

探究使用基於類神經網路之特徵於文本可讀性分類

Exploring the Use of Neural Network based Features for Text Readability Classification

曾厚強*、陳柏琳*、宋曜廷⁺

Hou-Chiang Tseng, Berlin Chen and Yao-Ting Sung

摘要

可讀性通常指的是閱讀題材可以被讀者理解的程度：當閱讀材料愈能夠被讀者所理解時，就愈能夠產生好的學習效果。為了能夠幫助讀者去適配符合自己閱讀能力的文件，研究人員長久以來持續發展各種能夠自動且精準地估測文本可讀性的模型來達到此目標。可讀性分類通常是透過分析文件上的資訊來轉化成一組可讀性特徵，再利用這些可讀性特徵來訓練出可讀性模型，以便能預測未知文件的可讀性。然而，傳統的可讀性模型所使用的特徵都需要根據專家的經驗來進行選取，這卻也限制其實用性。近年來隨著表示學習法技術的蓬勃發展，訓練可讀性模型所需要的特徵可以不再需要仰賴專家，這也使得可讀性模型的發展有了一個嶄新的研究方向。因此，本論文嘗試以卷積神經網路以及快速文本兩種技術分別來自動地擷取文本特徵，以訓練出一個能夠分析跨領域文件的可讀性模型，並可以因應文件內容多元主題的特性。經與現有方法的一系列實驗比較後，其結果確認了本論文所提可讀性模型的效能優勢。

關鍵字：可讀性、詞向量、卷積神經網路、表示學習法、快速文本。

Abstract

Text readability refers to the degree to which a text can be understood by its readers: the higher the readability of a text for readers, the better the the

* 國立臺灣師範大學資訊工程系

Department of Computer Science & Information Engineering, National Taiwan Normal University
E-mail: ouartz99@gmail.com, berlin@cise.ntnu.edu.tw

⁺ 國立臺灣師範大學教育心理與輔導學系

Department of Educational Psychology and Counseling, National Taiwan Normal University
E-mail: sungtc@ntnu.edu.tw

comprehension and learning retention can be achieved. In order to facilitate readers to digest and comprehend documents, researchers have long been developing readability models that can automatically and accurately estimate text readability. Conventional approaches to readability classification is to infer a readability model using a set of handcrafted features defined a priori and computed from the training documents, along with the readability levels of these documents. However, the use of handcrafted features requires special expertise and its applicability also is limited. With the recent advance of representation learning techniques, we can efficiently extract salient features from documents without recourse to specialized expertise, which offers a promising avenue of research on readability classification. In view of this, we in this paper propose two novel readability models built on top of a convolutional neural network based representation and the so-called fastText representation, respectively, which have the capability of effectively analyzing documents belonging to different domains and covering a wide variety of topics. A series of empirical experiments seem to demonstrate the utility of the proposed models in relation to several existing methods.

Keywords: Readability, Word Vector, Convolutional Neural Network, Representation Learning, fastText.

1. 緒論 (Introduction)

一般而言，可讀性(Readability)是指閱讀材料能夠被讀者所理解的程度(Dale & Chall, 1949; Klare, 1963, 2000; Mc Laughlin, 1969)，當讀者閱讀高可讀性的文件時，會產生較好的理解及學後保留效果(Klare, 1963, 2000)。由於文件的可讀性在知識傳遞扮演極為重要的角色，因此西方的可讀性公式發展的非常早，如：1923年 Bertha 等人就提出方法來探討教科書中字彙難度的問題(Bertha & Pressey, 1923)。另外，Vogel 和 Washburne 在 1928 年則是提出一個 Winnetka Formula 來評量小孩讀物的可讀性(Vogel & Washburne, 1928)。由於可讀性相關的研究非常重要。因此，據 Chall 與 Dale 在 1995 年的統計，到 1980 年為止相關的可讀性公式就已經超過 200 多個可讀性公式(Chall & Dale, 1995)。這些傳統的可讀性研究大多使用較淺層的語言特徵來發展線性的可讀性公式，例如 Flesch Reading Ease 採用詞彙的平均音節數與平均的句子長度(Flesch, 1948)或 Chall 和 Dale 計算難詞在文章中的比率(Chall & Dale, 1995)等，都是傳統可讀性公式代表之一。然而，傳統可讀性公式所採用的淺層語言特徵，並不足以反映文件難度。Graesser、Singer 和 Trabasso 便指出，傳統語言特徵公式無法反映閱讀的真實歷程，文件的語意語法只是文件的淺層語言特徵，沒有考量文件的凝聚特性(Graesser, Singer & Trabasso, 1994)。Collins-Thompson 亦指出傳統可讀性公式僅著重在文件的表淺資訊，而忽略文件重要的深層特徵。這也讓傳統可讀性公式在預測文本可讀性的結果常遭受到質疑(Collins-Thompson, 2014)。直至今日，可讀性的研究仍持續不斷。研究人員為了克服傳統可讀性公式的缺點，嘗試利用更細緻的

機器學習演算法來發展出非線性的可讀性模型，並納入更多元的可讀性指標來共同評量文本的可讀性，以提升可讀性模型的效能(Petersen & Ostendorf, 2009; Feng, Jansche, Huenerfauth & Elhadad, 2010; Sung *et al.*, 2015)。

然而可惜的是，研究人員發現採用一般語言特徵的可讀性模型在應用於特定領域文時，一般語言特徵並無法判斷詞彙在不同領域文本時背後所代表的意義。其原因在於特定領域文本的內容著重在闡述領域的「知識概念」，而這樣子的描述方式有別於一般語文的敘述文或故事體的結構。Yan 等人就明確指出在計算美國大型醫學資料庫(Medical Subject Headings, MeSH)中的專業術語去探討，發現語言特徵公式的音節數、字長與醫學類專業詞彙的困難度無相關。換句話說，採用淺層語言特徵的可讀性模型並無法反映特定領域文件中專業術語的難度(Yan, Song & Li, 2006)。針對一般語言特徵無法表徵特定領域知識結構的問題，開始有學者針對這個議題進行研究。例如，Yan 等人利用本體論的技術將美國國家醫學資料庫(Medical Subject Headings, MeSH)的醫學符號階層資料庫作為概念資料庫，從中找出每一個醫學類文件中的概念，並計算概念到此樹狀結構最底部的距離，得出每篇文件概念深度指標(Document Scope) (Yan *et al.*, 2006)。Borst 等人則是利用詞表的方式將每個詞彙的「類別複雜度」與「詞頻」兩個分數加總來計算詞彙複雜度，作為評估醫學類線上文件詞彙、句子及文件難度的依據(Borst, Gaudinat, Grabar & Boyer, 2008)。

由上述的研究可知，不論是過去一般語言特徵或是針對特定領域文本的知識結構所設計的文件表示(Document Representation)技術，長久以來都需要仰賴專家來研發，有著耗時費力等問題。近年來，有所謂表徵學習(Representation Learning)方法可以自動從原始資料中去抽取有用的資訊，能有助於建立分類模型和預測測試資料(Goodfellow, Bengio & Courville, 2016)。使得模型所需要的特徵可以逐漸不需仰賴專家，成功開啟了另一個研究的方向。因此本研究將基於近年來熱門的表徵學習法來自動從文本中抽取出可讀性模型所需要的特徵，訓練一個能夠分析跨領域文件的可讀性模型。本論文的內容安排如下：第二節將描述目前採用表示學習法來研發可讀性特徵或是訓練可讀性模型的相關研究。第三節將基於表示學習技術來訓練出一個能夠同時分析不同領域文件的可讀性模型。第四節將呈現本論文所提出可讀性模型的效能。最後第五節是總結及未來研究的方向。

2. 相關研究 (Related Work)

在文本可讀性的研究中，潛在語意分析(Latent Semantic Analysis, LSA)是早期非常受歡迎的語意分析技術之一(Landauer & Dumais, 1997; Landauer, Foltz & Laham, 1998)。其技術如圖 1 所示，潛在語意分析僅需要將詞彙-文章矩陣利用奇異值分解(singular value decomposition, SVD) 將維度縮減，便可以擷取出語料庫的語意空間來表達文件潛藏語意屬性，在取得潛在語意空間(U)後便可以去測量任意二個詞彙、句子、段落及文章之間的語意相似度。在過去，已經許多學者利用潛在語意分析這種表徵學習法應用在可讀性的相關研究。如 Graesser 等人在 Coh-Metrix 3.0 中提供了八個跟 LSA 相關的指標來測量句子或篇章的相似程度(Graesser, McNamara, Louwerse & Cai, 2004)。Truran 等人則是利用

潛在語意分析技術來研究醫學臨床文章的可讀性(Truran, Georg, Cavazza & Zhou, 2010)。François 和 Miltsakaki 利用潛在語意分析去計算詞彙之間的凝聚性來當成可讀性模型的語意指標，以分類法文為第二學習語言的書本可讀性(François & Miltsakaki, 2012)。Kireyev 和 Landauer 使用潛在語意分析來觀察字的成熟度(Word Maturity)，以估測出詞彙的年級難度(Kireyev & Landauer, 2011)。Chang 等人利用潛在語意分析於社會科和自然科的教科書，利用相似度的做法來歸納出每個年級的特定知識概念(Chang, Sung & Lee, 2013)。

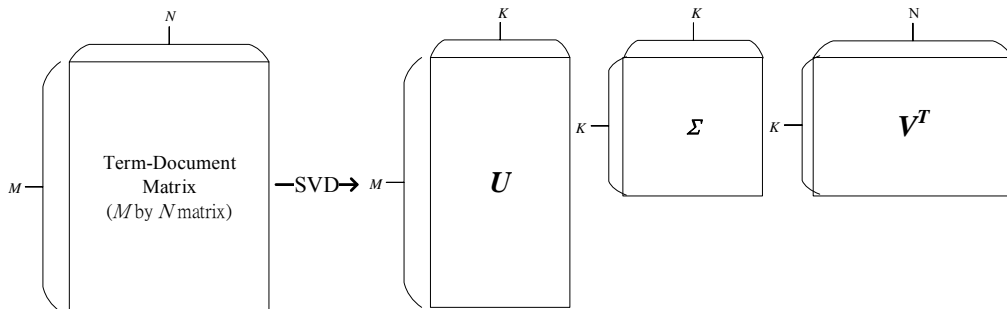


圖1. 潛在語意分析運用奇異值分解抽取潛在語意空間

[Figure 1. Latent semantic analysis using singular value decomposition to extract the latent semantic space.]

另一個常被應用於可讀性研究的表示學習法是：詞向量，詞向量表示的觀念最早由 Hinton 在 1986 年所提出，又被稱為詞表示(Word Representation or Word Embedding) (Hinton, 1986)。Bengio 在 2003 年提出回饋式類神經網路語言模型(Feed-forward Neural Network Language Model (FFNNLM)的訓練架構，從文件中詞彙前後相鄰的關係來求取詞向量表示(Bengio, Ducharme, Vincent & Jauvin, 2003)。而近期 Google 所發表的 Word2vec 則可視為 FFNNLM 的後繼方法(Mikolov, Chen, Corrado & Dean, 2013)。然而跟 FFNNLM 架構不一樣的是，Word2vec 去除了 FFNNLM 在訓練時最耗時的非線性隱藏層，僅保留輸入層、投影層和輸出層，使其架構更加簡單。Word2vec 提供了二種訓練方式，分別是連續詞袋模型(Continuous Bag-of-words Model, CBOW)及略詞模型(Skip-gram Model, Skip-gram)。連續詞袋模型主要的精神是由目標詞之外的前後文來預測目標詞的機率；而略詞模型的訓練方式正好相反，它是由目標詞本身來去預測前後文的機率，二種訓練模型示意圖如圖 2(a)及圖 2(b)所示。在 Word2Vec 中不論是連續詞袋模型還是略詞模型，在輸出層都可以採用 Hierarchical Softmax 或是 Negative Sampling 兩種模式來增進訓練的效能。其中 Hierarchical Softmax 指的是將訓練資料中不同詞彙都建置霍夫曼樹(Huffman tree)上，使得根節點(root)到每個詞彙都是唯一的路徑，接著在訓練的過程中，不斷得更新霍夫曼樹上每個節點所對應的權重外，也逐步更新詞彙所對應的向量。而 Negative Sampling 則是捨棄了霍夫曼樹的作法；在訓練前除了原本的正例的樣本外，還額外選了數個負例的樣本，在訓練的過程中不斷更新權重，使得正例樣本的機率最大化外，也同時降低了負例樣本的機率，讓詞彙所對應的向量可以逐步獲得修正。在目前也有學者已

經將 Word2vec 應用於可讀性模型，例如 Liu 等人便將 Word2vec 當成可讀性模型裡其中的一個特徵，以分析中、小學國文教科書及優良課外讀物的可讀性(Liu, Chen, Tseng & Chen, 2015)。Tseng 等人則是將 Word2vec 結合支向量機(Tseng, Sung, Chen & Lee, 2016a)或深層類神經網路(Tseng, Hung, Sung & Chen, 2016b)發展出一個能夠同時分析國文科、社會科及自然科等不同領域文本的可讀性模型。

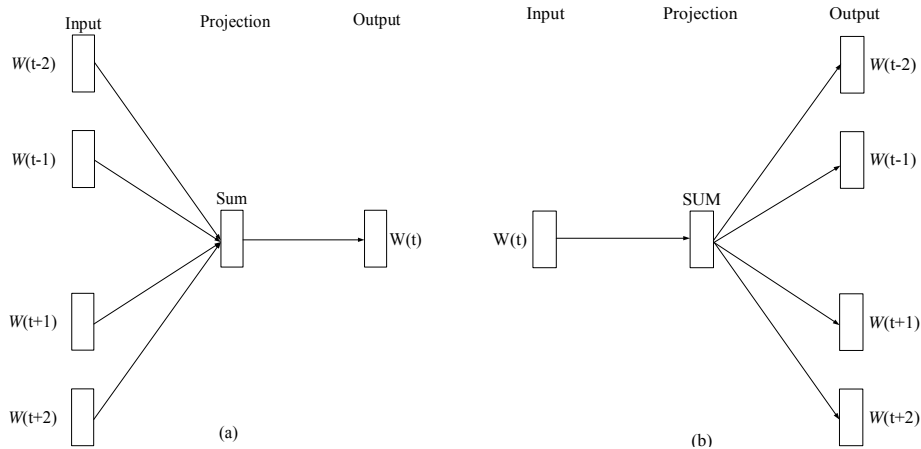


圖 2. (a) 連續詞袋模型訓練演算法。(b) 略詞模型訓練演算法。

[Figure 2. (a) The continuous bag-of-words model; (b) The skip-gram model.]

由上述可知研究可知不論是潛在語意分析或是 Word2vec，都可以在不需領域專家的介入之下，依據其演算法自動從文本中抽取出可讀性模型所需要的特徵。近年來表示學習法仍蓬勃發展，因此，本論文將嘗試以卷積神經網路(Convolutional Neural Network, CNN) (LeCun, 1989)或快速文本(fastText) (Joulin, Grave, Bojanowski & Mikolov, 2016)等不同的表示學習法來自動抽取文本特徵的技術，訓練出一個能夠分析跨領域文件的可讀性模型。

3. 基於表示學習技術之可讀性模型 (Readability Model Based on Representation Learning Techniques)

3.1 卷積神經網路 (Convolutional Neural Network)

卷積神經網路是一種分層式的分類類型結構，它的每個模組都是由卷積層(Convolutional Layer)和池化層(Pooling Layer)來組成(LeCun, 1989)，通過模組不斷的疊加或是加上多層的深層類神經網路後，以形成深度的模型。整個卷積神經網路透過三個重要的思想來幫助改進機械學習的系統：稀疏交互(Sparse Interactions)、參數共享(Parameter Sharing)及等變表示(Equi-variant Representations) (Goodfellow *et al.*, 2016)。稀疏交互又稱為稀疏連接(Sparse Connectivity)，主要是利用數個核(Kernel)基於自定的大小(Kernel Size)來局部連結兩層網路，使得整個模型所要儲存的參數變少，可以有效的減少計算量而提高計算效率。參數共享指的是在同一個核中，每一個元素在不同位置所作用的權重都是相同的。這意

味著在卷積運算的過程中，模型只需要學習一組固定的參數即可，而不是對於每個元素在不同的位置所用的權重都是獨立的。因此這也將大幅度提高模型的訓練效能。而參數共享的機制再加上適當的池化策略，也促成了卷積類神經網路對於局部平移有一些不變的特性可以應用於圖像的處理或語音辨識，尤其是在關心某個特徵是不是有出現，而不是關心它出現的具體位置時(Goodfellow *et al.*, 2016)。目前卷積類神經網路已經成功被應用於圖像分析(Cireşan, Meier, Gambardella & Schmidhuber, 2010; Cireşan, Meier, Masci & Schmidhuber, 2011; Ciresan, Giusti, Gambardella & Schmidhuber, 2012)、語音辨識(Abdel-Hamid, Deng & Yu, 2013; Deng, Abdel-Hamid & Yu, 2013; Deng *et al.*, 2013)和自然語言處理(Kim, 2014; Zhang & Wallace, 2015; Johnson & Zhang, 2014)。本研究也嘗試將卷積類神經網路用來自動抽取可讀性模型所需要的特徵，並利用深層類神經網路訓練出可讀性模型，其架構如圖 3 所示，在訓練的過程中將使用到 Dropout (Srivastava, Hinton, Krizhevsky, Sutskever & Salakhutdinov, 2014)的技巧來避免模型過度適配(overfitting)外，並利用 rectified linear units (ReLU) (Nair & Hinton, 2010)作為的激發函數(active function)，以避免典型的梯度消失(gradient vanish)問題。

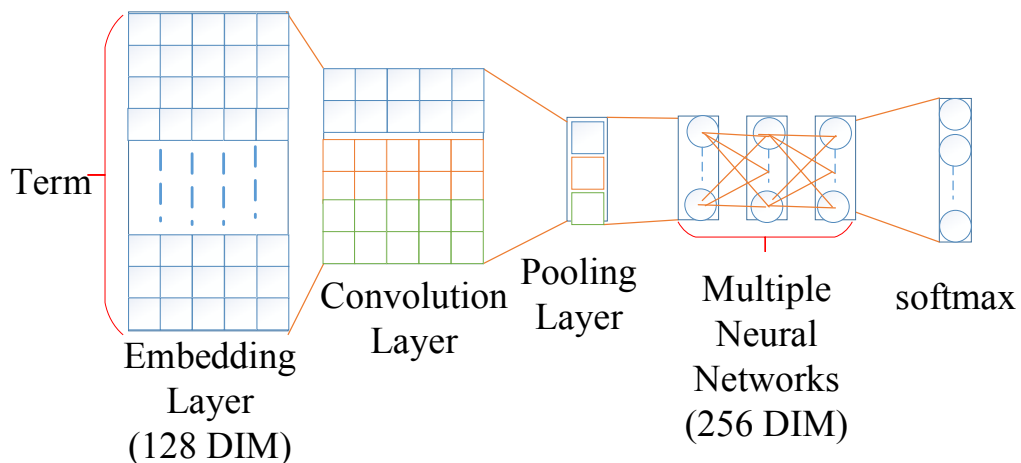


圖 3. 基於卷積類神經網路之可讀性模型架構

[Figure 3. A Framework of Readability Model Based on Convolutional Neural Networks.]

3.2 快速文本(fastText)方法 (fastText Library)

繼 Word2vec 之後，Joulin 等人持續改變 Word2vec 的架構發展出快速文本(Joulin *et al.*, 2016)。快速文本與 Word2vec 一樣有連續詞袋模型和略詞模型兩種架構，是屬於基於一個長度來看詞彙之間的關係，而透過滑動視窗(Sliding Window)來進行訓練的技巧可以彈性適用於不同長度的文本。但不一樣的地方在於將目標詞彙改換成訓練資料的類別，其示意圖如圖 4(a)及圖 4(b)所示。除此之外，在輸入層方面，也由 Word2vec 的 unigram 改

成 n -gram 的架構以供模型訓練時有更多的彈性。至於快速文本的輸出層是要採用 Hierarchical Softmax 或 Negative Sampling 則跟 Word2vec 一樣，可以自由選用。而快速文本這樣子的訓練技巧也有別於潛在語意分析與 Word2vec，直接將分類器的需求巧妙的整合至訓練的演算法之中。因此，本研究也嘗試將快速文本採用連續詞袋演算法來訓練可讀性模型。

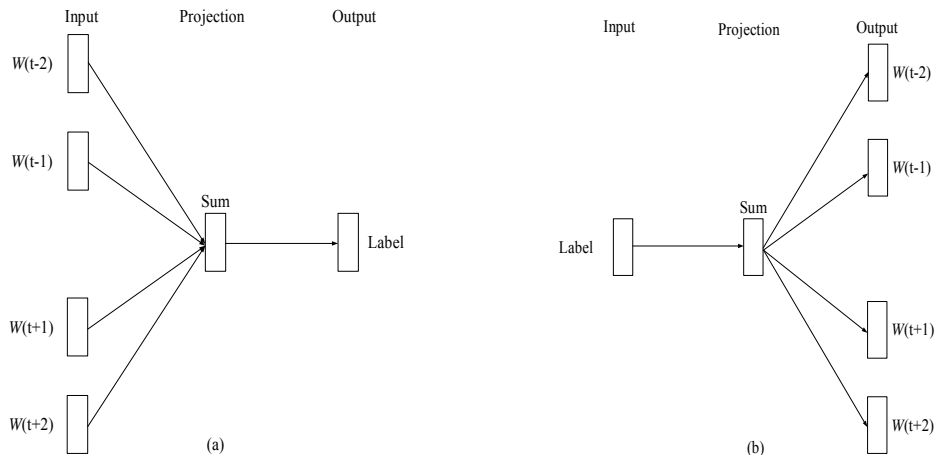


圖4. (a) 快速文本之連續詞袋模型。(b) 快速文本之略詞模型。

[Figure 4. (a) Continuous Bag-of-words Model of fastText. (b) Skip-gram Model of fastText.]

4. 實驗及結果 (Experimental Results)

4.1 實驗材料 (Materials)

本研究材料選自 98 年度臺灣 H、K、N 三大出版社所出版的 1-12 年級審定版的國語科、社會科、自然科及體育和健康教育等四個領域的教科書全部共計 6,230 篇，各版本教科書均經由專家根據課程綱要編制而成，其實驗材料的年級分佈如表 1 所示。

表1. 實驗材料在各年級的數量分佈
[Table 1. The statistics of the dataset.]

年級	1	2	3	4	5	6	7	8	9	10	11	12
社會科	0	0	80	74	85	81	389	407	325	340	331	270
自然科	0	0	72	67	67	62	172	175	157	211	355	295
國文科	24	67	61	71	69	70	37	34	28	84	41	47
體育及健康教育科	125	125	121	144	149	150	79	91	85	197	139	177

4.2 訓練可讀性模型 (Train Readability Models)

首先將實驗材料利用 WECAn(Chang, Sung & Lee, 2012)來進行中文斷詞的前處理程序，接著再將訓練資料利用卷積類神經網路或快速文本來抽取出可讀性模型所需要的特徵，而它的類別就是課文所屬的年級，並未再細分文本是屬於何種領域。本研究利用 Keras (Chollet, 2015)予以實作，整個實驗流程皆採用 5-fold 交互驗證的方式如圖 5 所示。在挑選訓練資料的時候，本研究分別從四大領域的文本中各年級文本數量依比例亂數挑選，接著將這些訓練資料輸入至斷詞的程序，各文本經過斷詞後的結果並不含詞性的標記(e.g. 「今天天氣很好」將會被斷詞成「今天 天氣 很 好」)。接著在訓練模型時，本研究將文本最大的長度設為 1,000，若文本的長度超過 1,000 時，程式將會自動截斷該文本後續的資訊。而在經過多次模型的參數調校，本研究將卷積神經網路和快速文本的嵌入層分別設為 128 與 100，而由於本研究並還未使用預訓練(pre-training)的技巧，因此嵌入層的權重將隨著模型訓練的過程中一起訓練。

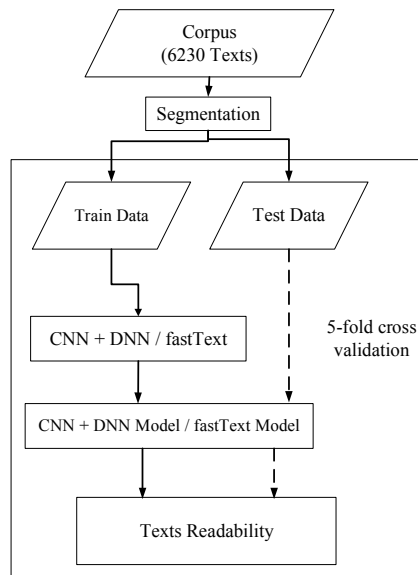


圖 5. 可讀性模型訓練及測試流程圖

[Figure 5. FlowChart of Training and Testing Readability models.]

4.3 實驗結果 (Results)

本論文分別採用卷積神經網路或快速文本來訓練可讀性模型，並與過去的研究進行比較。可讀性模型準確率如表 2，而錯誤矩陣分別如表 3 及表 4 所示。除了呈現準確率之外，本研究也放寬上、下一個年級的標準來統計出鄰近準確率，以觀察可讀性模型錯誤預測的程度是否嚴重。我們可以發現不論是以類神經網路或是快速文本來訓練可讀性模型，其準確率和鄰近準確率皆比支向量機(Support Vector Machine, SVM) (Vapnik & Chervonenkis, 1974)還好。而我們也可以發現不論是採用卷積神經網路、快速文本及

Word2vec 皆可以有效的表徵不同領域的文本來當成可讀性特徵，使得訓練出來的可讀性模型可以具有領域一般化的能力。

表2. 基於卷積神經網路及快速文本之可讀性模型效能比較
 [Table 2. Comparison Performance of Readability Models Based on Convolutional Neural Network and fastText.]

適用年級	適用領域	可讀性特徵	分類器	準確率	鄰近準確率
1-12 年級	國語科、社會科、自然科及體育和健康教育共計 6,230 篇	卷積神經網路	類神經網路 (一層)	67.62%	86.76%
			快速文本	69.63%	86.01%
		Word2vec	支向量機 (Tseng <i>et al.</i> , 2016b)	61.33%	82.2%
		Word2vec	類神經網路 (一層) (Tseng <i>et al.</i> , 2016b)	66.95%	85.26%

表3. 卷積神經網路可讀性模型之錯誤矩陣
 [Table 3. Confusion Matrices of the Convolutional Neural Network based Readability Model.]

		模型預估年級												準確率(%)
		1	2	3	4	5	6	7	8	9	10	11	12	
實際年級	1	74	52	21	0	1	0	0	0	0	0	0	1	49.66
	2	42	92	42	13	2	0	0	0	1	0	0	0	47.92
	3	6	36	197	63	17	11	1	2	1	0	0	0	58.98
	4	4	13	51	223	40	20	5	0	0	0	0	0	62.64
	5	0	1	22	54	165	88	29	4	2	3	1	1	44.59
	6	0	2	7	34	95	176	18	6	10	11	1	3	48.48
	7	2	0	1	2	16	26	526	36	26	39	1	2	77.70
	8	0	0	0	2	3	29	42	540	31	43	12	5	76.38
	9	1	0	0	0	1	23	29	29	454	41	15	2	76.30
	10	5	0	0	0	5	15	26	23	19	586	92	61	70.43
	11	0	0	0	0	2	7	7	9	9	142	630	60	72.75
	12	4	0	0	0	0	5	8	4	3	122	93	550	69.71

表 4. 快速文本可讀性模型之錯誤矩陣
[Table 4. Confusion Matrices of the fastText Readability Model.]

		模型預估年級												準確率(%)
		1	2	3	4	5	6	7	8	9	10	11	12	
實際年級	1	87	47	12	3	0	0	0	0	0	0	0	0	58.39
	2	46	97	43	6	0	0	0	0	0	0	0	0	50.52
	3	28	15	219	52	12	6	1	0	0	0	1	0	65.57
	4	10	5	62	209	33	25	10	1	1	0	0	0	58.71
	5	2	5	28	43	184	75	27	3	0	2	0	1	49.73
	6	0	4	22	23	79	188	18	5	13	7	2	2	51.79
	7	1	1	3	2	19	22	560	26	5	28	2	8	82.72
	8	10	0	0	1	12	21	28	571	12	32	7	13	80.76
	9	10	0	0	0	8	18	24	11	461	36	16	11	77.48
	10	3	4	0	1	7	11	47	39	13	555	95	57	66.71
	11	0	1	0	0	4	4	16	20	24	113	619	65	71.48
	12	1	3	0	0	1	4	11	6	4	84	87	588	74.52

此外，本論文也針對卷積類神經網路的可讀性模型去加深類神經網路的層數，以跟過去學者的結果進行比較。其結果如表 5 所示，本研究發現，以卷積類神經網路為特徵的可讀性模型，其效能並未隨著類神經分類器的層數增加而上升，且最佳的準確率低於 Word2vec 為特徵的可讀性模型 0.97%。然而，其鄰近準確率卻反而高過 0.65%。以整體而言，以 Word2vec 為特徵的可讀性模型其準確率是比較高的，但以卷積類神經網路為特徵的可讀性模型其鄰近準確率是較高的。最後綜合表 2 和表 5 而言，我們可以發現快速文本的準確率仍是所有可讀性模型中最高的，但鄰近準確率卻也是最低的。

表5. 深層類神經網路隱藏層的數量對於可讀性模型的影響
 [Table 5. The Influence of the Number of Hidden Layers of Neural Networks on the Readability Model.]

適用年級	適用領域	分類器-類神經網路層數量	可讀性特徵			
			卷積類神經網路		Word2vec (Tseng <i>et al.</i> , 2016b)	
			準確率(%)	鄰近準確率(%)	準確率(%)	鄰近準確率(%)
1-12 年級	國語、社會、自然、體育和健康教育共計 6,230 篇	1	67.62	86.76	66.95	85.26
		2	67.09	85.99	68.59	86.11
		3	66.5	86.31	68.33	85.54

5. 結論 (Conclusions and Future work)

過去可讀性模型所採用的特徵大多需要專家去設計，有著耗時費力等問題。有鑑於此，本論文基於表示學習演算法，提出以卷積類神經網路或快速文本來自動抽取文本的特徵去訓練可讀性模型，並以實證證明其效能與具領域一般化的能力。除此之外，以本研究的實驗材料而言，與支向量機同屬於淺層結構的機械學習演算法：快速文本，其效能並不輸給深層結構的機械學習演算法。針對此點發現，本研究未來將會納入更多的訓練資料及不同深、淺層結構的機械學習演算法來加以探討對於可讀性模型的影響。

除此之外，本研究也發現不同架構的可讀性模型所呈現出來的結果有很大的差異，如快速文本準確率雖然是最高的，但從表 3 和表 4 的比較可以發現，快速文本對於某些年級的文本在預測錯誤時，其錯誤誤差的程度非常嚴重；反觀卷積類神經網路預測錯誤誤差的程度就相對集中。而針對模型預測部分文本產生嚴重的誤差，本研究認為可能的原因是：對於體育和健康教育這個領域的教科書而言，為了讓國小低年級的幼童可以盡早認識與自己切身相關的知識，如：身體構造、身體自主權及生活環境、疾病...等等議題。雖然遣詞用字早就超過該年級的識字難度(以國文科相應年級課文所教授的生字而言)，但經由老師的介紹及圖片和注音的輔助，使得學生是可以理解文本的內容。相較之下，表徵學習法單純從文字所獲得的資訊就相當有限，因此當上述這些低年級的文本當成訓練資料時，一些高年級的測試資料如果用字是簡單時(如：白話文、介紹體育器材、介紹運動規則...等等)，這些文本很容易被誤判成國小低年級就可以閱讀。因此在未來的研究中，除了整合不同類型的類神經網路模型的優點來促使可讀性模型在預測錯誤時，其誤差也能夠盡可能的往適讀年級集中外；也將納入更多的特徵以輔助目前可讀性模型不足的地方。

Reference

- Abdel-Hamid, O., Deng, L. & Yu, D. (2013). Exploring convolutional neural network structures and optimization techniques for speech recognition. In *Interspeech 2013*, 3366-3370.
- Bertha, A. L. & Pressey, S. L. (1923). A method for measuring the "vocabulary burden" of textbooks. *Educational Administration and Supervision*, 9, 389-398
- Bengio, Y., Ducharme, R., Vincent, P. & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine learning research*, 3(Feb), 1137-1155.
- Borst, A., Gaudinat, A., Grabar, N. & Boyer, C. (2008). Lexically-based distinction of readability levels of health documents. *Acta Informatica Medica*, 16(2), 72-75.
- Chang, T. H., Sung, Y. T. & Lee, Y. T. (2012). A Chinese word segmentation and POS tagging system for readability research. In *Proceedings of the 42nd Annual Meeting of the Society for Computers in Psychology*.
- Chang, T. H., Sung, Y. T. & Lee, Y. T. (2013). Evaluating the difficulty of concepts on domain knowledge using latent semantic analysis. In *Proceedings of 2013 International Conference on Asian Language Processing (IALP)*, 193-196. doi: 10.1109/IALP.2013.58
- Chall, J. S. & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, Mass: Brookline Books.
- Chollet, F. (2015). Keras: Deep learning library for theano and tensorflow. URL: <https://keras.io>.
- Cireşan, D. C., Meier, U., Masci, J. & Schmidhuber, J. (2011). A committee of neural networks for traffic sign classification. In *Proceedings of The 2011 International Joint Conference on Neural Networks (IJCNN)*, 1918-1921. doi: 10.1109/IJCNN.2011.6033458
- Ciresan, D. C., Giusti, A., Gambardella, L. M. & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. In *Proceedings of the 25th International Conference on Advances in neural information processing systems(NIPS'12)*, 2843-2851.
- Cireşan, D. C., Meier, U., Gambardella, L. M. & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12), 3207-3220. doi: 10.1162/NECO_a_00052
- Collins-Thompson, K. (2014). Computational assessment of text readability: A survey of current and future research. *ITL-International Journal of Applied Linguistics*, 165(2), 97-135.
- Dale, E. & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26(1), 19-26.
- Deng, L., Abdel-Hamid, O. & Yu, D. (2013). A deep convolutional neural network using heterogeneous pooling for trading acoustic invariance with phonetic confusion. In *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6669-6673.

- Deng, L., Li, J., Huang, J. T., Yao, K., Yu, D., Seide, F., ... & Acero, A. (2013). Recent advances in deep learning for speech research at Microsoft. In *Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8604-8608.
- Feng, L., Jansche, M., Huenerfauth, M. & Elhadad, N. (2010). A comparison of features for automatic readability assessment. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10)*, 276-284.
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3), 221-233. doi: 10.1037/h0057532.
- François, T. & Miltsakaki, E. (2012). Do NLP and machine learning improve traditional readability formulas?. In *Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations (PITR '12)*, 49-57.
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning (adaptive computation and machine learning series)*. Cambridge, MA: The MIT Press.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Graesser, A. C., Singer, M. & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, 101(3), 371-395. doi: 10.1037/0033-295X.101.3.371
- Hinton, G. E. (1986). Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the cognitive science society, 1*, 1-12.
- Johnson, R. & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks. Retrieved from *arXiv preprint arXiv:1412.1058*.
- Joulin, A., Grave, E., Bojanowski, P. & Mikolov, T. (2016). Bag of tricks for efficient text classification. Retrieved from *arXiv preprint arXiv:1607.01759*.
- Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746-1751. Retrieved from *arXiv preprint arXiv:1408.5882*.
- Kireyev, K. & Landauer, T. K. (2011). Word maturity: Computational modeling of word knowledge. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 1*, 299-308.
- Klare, G. R. (1963). *Measurement of readability*. Ames, IA: Iowa State University Press.
- Klare, G. R. (2000). The measurement of readability: useful information for communicators. *ACM Journal of Computer Documentation (JCD)*, 24(3), 107-121. doi: 10.1145/344599.344630
- Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2), 211-240. doi: 10.1037/0033-295X.104.2.211

- Landauer, T. K., Foltz, P. W. & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259-284. doi: 10.1080/01638539809545028
- LeCun, Y. (1989). Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman, & L. Steels (Eds.), *Connectionism in perspective*. Zurich, Switzerland: Elsevier.
- Liu, Y. N., Chen, K. Y., Tseng, H. C. & Chen, B. (2015). A Study of Readability Prediction on Elementary and Secondary Chinese Textbooks and Excellent Extracurricular Reading Materials. In *Proceedings of the 27th Conference on Computational Linguistics and Speech Processing (ROCLING 2015)*, 71-86. [In Chinese]
- Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of reading*, 12(8), 639-646.
- Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space. Retrieved from *arXiv preprint arXiv:1301.3781*.
- Nair, V. & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807-814.
- Petersen, S. E. & Ostendorf, M. (2009). A machine learning approach to reading level assessment. *Computer speech & language*, 23(1), 89-106. doi: 10.1016/j.csl.2008.04.003
- Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1), 1929-1958.
- Sung, Y. T., Chen, J. L., Cha, J. H., Tseng, H. C., Chang, T. H. & Chang, K. E. (2015). Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning. *Behavior research methods*, 47(2), 340-354. doi: 10.3758/s13428-014-0459-x.
- Truran, M., Georg, G., Cavazza, M. & Zhou, D. (2010). Assessing the readability of clinical documents in a document engineering environment. In *Proceedings of the 10th ACM symposium on Document engineering (DocEng '10)*, 125-134. doi: 10.1145/1860559.1860585
- Tseng, H. C., Hung, H. T., Sung, Y. T. & Chen, B. (2016). Classification of Text Readability Based on Deep Neural Network and Representation Learning Techniques. In *Proceedings of 28th Conference on Computational Linguistics and Speech Processing (ROCLING 2016)*, 255-270. [In Chinese]
- Tseng, H. C., Sung, Y. T., Chen, B. & Lee, W. E. (2016). Classification of text readability based on representation learning techniques. In *Proceedings of the 26th Annual Meeting of the Society for Text & Discourse*.
- Vapnik, V. N. & Chervonenkis, A. Y. (1974). *Teoriya raspoznavaniya obrazov. Statisticheskie problemy obucheniya* (Theory of pattern recognition. Statistical problems of learning). Moscow, Russia: Nauka.

- Vogel, M. & Washburne, C. (1928). An objective method of determining grade placement of children's reading material. *The Elementary School Journal*, 28(5), 373-381.
- Yan, X., Song, D. & Li, X. (2006). Concept-based document readability in domain specific information retrieval. In *Proceedings of the 15th ACM international conference on Information and knowledge management (CIKM '06)*, 540-549. doi: 10.1145/1183614.1183692
- Zhang, Y. & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. Retrieved from *arXiv preprint arXiv:1510.03820*.

