model architecture. The modeling and retrieval is blue block, while yellow part is query intent.]

境，假設已知查詢與相關文件，如給定一個查詢，便有相關文件的正確答案。查詢與文件相關性可以用使用者的看到檢索結果後，點擊相關文件的資訊來代表。儘管點擊資訊可以反映出這篇文件與查詢是否相關，但要收集到這些正確資訊，本身就是一個浩大的工程。因此，我們提出第二種非監督式的情境，原先第一種情境的假設不存在，我們事前不知道相關文件與查詢的對應，只有一個多個查詢構成的集合，與不知道是否相關的文件集合。這樣的情境下，最自然的方法便經過一次查詢，得到相關性的排行後，取出前幾篇當作(準)相關的文件，這樣的策略就如同準相關回饋的方式。經過以上兩個情境的處理後，在訓練的階段，我們便有成對的查詢與相關文件，接著再利用準相關回饋處理這份資料，從中獲得每個查詢的相關性分佈$P(w|R)$，可利用擴展查詢的語言模型建模方法(如，RM、SMM、SWM)獲得。使用在 NRM 架構的神經網路配置如下，隱藏層的激勵函數為線性函數(linear function)，輸出層為 Softmax 函數(softmax function)，其中我們利用 Adam (Kingma & Ba, 2015)演算法尋找最佳解。在測試階段，每一個輸入的查詢皆會被表示為高維度的詞袋模型，經過一層編碼器$f(\cdot)$，將原先高維度表示的向量轉成低維度的詞嵌入向量，並保有原先資料點在高維度的關係，經過一層解碼器$g(\cdot)$，將原先被編碼成低維度表示的詞嵌入向量，依序解碼成 NRM 查詢語言模型，損失函數為交叉熵(Cross Entropy)。取得一個新的查詢語言模型後，我們再利用現有的 KLM 計算 NRM 查詢語言模型與文件語言模型的 KL 散度距離(divergence distance)。總結目前提到的好處，我們利用 NRM 的技術離線學習準相關回饋的建模方式，並在線直接推論新的表示方式。線上查詢時，系統可從原始缺乏使用者的資訊需求的簡短查詢中，重新建構出一個具有準相關回饋效能的的查詢語言模型，也因為耗時的處理過程皆在離線時做完，因此同時也解決耗時的缺點。

最後是建立在 NRM 的基礎之上，針對不同主題的訓練方式(圖 2 的黃色部分)，實作可分為兩個階段，首先是訓練階段，先利用 K-means 演算法將查詢分為 2、4、8 群，視為將查詢分成 2、4、8 個主題類別，接著將每一群當作訓練資料，分別訓練不同的 NRM。第二部分為測試階段，依據在訓練階段分好的 2、4、8 群，測試的查詢以 KL 散度距離計算該歸類在那群，並以該群訓練好的 NRM 模型預測新的查詢表示式，並進行檢索取得相關文件。

4. 實驗設定與結果 (EXPERIMENTS)

4.1 實驗設定 (EXPERIMENTAL SETUP)

我們使用 TDT2 (Topic Detection and Tracking collection)作為實驗數據 (Linguistic Data Consortium, 2000)。來自美國之聲的新聞廣播(Voice of America news broadcasts)的國語新聞(Mandarin news stories)被用來當作語音文件。所有的新聞故事都被標記上特定主題，以此作為評估效能時的相關與否。語音文件的平均詞錯誤率(Word Error Rate, WER)為 35% (Meng *et al.*, 2004)。測試時用來自新華社(Xinhua News Agency)的國語新聞當作檢索的查詢。更精確點來說，我們測試用的查詢可以分為兩個類別，使用新聞內容的長查詢

和使用新聞標題的短查詢。表 1 為 TDT2 一些基本的統計數據。評估檢索效能的方式，我們選用非內差的平均精確度 (non-interpolated mean average precision, MAP) (Manning *et al.*, 2008) (Baeza-Yates & Ribeiro-Neto, 2011)作為評判尺度。

表 1. TDT-2 的統計資訊
[Table 1. Statistics of the TDT2 collection]

# 語音文件	2,265 新聞, 46.03 小時的語音錄音			
# 測試用查詢	16 新華社新聞 (Topics 20001~20096)			
	最短	最長	中位數	平均
文件長度	23	4,841	153	287.1
短查詢的長度	8	27	13	14.0
長查詢的長度	183	2,623	329	532.9
# 測試用查詢的相關文件	2	95	13	29.3

4.2 實驗結果 (EXPERIMENTAL RESULTS)

首先， 我們探討 NRM 在已知相關性和未知相關性兩種不同的情境的表現，實驗結果呈現於圖 3 與圖 4。比較的方法為向量空間式的深層結構語意模型(DSSM) (Huang *et al.*, 2013)和局部保留本質向量模型(LPEV) (Chen *et al.*, 2017)。DSSM 使用點擊資訊當作相關性，訓練網路的參數使相關文件和查詢的似然機率最大化。另一方面，LPEV 則意旨學習一個更好的低維度表示空間，同時保留原始語意結構。

由圖 3 中呈現的數據來看，有幾個比較明顯的結果。首先，DSSM 不論是在人工轉錄還是自動語音轉錄的文件，皆優於 LPEV。原因是在於 DSSM 是利用點擊資訊代表相關性，目的是訓練網路正確地分辨相關和不相關的文件。LPEV 則是意旨在學習一個文件和查詢的特殊的表示空間。目的性的不同，也使得 LPEV 在檢索結果上，比起 DSSM 差強人意。其次，因為在先前的實驗中，SWM 本身的表現普遍比 RM、SMM 較為優異。因此將這樣的方法使用在 NRM 的架構中，其中結合 SWM 和 NRM 的方法，我們將表示為 NRM(SWM)。評估結果正如我們的預期，一樣是 SWM 的表現較為穩定。我們發現在 NRM 的架構之下，不論傳統向量空間模型式的方法，或是經典的語言模型式的方法，NRM 的檢索結果效能普遍都能明顯勝出。有趣的是，NRM 方法甚至能勝過現有的重構查詢的方法，如 RM、SMM、SWM。這樣的好結果或許得歸功於離線的計算過程。相較於 RM、SMM、SWM 利用一次查詢的準回饋，線上做準相關回饋的計算，NRM 在離線時利用點擊資訊，並透過 RM、SMM、SWM 捕捉相關性分佈。我們認為這不同之處，可能讓 NRM 可以學習到有效的特徵,並利用這些特徵在線上即時取得更好的查詢結果，同時也解決耗時的問題。最後，我們可以明顯觀察出，在 NRM 架構之下，不論是那一

個方法，在所有的情況下皆大幅勝出 LPEV，以及在大部分的情況下贏過 DSSM。以上的結果進一步地證明 NRM 在語音文件檢索的有效性。

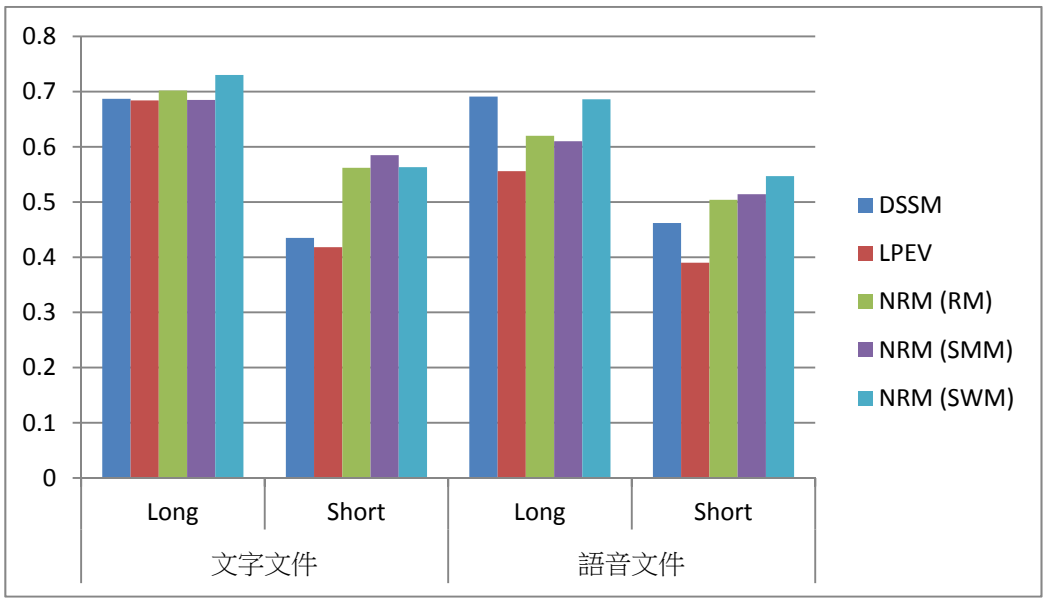

圖 3. 離線時利用點擊資訊訓練的 NRM 模型之檢索結果
[Figure 3. Retrieval results of the NRM offline trained on click-through information.]

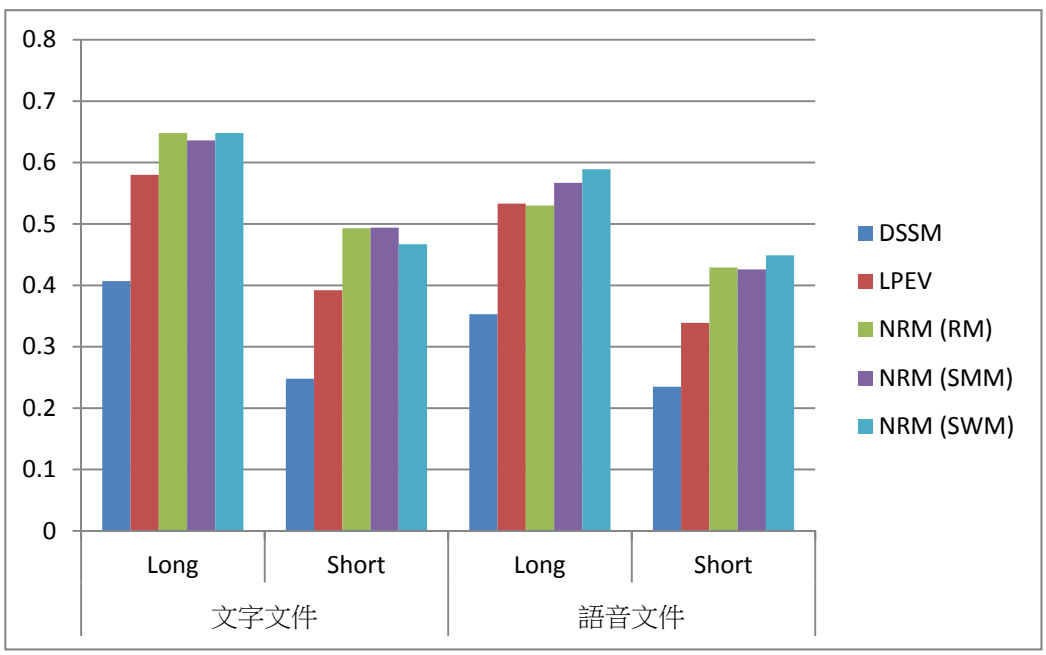

圖 4. 離線時利用準相關回饋訓練的 NRM 模型之檢索結果
[Figure 4. Retrieval results of the NRM trained with pseudo relevance feedback.]

接著是我們假設的第二個情境，查詢與相關文件的關係是未知。在這次的實驗中，我們查詢結果的前 10 篇當作相關文件，因此相關文件未必真的正確。我們可以從結果中觀察到幾個現象。首先，與表 2 不同，LPEV 在這次的實驗中，表現勝過 DSSM。因為 LPEV 的目標與 DSSM 不同，也因此辨識錯誤的相關文件的影響也較小，表現較為穩健。在語音文件的部分，SWM 的表現依舊可以勝過 SMM 和 RM，僅在手寫文本的部分略微下降。最後，整體來看，比較圖 3 與圖 4，這次的實驗普遍表現較差，可以觀察到正確答案對於以上幾個方法的影響，以 NRM 的結果來看，可以發現 NRM 非常依賴相關文件的信息是否正確。最後，即使是在非監督的環境下，NRM 仍然可在各種情境下，表現比 LPEV、DSSM 優異，再一次證明 NRM 的優秀之處。

表 2. 基於以上 NRM 之上，進一步利用使用者意向資訊
[Table 2. Based on NRM framework and further use query intent information.]

	點擊資訊				準相關回饋			
	文字文件		語音文件		文字文件		語音文件	
	Long	Short	Long	Short	Long	Short	Long	Short
NRM SWM	**0.730**	0.563	**0.686**	**0.547**	**0.648**	0.467	0.589	0.449
NRM SWM)-2	0.690	**0.571**	0.670	0.544	0.636	**0.475**	**0.593**	**0.470**
NRM(SWM)-4	0.694	0.562-	0.669	0.545	0.628	0.462	0.583	0.434
NRM(SWM)-8	0.712	0.564-	0.672	**0.547**	0.632	0.463	**0.593**	0.437

分群的實驗部分，我們假設查詢之間是有不同類別的關係，因此建立在原先 NRM 的基礎之上，利用簡單的分群演算法將查詢分群，不同的類別就訓練新的 NRM 模型，期望能達到 NRM 能學習到不同主題的特徵，進一步提升學習的效果。這裡的實驗，我們採用的是在第一個及第二個場景下皆表現較為亮眼的 SWM 方法。首先，我們將訓練資料中的查詢分群，並依據不同的群訓練新的 NRM，測試的長(短)查詢會依據不同歸屬的群，決定使用那個 NRM 預測新的查詢表示式，並以此新的模型結合舊有預測出來的模型(圖 3 與圖 4)線性組合，以此做為新的查詢，實驗結果呈現在表 2。從實驗結果可以看出，不論是那一種的情境，在分群後的訓練結果，大部分的效能都是下降或持平，少數幾個狀況下表現得比原先的結果較好。平均來看，分兩群的效果勝過四群與八群，四群的效能經常落在分兩群與分八群之間，偶而分八群的效果會是最佳的，但卻不會贏過太多。這樣的實驗結果，可以視為 SWM 是較為複雜的語言模型，會根據不同的查詢有不同的影響，所以當我們只將訓練資料分成兩群，對個別訓練的網路來說，可以學習到的訓練資料較為多樣，因此網路較能學習到 SWM 的準相關回饋的特性。反之，切割越多的資料後，讓網路的效能則變得較差。儘管分兩群的效果普遍勝於其他的設定， 但整體的效能比起舊有的模型，大部分仍是持平或退步。

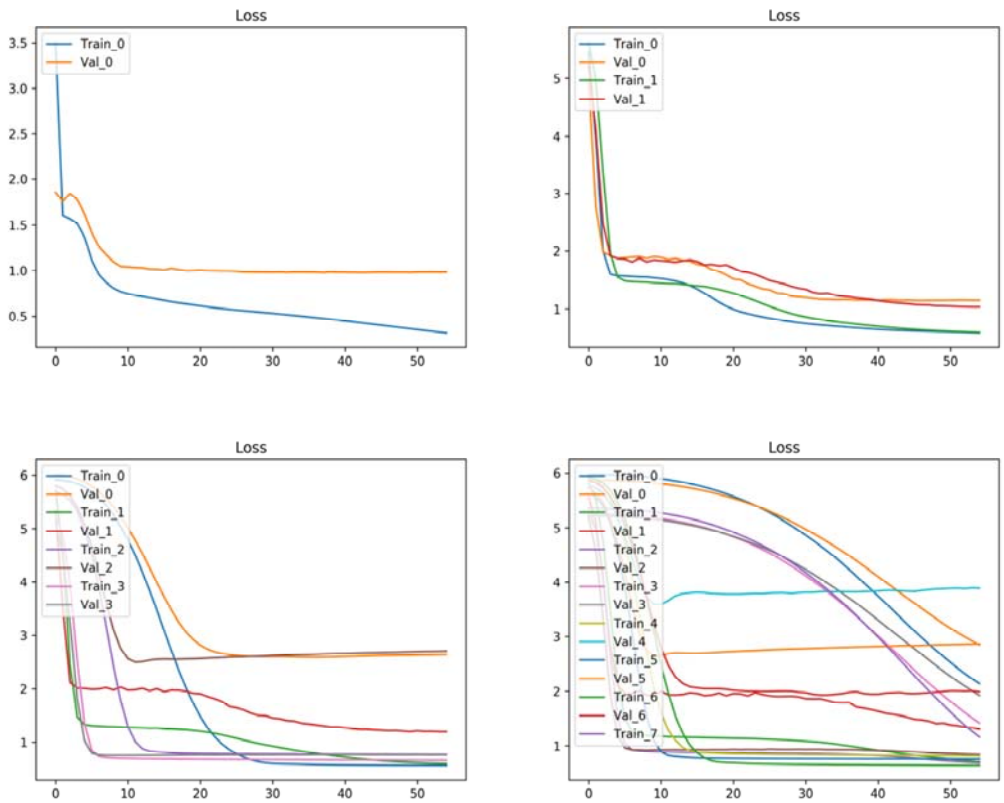

圖 5. 不同集群下，訓練集和驗證集的損失值
[Figure 5. Loss in training set and developing set with diffrernent clusters]

為了進一步研究這次分群對結果的影響，我們將網路訓練的損失值(loss)視覺化，記錄在圖 5。縱軸為損失值，橫軸為迭代次數。不同的線代表不同網路訓練時的兩個值，訓練集(Training Set)和驗證集(Validation Set)的損失值，由左圖到右圖，分別是不分群，以及二、四和八個集群的訓練結果，有些集群只分到一個查詢，集群大小為一就不切成訓練集和驗證集。從上圖中可以看出，分成二個集群的訓練過程較為穩定，訓練集和驗證集的損失值皆為穩定地下降。分成四個集群後，訓練情況有些不同，雖然訓練集的損失值仍是穩定下降，但驗證集的損失值中表現較差，尤其是第零群和第二群。分成八個集群後，這樣的現象就更為明顯，訓練集的損失值依舊穩定下降，但驗證集的損失值則沒有呈穩定的下降曲線，有些集群的損失值甚至有些上升，可以看出明顯地過度訓練(Overfitting)。除了過多的集群可能會造成過度訓練以外，我們也觀察到分越多集群，部分集群的表現(訓練集和驗證集)下降的比原先較少集群的多(如分成四個集群的第三個集群)，但這樣的表現並不是穩定的結果，可能是原先將查詢分群的依據是詞袋模型，過短的查詢造成語意不清，導致分群效果不彰，部分集群沒訓練好，連帶影響整體的成果。儘管這次實驗的結果不如預期，但我們依舊發現，透過簡單的查詢分類，可以更有效地

在 NRM 的架構下訓練模型，損失值可以降得比原先集群較少的狀況下更低(如分成四個集群的第三個集群)。因此用於捕捉使用者意向(Intent)的訓練方式，讓網路得到更多資訊的情況下，能更有效地訓練 NRM。總結先前所做的實驗，以上種種實驗揭示了一些訊息，不論是在語音文件檢索或資訊檢索的領域上，我們提出的 NRM 都可得到更好的結果，與最新穎的語言模型比較起來毫不遜色。

5. 結論 (CONCLUSIONS)

在這篇論文中，我們提出一個建立在 NRM 之上的查詢意向探索方法。在語音文件檢索的任務中，NRM 的方法能夠在不需要繁雜的準相關回饋的處理下，得到一個更有鑑別力的查詢語言模型，大幅提升檢索的效能。實驗的結果也證實，這樣的重構查詢語言模型的技術，比起過往的相關技術，檢索效能皆能穩定的勝出。儘管初步加入查詢意向的結果不盡理想，但實驗結果揭示仍有訓練 NRM 查詢意向的可能。未來的工作，我們希望能夠嘗試一些更複雜的類神經網路模型(如摺積神經網路(CNN)、遞歸神經網路(RNN)等)來作為 NRM 的骨幹。此外，我們也將嘗試加入一些新的特徵(如句法或韻律)，觀察網路獲得更多資訊的情況下能否增益學習效果。最後則是提供更複雜的查詢意向方法，以求更為細緻的查詢結果。

參考文獻 References

Baeza-Yates, R. & Ribeiro-Neto, B. (2011). *Modern information retrieval: the concepts and technology behind search*. Boston, MA: Addison-Wesley Professional.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(4-5), 993-1022.

Chelba, C., Hazen, T. J. & Saraclar, M. (2008). Retrieval and browsing of spoken content. *IEEE Signal Processing Magazine*, *25*(3), 39-49. doi: 10.1109/MSP.2008.917992

Chen, B., Chen, K.-Y., Chen, P.-N. & Chen, Y.-W. (2012). Spoken document retrieval with unsupervised query modeling techniques. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(9), 2602-2612. doi: 10.1109/TASL.2012.2208628

Chen, K. Y., Lee, H. S., Wang, H. M., Chen, B. & Chen, H. H. (2014). I-vector Based Language Modeling for Spoken Document Retrieval. In *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7083-7088. doi: 10.1109/ICASSP.2014.6854974

Chen, K. Y., Liu, S. H., Chen, B. & Wang, H. M. (2016). Learning to Distill: The Essence Vector Modeling Framework. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016)*, 358-368.

Chen, K. Y., Liu, S. H., Chen, B. & Wang, H. M. (2017). A locality-preserving essence vector modeling framework for spoken document retrieval. In *Proceedings of ICASSP 2017*, 5665-5669. doi: 10.1109/ICASSP.2017.7953241

Chen, K. Y., Liu, S. H., Chen, B., Wang, H. M. & Chen, H. H. (2016). Exploring the use of unsupervised query modeling techniques for speech recognition and summarization. *Speech Communication*, *80*, 49-59. doi: 10.1016/j.specom.2016.03.006

Chen, Y. W., Chen, K. Y., Wang, H. M. & Chen, B. (2017). Exploring the use of significant words language modeling for spoken document retrieval. In *Proceedings of INTERSPEECH 2017*. doi: 10.21437/Interspeech.2017-612

Croft, W. B. & Lafferty, J. (2003). *Language modeling for information retrieval*. Dordrecht, the Netherlands : Kluwer Academic Publishers. doi: 10.1007/978-94-017-0171-6

Dehghani, M., Azarbonyad, H., Kamps, J., Hiemstra, D. & Marx, M. (2016). Luhn revisited: significant words language models. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16)*, 1301-1310. doi: 10.1145/2983323.2983814

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, *39*(1), 1-38.

Guo, J.-F., Fan, Y., Ai, Q. & Croft, W. B. (2016). A deep relevance matching model for ad-hoc retrieval. In *Proceedings of CIKM '16*, 55-64. doi: 10.1145/2983323.2983769

Huang, P. S., He, X., Gao, J., Deng, L., Acero, A. & Heck, L. (2013). Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of CIKM '13*, 2333-2338. doi: 10.1145/2505515.2505665

Huang, C. L., Ma, B., Li, H. & Wu, C.-H. (2011). Speech indexing using semantic context inference. In *Proceedings of INTERSPEECH 2011*, 717-720.

Jones, K. S., Walker, S. & Robertson, S. E. (2000). A probabilistic model of information retrieval: development and comparative experiments (Parts 1 and 2). *Information Processing and Management*, *36*(6), 779-840. doi: 10.1016/S0306-4573(00)00015-7

Kingma, D. & Ba, J. (2015). ADAM: A method for stochastic optimization. In *Proceedings of ICLR 2015*.

Lavrenko, V. & Croft, W. B. (2001). Relevance based language models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '01)*, 120-127. doi: 10.1145/383952.383972

Linguistic Data Consortium. (2000). Project of Topic Detection and Tracking.

Le, Q. & Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of ICML '14*, 1188-1196.

Lee, L. S. & Chen, B. (2005). Spoken document understanding and organization. *IEEE Signal Processing Magazine*, *22*(5), 42-60. doi: 10.1109/MSP.2005.1511823

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, *2*(2), 159-165. doi: 10.1147/rd.22.0159

Manning, C. D., Raghavan, P. & Schutze, H. (2008). *Introduction to Information Retrieval*. New York, NY: Cambridge University Press.

Meng, H., Chen, B., Khudanpur, S., Levow, G.-A., Lo, W.-K., Oard, D., …Wang, J. (2004). Mandarin-English information (MEI): investigating translingual speech retrieval. *Computer Speech and Language*, *18*(2), 163-179. doi: 10.1016/j.csl.2003.09.003

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space. Retrieved from arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS '13*, 3111-3119.

Mitra, B., Diaz, F. & Craswell, N. (2017). Learning to match using local and distributed representations of text for web search. In *Proceedings of WWW '17*, 1291-1299. doi: 10.1145/3038912.3052579

Ponte, J. M. & Croft, W. B. (1998). A language modeling approach to information retrieval. In *Proceedings of SIGIR '98*, 275-281. doi: 10.1145/290941.291008

Salton, G., Wong, A. & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613-620. doi: 10.1145/361219.361220

Shen, D., Pan, R., Sun, J. T., Pan, J. J., Wu, K., Yin, J., & Yang, Q. (2006). Query enrichment for web-query classification. *ACM Transactions on Information Systems (TOIS)*, *24*(3), 320-352. doi: 10.1145/1165774.1165776

Song, F. & Croft, W. B. (1999). A general language model for information retrieval. In *Proceedings of CIKM '99*, 316-321. doi: 10.1145/319950.320022

Zhai, C. & Lafferty, J. (2001). Model-based feedback in the language modeling approach to information retrieval. In *Proceedings of CIKM '01*, 403-410. doi: 10.1145/502585.502654