

應用詞向量於語言樣式探勘之研究

Mining Language Patterns Using Word Embeddings

肖湘 Xiang Xiao

福州大學數學與計算機科學學院

元智大學大數據與數位匯流創新中心

College of Mathematics and Computer Science, FuZhou University

Innovation Center for Big Data and Digital Convergence, Yuan Ze University

n150320034@fzu.edu.cn

葉少珍 Shao-Zhen Ye

福州大學數學與計算機科學學院

College of Mathematics and Computer Science, FuZhou University

yeshzh@fzu.edu.cn

禹良治 Liang-Chih Yu

元智大學資訊管理學系

元智大學大數據與數位匯流創新中心

Department of Information Management, Yuan Ze University

Innovation Center for Big Data and Digital Convergence, Yuan Ze University

lcyu@staurn.yzu.edu.tw

賴國華 K. Robert Lai

元智大學資訊工程學系

元智大學大數據與數位匯流創新中心

Department of Computer Science & Engineering, Yuan Ze University

Innovation Center for Big Data and Digital Convergence, Yuan Ze University

krilai@cs.yzu.edu.tw

摘要

負面生活事件是引發抑鬱情緒的重要原因，比如家庭成員的去世、與伴侶的爭吵、被老闆開除或被老師責備等。主語和負面生活事件的組合稱為負面生活事件語言樣式。能否通過自動準確識別這些負面的生活事件來理解那些有抑鬱傾向的網絡文本對建立有效實用的精神病學網絡服務是至關重要的。如問題（父母：離婚::男友：？），本文使用分佈式詞向量模型來挖掘負面生活事件語言樣式，實驗結果 MRR 值達 0.71。

Abstract

Negative life event is an important reason of causing depression, such as the death of family members, quarreled with the spouse, fired by the boss or blamed by the teacher. The

combination of a subject and a negative life event is called a language pattern of negative life events. Therefore, whether it can understand those web texts with depression trend by identifying these negative life event language patterns automatically and accurately, which is important to establish effective and practical psychiatric network services. Such as the question (parent: divorce:: boyfriend:?), this study applied the distributional word embeddings to mine the language patterns of negative life events. The experimental results showed that the MRR index reached 0.71.

關鍵詞：自然語言處理，詞向量，語言樣式，負面生活事件

Keywords: Natural Language Processing, Word Embeddings, Language Patterns, Negative Life Events.

一、緒論

近年來，患抑鬱症的人群急劇擴大，抑鬱症患者自殺的消息轟動社會，抑鬱症受到了廣大關注。許多關注抑鬱症的網站和系統也應運而生，抑鬱症患者可以在網路上諮詢醫生，這些網站包含了大量的關於抑鬱症的資源，如何從中獲取到有價值的資訊，對是否能為患者提供更好的治療方式及能否有效地降低和預防抑鬱症都是至關重要的。在有關的醫療問答的網站上，受抑鬱情緒困擾的患者可以在網站上發佈他們的心情和發生的負面事件，會有專業的心理醫生在網站上進行診斷和回復。但是這個回復的時間取決於問題的數量，通常會是幾天的時間。而這個時間差對於抑鬱症患者而言是一個不可容忍的時間，尤其是對於那些有自殘自殺傾向的患者來說，可能就會錯失一個挽救的機會。如某位用戶的發文：“頭疼的現象很想拿把刀子把胸口剖開來透透氣”，表達了強烈的自殘的欲望。如果能夠自動識別問題中的負面生活事件，並通過患者發佈的問題迅速找到以往類似問題的答案，將其回饋給提問者，這會是一個非常有益的事情，患者可以看到也有很多人經歷過同樣的事情，承受過同樣的壓力，不會再感覺自己孤單一人，也會從類似問題的解答中尋找到適合自己的緩解辦法。

網路上的有關抑鬱症的問答文本通常會包含許多負面生活事件的語言樣式，但這些文本包含大量的非結構化的自然語言，負面生活事件也通常不是連續的詞，如“兄長因失戀沮喪辭公家單位優渥的工作”這一句話潛含了兩個負面事件“失戀”和“辭職”，可與主語組成兩個語言樣式（“兄長”，“失戀”）和（“兄長”，“辭職”）。負面生活

事件語言樣式就是指一個主語和一件負面事件構成的詞組。需要自動識別這些文本中的負面情緒，構造負面生活事件語言模式是關鍵。

詞向量是將自然語言符號數學化的關鍵技術，目前常見的有兩種表示方法：獨熱表示和分佈表示。前者向量的維度表示詞彙表大小，只有一位是 1 表示當前詞，其餘為 0，簡單直觀但易造成維數災難；後者每一維代表當前詞的一個潛在特征，是稠密的低維實數向量。詞語之間的關係和規律是通過計算相似度來衡量的。詞之間至少有關係相似度和特徵相似度兩種相似性，其中當兩個詞的特徵相似度很高的時候被稱為同義詞；當兩組詞的關係相似度很高的時候被稱為類比[1]。基於統計方法的點互信息 (Pointwise Mutual Information, PMI) 是最直接的計算詞語之間相似度的方法，只需要統計語料中固有的詞語頻率和共現資訊，很好地模擬了人類的大腦記憶[2]，且計算簡單易懂而在眾多自然語言處理任務中被廣泛使用[3-5]。Mikolov[6,7]通過對分佈式詞向量的矢量加減運算來衡量詞之間的關係相似度，分別對語義和詞法兩大類進行類比實驗，並獲得了高準確率。Levy 和 Goldberg[8]提出了計算關係相似度方法的變型，將向量加減法變型為乘除法，並在相同的詞對類比任務中獲得了更好的結果。Qiu 和 Zhang[9]等人應用這兩種類比方法將英文類比資料翻譯成中文進行了中文的類比研究，並表明可以通過過濾候選詞來獲得更好的結果。類比推理在計算機科學領域尤其是人工智慧方面吸引了眾多關注[10]，在自然語言處理領域，同樣應用廣泛，如改善問答 (Question Answering, Q-A) 任務中的答案排名準確率[11-13]；推薦系統[14]；詞語之間的潛在關係分析 (Latent Relational Analysis, LRA) [15,16]；以及不需任何預訓練和詞彙知識的中文分詞方法[17]等。類比在語言學習過程中也能幫學習者注意兩個概念間的關鍵對比[18]。

語言樣式的挖掘屬於模式挖掘的範疇，常見的模式挖掘多是使用關聯規則的方法[19,20]。對於負面生活語言樣式的探勘研究，Yu et al.結合進化推理演算法 (Evolutionary Inference Algorithm, EIA) 和 HAL (Hyperspace Analog to Language) 高維語義空間，從少量種子樣式集迭代地推論出額外的相關樣式，並加入了相應的反饋機制[21]；接著透過關聯規則的資料採擷演算法，發現句子中的頻繁項集來產生關聯語言樣式，並將找出來的負面事件語言樣式來幫助句子分類，同時提出了結合監督式資料採擷演算法和非監督式分佈式語言模型來發現抑鬱症文本中標記有負面生活事件句子的小語料庫中的關聯語言樣式[22,23]。

本文結構一共分為四章，接下來第二章主要介紹實驗用到的詞向量模型和相應的實驗方

法；第三章介紹資料集、實驗流程以及結果分析；第四章是總結以及對未來工作的展望。

二、詞向量模型和探勘方法

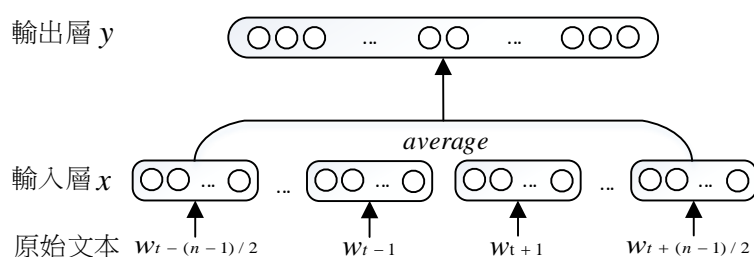
(一)、詞向量模型

語言模型是描述文字內在規律的數學模型，一般形式是給定一組詞，求解下一個詞的條件概率。常見的有基於統計的傳統語言模型和基於神經網絡的語言模型。**N-gram** 模型假設不改變上下文順序的前提下，距離相近的詞關係越近，當距離足夠遠時認為詞語之間沒有關聯，因而無法完全利用語料的信息。神經網絡語言模型是由 **Bengio[24]**正式提出的，通過一個三層（輸入層、隱藏層、輸出層）的神經網絡結構來構建語言模型，其中由隱藏層到輸出層的計算量是影響訓練效率的主要因素。神經網絡語言模型通過學習訓練語料庫獲取詞向量，捕獲了語義和詞法的特征信息。

Mikolov[6]提出的 **Skip-gram** 模型和 **CBOW** 模型，對神經網絡語言模型做了簡化，去掉了隱藏層，模型從神經網絡結構轉變為對數線性結構，大幅度減少了運算量，提高了訓練模型的效率。模型可以迅速有效地訓練大型語料庫，且得到的是分佈式的詞向量表示。

1、CBOW

CBOW 模型的結構如圖一所示，是根據給定上下文序列 $S = \{w_{t-(n-1)/2}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+(n-1)/2}\}$ 直接對目標詞 w_t 進行預測。



圖一、CBOW 模型結構圖

上下文所有詞對當前詞出現概率的影響權重一樣，不考慮詞序。神經網路語言模型的輸入是上文各詞向量首尾拼接而成，**CBOW** 模型中是上下文各詞向量的平均值。輸入為：

$$x = \frac{1}{n-1} \sum_{w_j \in S} C(w_j) \quad (1)$$

其中 w_j 表示目標詞 w_t 上下文的任一詞， $C(w_j)$ 表示詞 w_j 的詞向量。對目標詞進行預測：

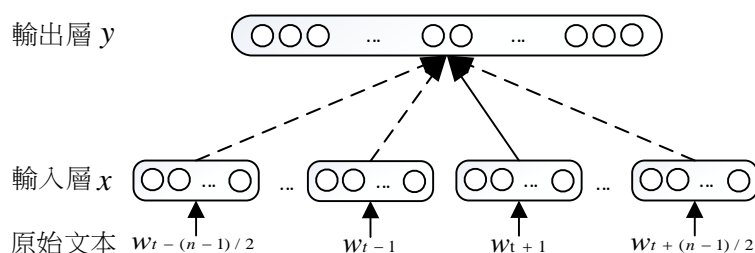
$$P(w_t | S) = P(w_t | w_{t-(n-1)/2}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+(n-1)/2})$$

$$= \frac{\exp(e'(w_t)^T x)}{\sum_{w' \in V} \exp(e'(w')^T x)} \quad (2)$$

2、Skip-gram

Skip-gram 模型結構如圖二所示，每次是從目標詞的上下文中選一個詞，將這個詞的向量表示作為模型的輸入作恒等投影，並認為與目標詞距離越遠的詞關係越小，賦予更小的權重。預測目標詞概率：

$$P(w_t | w_j) = \frac{\exp(e'(w_t)^T e(w_j))}{\sum_{w' \in V} \exp(e'(w')^T e(w_j))} \quad (3)$$



圖二、Skip-gram 模型結構圖

(二)、樣式探勘方法

詞語之間的語義及語法相關性可以通過詞向量的矢量運算來體現，詞向量的語義合成性 [7] 由矢量加法體現，如 $\text{vector}(\text{“俄羅斯”}) + \text{vector}(\text{“河流”}) \approx \text{vector}(\text{“伏爾加河”})$ ，類比相似性可以也可由簡單的數學運算來表現，如 $\text{vector}(\text{“北京”}) - \text{vector}(\text{“中國”}) \approx \text{vector}(\text{“巴黎”}) - \text{vector}(\text{“法國”})$ 。Mikolov [6] 在實驗中應用詞向量對英文五類語義和九類語法問題關係進行了測試，採用類比推理的原理，證明了其詞向量在捕獲語義和語法特征方面的有效性。

本文所針對的負面生活事件語言樣式符合類比推理的條件，對於存在類比關係的四個詞 “a:b::a*:b*” ，a 之於 b 與 a* 之於 b* 的關係一致，它們的詞向量之間存在 $V(b) - V(a) \approx V(b^*) - V(a^*)$ ，即 $V(b^*) \approx V(b) - V(a) + V(a^*)$ 的關係。問題可以轉化成計算候選詞與給定詞間相似度：

$$similarity = sim(b^*, b - a + a^*) \quad (4)$$

Mikolov 在實驗中選擇最常見的餘弦相似度，通過計算向量間夾角的餘弦值來衡量詞語之間的相似度，夾角為 0 時相似度為 1，夾角為 90 度時相似度為 0。COSINE 公式為：

$$similarity = \cos(b^*, b - a + a^*) \quad (5)$$

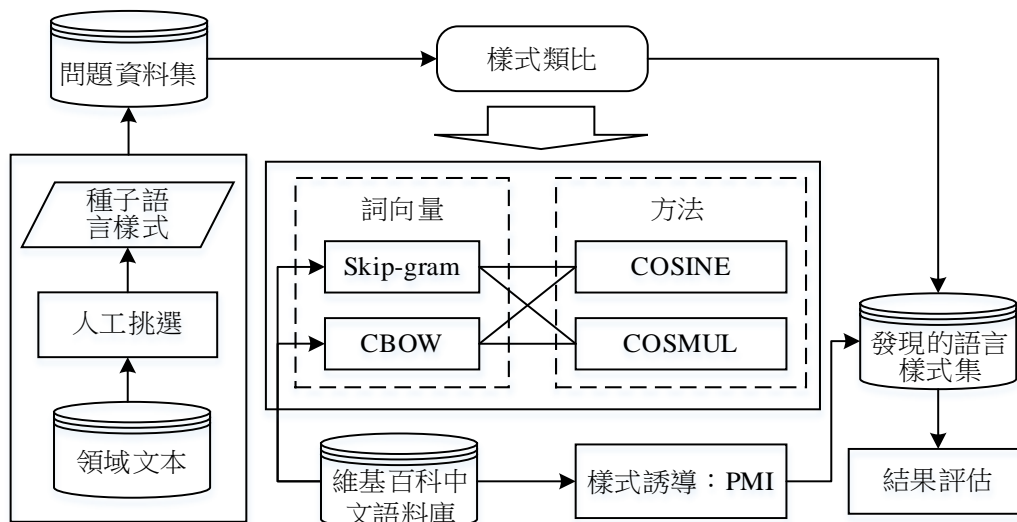
Levy 和 Goldberg[8]進一步提出了將加減法轉換成乘除法的變換式 COSMUL：

$$similarity = \frac{\cos(b^*, b) \cos(b^*, a^*)}{\cos(b^*, a) + \varepsilon} \quad (6)$$

其中調節因數 $\varepsilon = 0.001$ 避免除數為 0 的情況。

三、語言樣式探勘實驗

本文應用分佈式詞向量來進行生活中的負面事件的中文語言樣式探勘實驗，實驗的總體架構如圖一所示，主要分為三個階段：1) 實驗前期，需通過閱讀從網路獲取的抑鬱症文本，挑選出其中存在的負面生活語言樣式，構成種子語言樣式，再通過人工標記產生合理的主語和事件的組合，並產生問題資料集；2) 對維基百科中文語料庫進行中文繁體轉換、分詞、過濾特殊符號及停用詞等預處理，用不同的語言模型訓練出不同的詞向量模型；3) 對不同詞向量模型進行語言樣式探勘實驗，用之前生成的問題資料集按照相似度排名找出答案，並對結果進行評估。



圖一、語言樣式探勘實驗架構

同時設置一組 PMI 對比實驗，與使用詞向量結果進行比較。PMI 無需建立詞向量模型，直接統計語料庫中各詞的出現頻率和共同出現的頻率，從一個主語誘導出能夠構成語言樣式的負面事件。

負面生活語言樣式的探勘每一個問題都對應著多個答案，是一對多類型的推理。為了獲得更好的結果，對結果進行詞性過濾，因為本實驗中考慮的所有負面事件詞性均為動詞，因而在實驗中只對動詞進行推論，即只考慮“名詞：動詞”的語言樣式。採用兩個常用的評價指標，分別是序位倒數均值（Mean Reciprocal Rank，MRR）和前 n 個結果中的正確率 Precision@n。

（一）、語言樣式問題資料集

在精神病、心理學、抑鬱症等相關的網路論壇或諮詢網站上有很多的資源，包括用戶發表的生活中遇到的困難，心中的壓抑和焦慮以及對負面情緒的發洩等。在這些資源中隱含了很多生活負面事件的語言樣式，但是由於大多是以自然語言的描述方式，這些語言樣式通常不是很明顯的表示，可能是連續的詞語，但大部分是非連續的詞語，可能一個句子中會包括多個語言樣式，或者一段話中僅有一個語言樣式，需要一定的語義理解能力才能從中找出潛在的負面生活樣式。如表一舉例說明瞭原文和潛含的負面生活事件語言樣式。

表一、原文和潛含語言樣式舉例

類別	原文	語言樣式
家庭 Family	我好痛苦，父母總是用自己的角度來批判我。	（父母，批判）
愛情 Love	交往三年的男友突然要跟我分手，讓我整個跌入穀底。	（男友，分手）
學校 School	我在學校有被同學排斥使我更不會喜歡說話。	（同學，排斥）
工作 Work	上班也總是被老闆抓到把柄，責備了我好幾次。	（老闆，責備）
社會 Social	跟朋友的關係鬧得很僵，經常說幾句就吵架。	（朋友，吵架）

本實驗中研究的是由名詞和動詞構成的語言樣式，其中名詞是指是日常生活中常見的人物角色的名稱，動詞是指日常生活中常見的負面事件，並且兩者要滿足合理的邏輯關係，即該主語發生該動作是可能的並且是合理的。

為了獲得初始的種子語言樣式，我們人工閱讀了 500 篇抑鬱症相關文本，找出其文本中暗含的負面生活語言樣式，共找出了 132 組這樣的主語和動作的組合，作為種子語言樣

式。並按主語所屬生活領域將語言樣式分為家庭、愛情、學校、工作和社會五個類別。將種子語言樣式拆分成主語和動作兩部分，並分別對人物名稱和負面生活事件進行了一定擴充。由三位研究生分別填寫列頭為主語、行頭為動作的表格，如果認為對應的主語發生該動作是合理的，則在對應的區域標記為 1，否則不標記。實際完成的表格是一個 152×54 的矩陣。對於同一個位置，若有兩人及以上的人標記了 1，則認為該語言樣式成立。表二是語言樣式標記示例，對於“生病”和“吵架”兩個事件而言，對所有的主語皆成立，而“離異”、“分手”、“輟學”、“辭職”這四個事件則不是對所有主語成立。

表二、語言樣式標記示例

	父母	男友	同學	同事	朋友
生病	1	1	1	1	1
吵架	1	1	1	1	1
離異	1				
分手		1			
輟學		1	1		1
辭職	1	1		1	1

通過整理三份不同的標記結果，最後生成了共 6002 組合理的主語和動作的語言樣式。本研究中的語言樣式探勘是指在給定 a、b、a* 三個對象的前提下來找出相似度最高的第四個對象 b*。上文中已經完成了所有的主語與動作的有效組合 (a:b 對)。要進行類比樣式探勘，需要先生成問題資料集，即所有的“a:b :: a*:?” 組合。其中 a 和 a* 都是指日常生活中常見的人物角色名稱，b 是日常生活中常見的負面壓力事件。生成問題資料集的方式是生成所有“a:b”對與 a* (在同一個問題中不同於 a) 的排列組合。問題資料集的形式如表三所示，當同一個問題中 a 和 a* 來自於同一類別的時候稱為同類別類比，否則稱為非同類別類比。

表三、問題及答案示例

	a:b :: a*:?	b*
同類別	哥哥:輟學::弟弟:?	退學、休學...
	同學:嘲諷::老師:?	嘲弄、取笑、譏諷...
非同類別	朋友:背叛::老師:?	背棄、背離...
	妻子:鬥嘴::同事:?	吵架、爭執、爭吵...

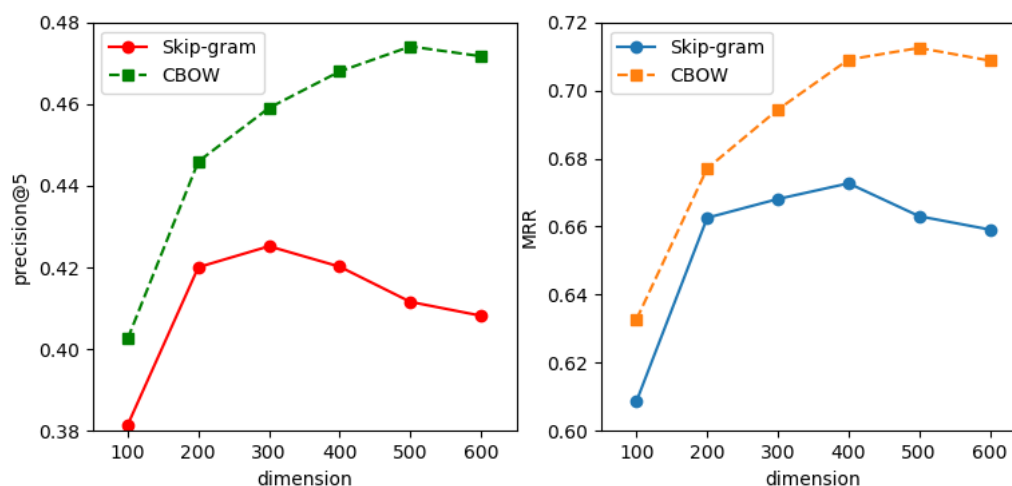
要對不同的向量模型找出來的語言樣式進行測評，必須要建立一份標準答案。在本實驗中定義答案的標準是該動作是主語 a* 所在類別能夠合理發生的事件。為了盡量減少漏掉的正確答案，本實驗中通過廣義知網 (E-HowNet) [25] 對答案進行了擴充。

(二)、語言樣式探勘實驗

問題資料集分為開發集 (Development Set) 和測試集 (Test Set)，其中開發集用來調節訓練詞向量過程中的參數，選擇其最優的參數進行測試實驗。

1、詞向量模型調參

Skip-gram 模型和 CBOW 模型通過開發集來調節詞向量的維度，圖二是調參結果，橫軸表示維度，縱軸表示評估結果，圖中曲線隨維度上升都呈現先上升後下降的趨勢，其中 CBOW 模型後期變化曲線趨於平緩，下降趨勢不太明顯。Skip-gram 模型詞向量取 300 維時 precision@5 結果最好，取 400 維時 MRR 結果最好；CBOW 模型取 500 維時 precision@5 和 MRR 評估指標的結果同時最好。



圖二、訓練詞向量調參結果

2、實驗結果及分析

調參過程結束後，已經確定了各向量模型將在測試階段用於不同評估指標的參數。測試問題集根據給定兩個主語的所屬類別，分為 25 個子集，其中同類別類比問題集有 5 個，其餘 20 個為非同類別問題集。測試實驗將分別對每一個測試問題子集進行測試。每一個子實驗用模型名字+類比函數名字標識，如 Skip-gram 模型用餘弦相似度進行計算的實驗，表示為 “Skip-gram+COSINE”。表四表示兩個向量模型在使用 COSINE 方法進行同類別實驗和非同類別類比實驗的平均結果。

表四、COSINE 實驗同類別和非同類別實驗結果

評估指標	同類別		非同類別	
	Skip-gram	CBOW	Skip-gram	CBOW
Precision@5	0.4345	0.4904	0.4215	0.4684
Precision@10	0.3824	0.43	0.3705	0.4129
MRR	0.6683	0.7287	0.6683	0.7037

從表中可以看出，同類別實驗結果優於非同類別實驗，CBOW 模型表現比 Skip-gram 要好，其中同類別實驗最優，Precision@5 和 MRR 結果分別領先 Skip-gram 模型 5.59% 和 6.04%，非同類別實驗分別領先 Skip-gram 模型 4.69% 和 3.54%。

表五表示兩個向量模型在使用 COSMUL 方法進行同類別實驗和非同類別類比實驗的平均結果。總體而言同類別的類比實驗結果比非同類別的實驗表現稍好，三個評估指標中的兩個 Precision@n 指標值都高於非同類別實驗結果，Skip-gram 模型和 CBOW 模型的 Precision@5 指標分別比非同類別類比實驗高 2.1% 和 2.68%。CBOW 模型的 MRR 結果差距較明顯，同類別實驗結果高出非同類別實驗 3.76%。

表五、COSMUL 實驗同類別和非同類別實驗結果

評估指標	同類別		非同類別	
	Skip-gram	CBOW	Skip-gram	CBOW
Precision@5	0.3961	0.4272	0.3751	0.4004
Precision@10	0.3516	0.3809	0.3321	0.3589
MRR	0.6286	0.6654	0.6197	0.6278

在本研究中一共做了 5 組測試實驗，其中 PMI 是作為對比實驗的語言樣式誘導實驗，其餘 4 組則是使用詞向量的實驗，每一組類比實驗測試了 25 個問題集，包括 5 個同類別實驗和 20 個非同類別實驗，表六中對五組實驗的總體結果進行了一個對比，其中 4 組類比實驗是 25 個子實驗結果的平均值。對比表中的五組實驗結果，四組使用詞向量的結果都比 PMI 好，說明這探勘語言樣式這個領域，使用詞向量的效果比簡單的樣式誘導好。總體實驗結果表明 CBOW 模型的表現優於 Skip-gram 模型，COSINE 方法表現優於 COSMUL 方法，所有實驗中最好的是 CBOW+COSINE 這一組結果，Precision@5 平均結果達 0.47，MRR 值達到 0.71。

表六、總體實驗結果對比

實驗名稱	Precision@5	Precision@10	MRR
PMI	0.1296	0.1241	0.2589
Skip-gram+COSINE	0.424	0.3729	0.6683
CBOW+COSINE	0.4728	0.4163	0.7087
Skip-gram+COSMUL	0.3793	0.336	0.6215
CBOW+COSMUL	0.4057	0.3633	0.6353

四、結論與展望

本論文應用詞向量針對負面生活事件進行語言樣式探勘，選取近年來熱門的 Word2Vec 中兩種低維分佈式詞向量模型 Skip-gram 和 CBOW 模型，使用 Mikolov 實驗中用到的也是最常見的餘弦相似度，以及 Levy 和 Goldberg 提出的餘弦相似度的乘除變型的方法。從網絡抑鬱症文本中挑選出負面生活語言樣式，經人工判斷生成問題資料集，再通過廣義知網擴充答案集。使用維基百科中文語料庫訓練詞向量模型，樣式探勘實驗主要分為調參和測試實驗。為了使實驗更完整，設置了一組 PMI 基礎實驗作為對比。共 5 大組實驗，所有使用詞向量的實驗結果優於 PMI 實驗，說明使用詞向量的方法更適用於語言樣式探勘。同類別實驗結果好於非同類別實驗結果，說明詞向量捕獲的同類別語義特徵更準確。CBOW 模型平均結果最好，說明其訓練出的詞向量捕獲的語義資訊與負面生活事件語言樣式的要求更契合。

本研究是首次將詞向量模型結合類比推理的思想應用於語言樣式探勘研究領域，並且結果好於傳統的 PMI 方法。在語言樣式探勘實驗中，具體探討了不同向量模型在捕捉語義資訊方面的準確性高低，以及不同詞向量方法的優劣。接下來的工作，我們會從三個方面進一步完善研究：一是將納入其他熱門的詞向量模型和類比方法，探討不同向量模型和方法的效能優劣；二是會對更多不同的語料進行樣式探勘，如抑鬱症專業領域的語料，或基本語料庫與抑鬱症專業領域語料相結合等；三考慮將語言樣式建立標準資料集並投入到進一步的研究中如文本分類等。

參考文獻

- [1] Turney, P. D., "Similarity of semantic relations," *Computational Linguistics*, vol. 32, no.3, pp. 379-416, 2006.
- [2] G. Recchia, and M. N. Jones, "More data trumps smarter algorithms Comparing pointwise mutual information with latent semantic analysis." *Behavior research methods*, vol. 41, no.3, pp. 647-656, 2009.
- [3] P. D. Turney, "Mining the web for synonyms PMI-IR versus LSA on TOEFL," *European Conference on Machine Learning*, Springer Berlin Heidelberg, pp. 491-502, 2001
- [4] E. Terra, and C. L. Clarke, "Frequency estimates for statistical word similarity measures," *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, pp.165-172, May, 2003.
- [5] T. Van de Cruys, "Two multivariate generalizations of pointwise mutual information," *Proceedings of the Workshop on Distributional Semantics and Compositionality*, Association for Computational Linguistics, pp. 16-20, June, 2011.
- [6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv1301.3781*, 2013.
- [7] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, pp. 3111-3119, 2013.
- [8] O. Levy, Y. Goldberg, and I. Ramat-Gan, "Linguistic Regularities in Sparse and Explicit Word Representations," *CoNLL*, pp. 171-180, 2014.
- [9] L. Qiu, Y. Zhang, Y. Lu, "Syntactic Dependencies and Distributed Word Representations for Chinese Analogy Detection and Mining," *EMNLP*, 2015.
- [10] H. Prade, and G. Richard, "A short introduction to computational trends in analogical reasoning," *Computational Approaches to Analogical Reasoning: Current Trends*. Springer Berlin Heidelberg, pp. 1-22, 2014.
- [11] H. Toba, M. Adriani, and H. M. Manurung, "Predicting Answer Location Using Shallow Semantic Analogical Reasoning in a Factoid Question Answering System," *PACLIC*, pp.

- 246-253, 2012.
- [12] X. Tu, D. Feng, X. J. Wang, and L. Zhang, "Analogical reasoning for answer ranking in social question answering," *IEEE Intelligent Systems*, vol. 27, no. 5, pp. 28-35, 2012.
- [13] V. K. Chaudhri, S. Heymans, A. Overholtzer, and M. Wessel, "Large-Scale Analogical Reasoning," *AAAI*, pp. 359-365, 2014.
- [14] N. Hug, H. Prade, and G. Richard, "Experimenting Analogical Reasoning in Recommendation," *International Symposium on Methodologies for Intelligent Systems*, Springer International Publishing, pp. 69-78, 2015.
- [15] N. T. Duc, D. Bollegala, and M. Ishizuka, "Using relational similarity between word pairs for latent relational search on the web," *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. IEEE, vol.1, pp. 196-199, 2010.
- [16] C. Liang, and Z. Lu, "Chinese Analogy Search Considering Multi Relations," *Cloud and Service Computing (CSC), 2012 International Conference on*. IEEE, pp.193-197, 2012.
- [17] Z. Zheng, Y. Wang, and Y. Lepage, "Chinese word segmentation based on analogy and majority voting," *PACLIC*, 2015.
- [18] D. Gentner, and L. L. Namy, "Analogical processes in language learning," *Current Directions in Psychological Science*, vol. 15 no. 6, pp. 297-301, 2006.
- [19] J. T. Chien, H. Y. Chen, "Mining of association patterns for language modeling," *INTERSPEECH 2004 - Icslp, International Conference on Spoken Language Processing*, Jeju Island, Korea, October, 2004.
- [20] A. C. Mendes, and C. Antunes, "Pattern Mining with Natural Language Processing: An Exploratory Approach," *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, Springer Berlin Heidelberg, pp. 266-279, 2009.
- [21] L. C. Yu, C. H. Wu, J. F. Yeh, and F. L. Jang, "HAL-Based Evolutionary Inference for Pattern Induction From Psychiatry Web Resources," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 2, pp. 160-170, 2008.
- [22] L. C. Yu, C. L. Chan, C. H. Wu, and C. C. Lin, "Mining association language patterns for negative life event classification," *Proceedings of the ACL-IJCNLP 2009 Conference*

Short Papers, Association for Computational Linguistics, pp. 201-204, 2009.

- [23] L. C. Yu, C. L. Chan, C. C. Lin and I. C. Lin, “Mining association language patterns using a distributional semantic model for negative life event classification,” *Journal of Biomedical Informatics*, vol.44, no. 4, pp. 509-518, 2011.
- [24] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, “A Neural Probabilistic Language Model,” *Journal of Machine Learning Research*, pp.1137-1155, 2003.
- [25] 中文詞知識庫小組, “廣義知網知識本體架構 2.0 版,” <http://ehownet.iis.sinica.edu.tw> , 2011.