

基於深層類神經網路及表示學習技術之文件可讀性分類

Classification of Text Readability Based on Deep Neural Network and Representation
Learning Techniques

曾厚強 Hou-Chiang Tseng

國立臺灣師範大學資訊工程學系

Department of Computer Science and Information Engineering

National Taiwan Normal University

ouartz99@gmail.com

洪孝宗 Hsiao-Tsung Hung

國立臺灣師範大學資訊工程學系

Department of Computer Science and Information Engineering

National Taiwan Normal University

alexhung@ntnu.edu.tw

宋曜廷 Yao-Ting Sung

國立臺灣師範大學教育心理與輔導學系所

Department of Educational Psychology and Counseling

National Taiwan Normal University

sungtc@ntnu.edu.tw

陳柏琳 Berlin Chen

國立臺灣師範大學資訊工程學系

Department of Computer Science and Information Engineering

National Taiwan Normal University

berlin@ntnu.edu.tw

摘要

隨著網路快速發展，帶來的是資訊流通的便利性。然而，如此爆炸性成長的網路資訊卻令人忽略最重要的根本：閱讀的材料是否能夠被理解。當閱讀材料愈能夠被讀者所理解，就愈能夠產生好的學習效果。而合理估測文本能夠被理解的程度是可讀性公式的研究目標，它主要透過分析文件上的資訊來轉化成可讀性特徵，再利用這些可讀性特徵來訓練出可讀性模型，以便能自動化預測文件的可讀性。而近年來隨著深層類神經網路的崛起，不論應用在語音辨識、圖像處理及自然語言都有明顯效能上的提升。因此，本論文將研究這個技術應用於文本可讀性這個有趣且富實際應用的議題，提出一個基於深層類神經網路技術及詞向量表示法來建構出一個能夠分析跨領域文件的可讀性模型，以符合文件

內容多元主題的特性。使可讀性模型能夠更精確分析文件的可讀性之外，也能夠兼具領域一般化的能力。

Abstract

The development of the internet has facilitated the flow of information. However, this explosive growth of information has led to fundamental importance being overlooked: Reading material can be understood. Research on readability formulas aims to predict, to a reasonable extent, the degree to which a text can be understood. It does so mainly by analyzing and translating the information within a text into readability features, which are used to train a readability model, in order to automatically predict the readability of a given text. In recent years, the development of deep neural networks, applied to speech recognition, image processing and natural language processing has improved significantly on the performance. Therefore, this paper proposes a readability model built with deep neural network and word vector representation, and which is capable of analyzing cross-domain texts, in accordance with the diverse topics of text contents. The authors aim to make the readability model capable of analyzing text readability with more accurate, as well as possess domain generalization capacity.

關鍵詞：可讀性，詞向量，分類，深層類神經網路，支向量機

Keywords: Readability, Word2vec, Classification, Deep Neural Network, Support Vector Machine.

一、緒論

可讀性(Readability)是指閱讀材料能夠被讀者理解的程度[1],[2],[3],[4]，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[2],[3]。由於文件的可讀性是如此重要，因此早在 1923 年 Lively 和 Pressey 就提出方法來探討教科書中字彙難度的問題[5]。在 1928 年 Vogel 和 Washburne 則是提出一個 Winnetka Formula 來評量小孩讀物的可讀性[6]。可讀性研究一直持續不斷的發展，據 Chall 與 Dale 在 1995 年的統計，到 1980 年為止相關的可讀性公式已經超過 200 多則文件可讀性公式[7]。這些傳統的可讀性公式大多都是使用語言特徵來評量文件的可讀性，例如：著名的 Flesch Reading Ease 公式以詞彙音節數做為語意的指標，以句子的長度作為語法的指標，計算詞彙的平均音節數與文件的平均句子長度來評估文件難度，當文件的詞彙音節數愈多、句子愈長，則該文

件愈困難[8]。Chall 和 Dale(1995)加入了「難詞比率」做為評估文件難度的方式，難詞出現愈多，表示文件愈困難[7]。至今，可讀性模型的發展依舊蓬勃發展，並隨著機器學習演算法的崛起，研究人員得以用更細緻的演算法讓可讀性模型可以納入更多元的可讀性指標，以提升模型準確率 [9],[10],[11]。

雖然機器學習演算法提升了可讀性模型的準確率，然而模型所採用的特徵仍舊為過去的一般語言特徵，而一般語言特徵只單純考量語意、語法和難詞比率等變項並不足以反映文件難度。Graesser、Singer 和 Trabasso(1994)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[12]。Collins-Thompson(2014)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[13]。此外，一般語言特徵亦無法判斷詞彙在不同領域時，其詞彙背後所代表的意義。因為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(2006)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，採用一般語言特徵的可讀性公式無法反映特定領域文件中專業術語的難度[14]。

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(2006)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 Scop)[14]。Borst 等人(2008)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[15]。Chang 等人(2013)則是將潛藏語意分析(Latent Semantic Analysis, LSA)應用在特定領域的教科書上，透過奇異值分解(Singular Value Decomposition, SVD)將維度縮減以擷取出語料庫的語意空間來表達文件潛藏語意屬性，接著再以餘弦測量(cosine measure)的方式來獲得領域文件中不同難度的概念詞彙[16]。

透過表徵特定領域的知識結構來做為文件表示(Document Representation)，雖然可以增進可讀性模型的效能。但卻也限制可讀性模型的一般化(Generalization)能力：必須先

確認文件所屬的領域後，才能使用相對應的可讀性模型來評估文件可讀性。而此限制除了將造成可讀性模型在實際上應用的問題外，如何事先對於文件所屬領域的界定更是一大考驗。例如網路文本或者課外讀物的內容非常多元，因此難以區別文件所屬的領域。所以發展一個跨領域文件的可讀性模型是必須的。

本論文的内容安排如下：第二節將描述目前可讀性模型應用於評量網路文本的相關研究，並說明其模型可改進的空間。第三節將提出基於深層類神經網路及詞向量技術來建構出一個能夠同時分析不同領域文件可讀性的模型。第四節將呈現本論文所提出模型的效能。最後第五節是總結及未來研究的方向。

二、相關研究

如同 Collins-Thompson 在 2014 所述：讀者是否能理解文件常被視為文件的重要價值之一，然而在設計網頁文本時卻常忽略這個基本的價值而導致讀者在閱讀上的困難[13]，而這也正好凸顯了文件可讀性的重要性和問題所在。在這資訊爆炸的現代，網路文本更是讀者獲取資訊的主要來源。然而目前常看到評量網頁可讀性的服務如 Readability Test [17]、The Readability Test Tool[18]依舊利用傳統的可讀性公式如 Flesch Reading Ease[8]、Flesch Kincaid Grade Level[19]及 Gunning Fog Score [20]來評量不同領域的網路文本可讀性。然而，傳統可讀性公式僅以詞彙（如音節數與難詞比率）或句子（如句長）來當作可讀性指標，但這類表淺的語言指標卻經常被質疑詞彙越長的詞不一定就越困難，並且以句子長度當成句法複雜的指標則過於直覺不夠精緻[21]。Kidwell、Lebanon 和 Collins-Thompson (2011)指出當文本字數少於 300 字時，其傳統可讀性公式是不適用的。而這個問題對於字數難以控制的網路文本而言顯然是一個重大的限制[22]。

目前已有許多研究指出傳統可讀性公式並不適合評量網頁文本的可讀性 [23],[24],[25]，因此如何才改善可讀性模型來適用於在網頁文本便是值得研究的議題。Miltsakaki 在 2007 年提出 Read-X 系統來針對不同類別網頁文本進行可讀性的評估，它是利用三種可讀性公式：Lix readability formula、Rix readability formula 和 Coleman-Liau readability formula 等來評量文件的可讀性難度，但 Read-X 系統未將三種可讀性的難度進行整合，這將造成使用者對於文件可讀性難度造成疑慮[26],[27]。直到 2009 年 Miltsakaki 利用平均將 Lix Readability Formula、Rix Readability Formula 和 Coleman-Liau Readability

Formula 的難度進行整合，然而卻也發現這樣子的整合方式無法有效區分出 9-10 年級及 11-13 年級的難度[28]。Eickhoff 等人則是藉由偵測主題的方式來區分哪種網頁適合小孩子閱讀[29]。然而此種判定文件可讀性的方式是否適用於螺旋式教學(Spiral Curriculum)的文本值得商榷。Kanungo 和 Orr(2009)則是針對搜尋引擎所搜尋出來的網頁摘要來評量可讀性，可惜所採用的可讀性特徵仍舊為一般的語言特徵，並無法表徵特定領域文件的知識結構，因此其文件可讀性結果是否能進一步對應網頁適讀的年齡則需進一步的驗證[30]。

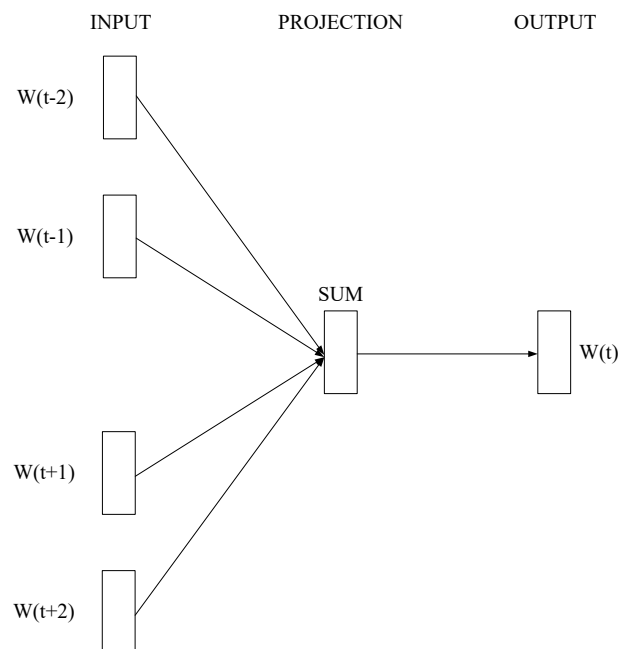
除了可讀性指標的問題之外，受益於自然語言處理技術與機械學習演算法的崛起，研究人員得以用更精緻的模型演算法來測量文本的可讀性，使可讀性模型不僅可以納入更多元的可讀性指標，並且對於模型的效能亦有明顯的提升[31],[32],[34]，而其中所採用分類的工具又以支向量機(Support Vector Machine, SVM)最為常見。然而支向量機相對於深層類神經網路(Deep Neural Network, DNN)而言，是屬於一種淺層的結構，目前已有研究證明這種淺層結構的機械學習演算法在解決簡單或限制較多的分類問題上，是可以得到不錯的效果。但是受限於模型建模和表示的能力，當處理的問題是更為複雜的情形下，就會面臨各種的困難[35]。因此，從上述的研究可以發現，將可讀性模型運用至網路文件是一個必然的趨勢，但也因為網路文件有著許多複雜且無法掌控的因素，致使可讀性模型的發展需要考慮的更加周延。因此，本論文將利用深層類神經網路及詞向量技術來建構跨領域文件的可讀性模型。

三、基於深層類神經網路及詞向量表示法之可讀性模型建立

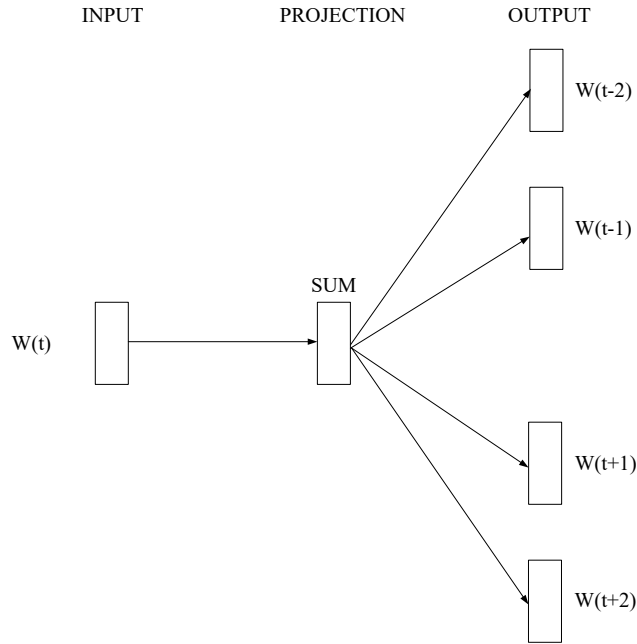
(一)、詞向量表示

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)[36]。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[37]。而近期 Google 所發表的 Word2Vec 則可視為 FFNNLM 的後繼方法[38]。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 1 和圖 2 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。然而，不論是連續詞袋模型還是略詞模型都是基於一個長度來看詞彙之間的關係，即所謂的 Shallow Window-Based 的方法。Jeffrey Pennington 則是在 2014 年提出一個 GloVe 的演算法來同時考慮全域及區域詞彙之間的關係，以提升 Word Embedding 的效果[39]。而根據過去的實驗，連續詞袋模型、略詞模型及 GloVe 在可讀性研究的效能差異不大的情況下，本論文將基於連續詞袋模型詞向量表示方式來搭配深層類神經網路建構出一個跨領域文件的可讀性模型[40]。



圖一、連續詞袋模型訓練演算法



圖二、略詞模型訓練演算法

(二)、深層類神經網路

近年來深層學習的相關研究將類神經網路成功應用在各領域，主要原因為一系列的訓練演算法及模型架構被提出，克服傳統的多層類神經網路訓練困難問題。例如利用 Deep Belief Network(DBN)來初始化深層類神經網路參數，常常會得到比隨機初始化的方法來得到更好的結果[35],[41]。又如 Hinton 利用 rectified linear units (ReLU) 作為隱藏層的激發函數，以克服當網路在兩個方向都接近飽和時，造成梯度變化很小，整個網路的學習變的很慢[35],[42]。本論文的模型利用 ReLU 作為隱藏層的激發函數。此函數可以表示為 $\max(0, x)$ ，即保留正數部分。由於其導函數為常數，可以避免典型的梯度消失(gradient vanish)問題。本論文採用網路模型架構為 1024 維 ReLU 隱藏層和 12 維 softmax 作為輸出層，預測文本屬於各年級教材機率。我們已知課程難易度是漸進關係，隨著年級提升而內容更複雜，但典型的類神經網路沒有考量此因素。因此，我們提出目標函數應加入一正則項，滿足前述的觀察。令輸出層網路參數為矩陣 W^L ，可以表達為 12 個行項量 (column vectors)，分別對應各年級，如下式(1)：

$$W^L = |\mathbf{w}_1^L \mathbf{w}_2^L \dots \mathbf{w}_{12}^L| \quad (1)$$

本論文提出的方法為限制鄰近年級的行項量距離，相鄰的年級應該有相近的向量，可以自然地表達年級的連續性。此正則項可以表達為下式(2)：

$$R(W^L) = \sum_{i=1}^{11} \|\mathbf{w}_i^L - \mathbf{w}_{i+1}^L\| \quad (2)$$

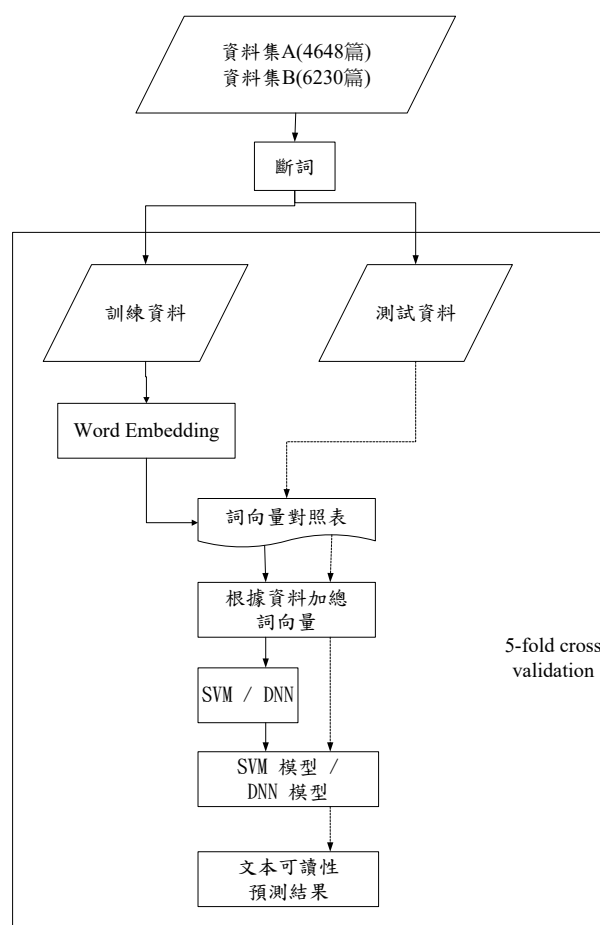
我們希望此正則項可以限制鄰近年級向量的距離，而完整的減損函數為交叉熵(cross entropy)和正則項的結合，如下式(3)：

$$L(\theta) = -\sum_{i=1}^{12} (y_i - \log v_i^L) - \sum_{j=1}^{11} \|\mathbf{w}_j - \mathbf{w}_{j+1}\| \quad (3)$$

其中 y_i 為年級標記， v_i^L 為模型輸出年級 i 的機率。我們藉此鼓勵輸出層參數呈現 12 年級的流形(manifold)，即相鄰兩個年級彼此在輸出層空間中相近。

(三)、訓練可讀性模型

建置跨領域文件的可讀性模型流程如圖 3 所示，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。本研究實驗拆成兩個資料集分別為：一、資料集 A：由國語科、社會科及自然科等三個領域的教科書共計 4,648 篇。二、資料集 B：由國語科、社會科、自然科及體育和健康教育等四個領域的教科書共計 6,230 篇。藉此觀察在資料集愈複雜的情況下，對於支向量機及深層類神經網路所造成的影響為何。而整實驗的流程皆採用 5-fold 交互驗證的方式進行，首先將文件利用 WECA[43]來進行中文文件的斷詞，再將訓練資料利用 Word2Vec[38]來分別得到連續詞袋模型詞向量對照表。接著將訓練資料的每一篇課文依據使用到的詞彙從詞向量對照表中取出向量，並將這些向量全部相加，最後所得到的向量便是這一篇課文的可讀性特徵，而它的類別就是課文所屬的年級。本研究分別利用 Keras [44] 和 LIBSVM[45]來訓練出深層類神經網路及支向量機可讀性模型。在驗證可讀性模型的階段，本研究將測試資料使用到的詞彙一樣從詞向量對照表取出向量，並將這些向量全部相加，如遇到詞向量對照表沒有的詞彙時，則不處理該詞彙。在取得測試資料的可讀性特徵後，便可輸入至已訓練好的可讀性模型來預測文件的年級值。



圖三、可讀性模型訓練及測試流程圖

四、實驗結果

本研究的實驗結果如表一和表二所示，而四種可讀性模型的錯誤矩陣分別如表三、表四、表五及表六。從結果可以發現不論是在三種領域文本還是四種領域文本的情況下，深層類神經網路的準確率都優於支向量機。然而我們也不難發現在加入 1,582 篇的體育和健康教育後，模型分類的難度大幅度的上升。對支向量機模型而言，準確率減少了 9.5% 的準確率，而深層類神經網路只有減少了 7.32% 的準確率。這顯示在文本可讀性分類的這個研究領域中，深層類神經網路比支向量機更能夠處理更為複雜的資料。

表一、實驗一：連續詞袋模型 100 維度之三種領域文本效能比較

適用年級	適用領域	分類演算法	準確率(%)
1-12 年級	國語、社會、自然共計 4,648 篇	支向量機	70.83
		深層類神經網路	74.27

表二、實驗二：連續詞袋模型 100 維度之四種領域文本效能比較

適用年級	適用領域	分類演算法	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	支向量機	61.33
		深層類神經網路	66.95

表三、實驗一支向量機模型之錯誤矩陣

		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	17	7	0	0	0	0	0	0	0	0	0	0	70.83%
	2	2	60	5	0	0	0	0	0	0	0	0	0	89.55%
	3	1	6	156	35	6	5	4	0	0	0	0	0	73.24%
	4	0	3	46	124	17	11	7	0	1	2	1	0	58.49%
	5	0	0	11	27	119	28	26	2	2	2	4	0	53.85%
	6	0	1	16	16	33	95	24	9	16	3	0	0	44.60%
	7	0	0	3	0	12	7	499	28	9	20	11	9	83.44%
	8	0	0	0	1	5	4	39	502	18	30	7	10	81.49%
	9	0	0	0	0	3	10	32	18	377	43	25	2	73.92%
	10	1	2	2	2	1	3	35	26	30	404	76	53	63.62%
	11	0	0	0	0	5	0	22	15	37	69	502	77	69.05%
	12	0	0	1	1	1	0	8	10	7	59	71	454	74.18%

表四、實驗一深層類神經網路模型之錯誤矩陣

		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	3	15	5	0	0	0	0	1	0	0	0	0	12.50%
	2	1	45	13	6	1	0	0	1	0	0	0	0	67.16%
	3	1	16	143	26	10	10	4	2	0	1	0	0	67.14%
	4	0	4	29	133	25	14	5	0	0	1	0	1	62.74%
	5	0	3	7	30	113	40	17	4	4	3	0	0	51.13%
	6	0	2	11	15	37	107	11	8	18	4	0	0	50.23%
	7	0	0	2	0	7	8	527	13	9	20	8	4	88.13%
	8	0	1	1	2	2	6	23	528	15	21	12	5	85.71%
	9	0	0	0	0	3	5	14	17	424	23	23	1	83.14%
	10	0	1	2	1	4	3	29	30	23	434	71	37	68.35%
	11	0	0	0	0	3	0	12	17	27	82	537	49	73.87%
	12	0	0	1	0	1	1	9	8	7	57	70	458	74.84%

表五、實驗二支向量機模型之錯誤矩陣

		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	106	27	7	5	0	0	0	0	0	2	0	2	71.14%
	2	65	86	27	13	0	0	0	0	0	1	0	0	44.79%
	3	34	30	172	65	12	12	5	0	0	3	0	1	51.50%
	4	18	14	66	179	42	19	9	1	1	7	0	0	50.28%
	5	2	1	35	59	142	64	34	6	3	13	6	5	38.38%
	6	1	2	19	37	72	147	34	10	23	14	0	4	40.50%
	7	2	0	2	0	18	19	513	40	11	44	14	14	75.78%
	8	9	0	0	1	3	6	57	527	24	55	7	18	74.54%
	9	10	0	0	1	3	15	32	30	383	74	33	14	64.37%
	10	2	3	3	1	5	5	35	33	32	520	116	77	62.50%
	11	0	0	0	3	7	4	15	17	37	148	530	105	61.20%
	12	4	0	0	0	5	1	19	9	3	128	104	516	65.40%

表六、實驗二深層類神經網路模型之錯誤矩陣

		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	80	49	15	1	1	1	1	0	0	0	0	1	53.69%
	2	37	101	39	7	4	3	0	0	0	0	1	0	52.60%
	3	13	42	191	56	13	12	2	1	0	2	2	0	57.19%
	4	6	19	66	185	48	20	11	0	0	1	0	0	51.97%
	5	0	3	36	50	168	66	20	6	3	7	8	3	45.41%
	6	1	4	24	29	65	158	23	16	22	16	4	1	43.53%
	7	0	1	1	3	12	18	553	35	8	29	11	6	81.68%
	8	2	1	0	3	9	14	45	552	23	35	16	7	78.08%
	9	2	1	1	0	4	16	29	25	449	38	22	8	75.46%
	10	1	2	1	1	6	6	37	36	27	545	113	57	65.50%
	11	0	0	0	1	5	1	18	21	42	102	611	65	70.55%
	12	2	0	0	0	0	0	9	7	11	73	109	578	73.26%

除此之外，本研究也比較深層類神經網路輸出層在考量教科書年級的 manifold 的情況下對於模型準確率的影響，其結果如表七所示。從結果可以發現，在考慮的情況下，可以小幅度的提升可讀性模型的準確率，而這樣的考量也凸顯文本的可讀性分類，其類別之間可能是存在著關係，而這樣子的關係對於分類準確率是有所幫助的。

表七、實驗三：深層類神經網路考量教科書年級的 manifold 對於模型準確率的影響

適用年級	適用領域	是否考慮教科書年級 manifold	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	考慮	66.95
		不考慮	66.48

最後，本研究所要探討的是深層類神經網路在考慮教科書年級的 manifold 的情況下，隱藏層的層數多寡是否會對可讀性模型的準確率造成影響。其結果如表八所示，我們可以發現二層與三層的深層類神經網路的準確率會比單層的好，但卻不保證一直增加隱藏層的數量，其模型的準確率能夠穩定向上提升。雖然如此，若跟過去的研究相比如表九所示，本研究也成功維持相近準確率的情況下，將可讀性模型的領域一般化能力又多拓展了一個領域。在未來本研究可以往更精緻的深層類神經網路來發展，例如用長短期記憶深層類神經網路(Long-Short Term Memory, LSTM)來模擬人閱讀文本順序時，這種時間序列對於文本可讀性的影響。

表八、實驗四：深層類神經網路隱藏層的數量對於可讀性模型的影響

適用年級	適用領域	隱藏層數量	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	66.95
		2	68.59
		3	68.33

表五、實驗五：與相關研究的可讀性模型之效能比較

	適用年級	適用領域	使用特徵	分類演算法	準確率(%)
[40]	1-12 年級	國語、社會、自然共計 4,648 篇	GloVe 100 維	支向量機	68.33
本論文				深層類神經網路	68.59

五、結論

過去可讀性模型在發展的過程中，從一開始的線性迴歸式到現在運用許多非線性的機械學習分類演算法來增進模型的效能，而其中又以支向量機最廣為研究人員所喜愛。然而，從本實驗的結果可以發現，在相同的條件下，當分類資料更為複雜時，支向量機這種淺層結構的機械學習演算法的效能會快速下降。有鑑於此，本論文基於深層的類神經網路提出一個訓練可讀性模型的架構，除了預測效能能夠更勝支向量機外，所訓練出來的可讀性模型也都能夠適用於多個領域的文本。

在未來，本研究將提升詞向量的維度來增進可讀性模型的效能外，並將整合更多元的指標，如 Sentence Embedding[46]來讓文件的表達(Document Representation)可以更加的精確和有意義外，也希望能整合 Sung 等人所提出 Multilevel Linguistic Features 的概念[11]，讓可讀性模型能夠從更多元的角度來考量文件的可讀性，使評量的結果能夠更加的客觀和精確。

參考文獻

- [1]E. Dale and J. S. Chall, "The concept of readability," *Elementary English*, vol. 26, pp. 19–26, 1949.
- [2]G. R. Klare, *Measurement of Readability*, 1963.
- [3]G. R. Klare, "The measurement of readability: useful information for communicators," *ACM Journal of Computer Documentation (JCD)*, vol. 24, pp. 107-121, 2000.
- [4]G. H. McLaughlin, "SMOG grading: A new readability formula," *Journal of reading*, vol. 12, pp. 639–646, 1969.
- [5]B. A. Lively and S. L. Pressey, "A method for measuring the vocabulary burden of textbooks," *Educational administration and supervision*, vol. 9, pp. 389–398, 1923.
- [6]M. Vogel and C. Washburne, "An objective method of determining grade placement of children's reading material," *The Elementary School Journal*, pp. 373–381, 1928.
- [7]J. S. Chall and E. Dale, *Readability Revisited: The new Dale-Chall Readability Formula*, Brookline Books, 1995.
- [8]R. Flesch, "A new readability yardstick," *Journal of applied psychology*, vol. 32, p. 221, 1948.
- [9]S. E. Petersen and M. Ostendorf, "A machine learning approach to reading level assessment," *Computer Speech & Language*, vol. 23, pp. 89–106, 2009.

- [10] L. Feng, M. Jansche, M. Huenerfauth, and N. Elhadad, "A comparison of features for automatic readability assessment," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 2010, pp. 276–284.
- [11] Y.-T. Sung, J.-L. Chen, J.-H. Cha, H.-C. Tseng, T.-H. Chang, and K.-E. Chang, "Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning," *Behavior research methods*, vol. 47, pp. 340–354, 2014.
- [12] A. C. Graesser, M. Singer, and T. Trabasso, "Constructing inferences during narrative text comprehension," *Psychological Review*, vol. 101, p. 371, 1994.
- [13] K. Collins-Thompson, "Computational assessment of text readability: A survey of current and future research," *International Journal of Applied Linguistics*, vol. 165, pp. 97–135, 2014.
- [14] X. Yan, D. Song, and X. Li, "Concept-based document readability in domain specific information retrieval," in *Proceedings of the 15th ACM international conference on Information and knowledge management*, 2006, pp. 540–549.
- [15] A. Borst, A. Gaudinat, C. Boyer, and N. Grabar, "Lexically based distinction of readability levels of health documents," *Acta Informatica Medica*, vol. 16, pp. 72–75, 2008.
- [16] T.-H. Chang, Y.-T. Sung, and Y.-T. Lee, "Evaluating the difficulty of concepts on domain knowledge using latent semantic analysis," in *Proceedings of the International Conference on Asian Language Processing (IALP)*, pp. 193–196, 2013.
- [17] Juicy Studio. (2015). *Readability Test*. Retrieved from <http://juicystudio.com/services/readability.php>.
- [18] David Simpson. (2015). *The Readability Test Tool*. Retrieved from <http://read-able.com/>.
- [19] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," *DTIC Document*, 1975.
- [20] R. Gunning, *The Technique of Clear Writing*, 1952.
- [21] A. Bailin and A. Grafstein, "The linguistic assumptions underlying readability formulae: A critique," *Language & Communication*, vol. 21, pp. 285–301, 2001.
- [22] P. Kidwell, G. Lebanon, and K. Collins-Thompson, "Statistical estimation of word acquisition with application to readability prediction," *Journal of the American Statistical Association*, vol. 106, pp. 21–30, 2011.
- [23] L. Si and J. Callan, "A statistical model for scientific readability," in *Proceedings of the International Conference on Information and Knowledge Management*, pp. 574–576, 2001.
- [24] K. Collins-Thompson and J. P. Callan, "A language modeling approach to predicting reading difficulty," in *Proceedings of HLT-NAACL*, pp. 193–200, 2004.

- [25] L. Feng, N. Elhadad, and M. Huenerfauth, “Cognitively motivated features for readability assessment,” in *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 229–237, 2009.
- [26] E. Miltsakaki and A. Troutt, “Read-x: Automatic evaluation of reading difficulty of web text,” in *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pp. 7280–7286, 2007.
- [27] E. Miltsakaki and A. Troutt, “Real-time web text classification and analysis of reading difficulty,” in *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, 2008, pp. 89–97.
- [28] E. Miltsakaki, “Matching readers' preferences and reading skills with appropriate web texts,” in *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session*, pp. 49–52, 2009.
- [29] C. Eickhoff, P. Serdyukov, and A. P. de Vries, “Web page classification on child suitability,” in *Proceedings of the ACM International Conference on Information and Knowledge Management*, pp. 1425–1428, 2010.
- [30] T. Kanungo and D. Orr, “Predicting the readability of short web summaries,” in *Proceedings of the ACM International Conference on Web Search and Data Mining*, pp. 202–211, 2009.
- [31] S. Petersen and M. Ostendorf, “A machine learning approach to reading level assessment,” *Computer Speech & Language*, vol. 23, no. 1, pp. 89–106, 2009.
- [32] L. Feng, M. Jansche, M. Huenerfauth, and N. Elhadad, “A comparison of features for automatic readability assessment,” In *Proceedings of the International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, pp. 276–284, 2010.
- [33] T. François and E. Miltsakaki, “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*. Association for Computational Linguistics, pp. 49-57, 2012.
- [34] Y. Sung, J. Chen, J. Cha, H. Tseng, T. Chang and K. Chang, “Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning,” *Behavior research methods*, vol. 47, no. 2, pp. 340–354, 2014.
- [35] L. Deng, and D. Yu, “Deep Learning Methods and Applications,” vol. 7, no. 3–4, pp 197–387, *Foundations and Trends in Signal Processing*, 2014.
- [36] G. E. Hinton, “Learning distributed representations of concepts,” in *Proceedings of the Annual Conference of the Cognitive Science Society*, pp. 1–12, 1986.
- [37] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, “A neural probabilistic language model,” *The Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

- [38] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [39] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the Empirical Methods in Natural Language Processing*, vol. 12, pp. 1532–1543, 2014.
- [40] H. C. Tseng, Y. T. Sung, B. Chen, W. E. Lee, “Classification of text readability based on representation learning techniques,” in *Proceedings of the Annual Meeting of the Society for Text & Discourse, kassel*, 2016.
- [41] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, and T. N. Sainath, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine*, vol. 29, pp. 82–97, 2012.
- [42] Nair, Vinod, and Geoffrey E. Hinton. “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the International Conference on Machine Learning*, pp. 807–814. 2010.
- [43] T. H. Chang, Y. T. Sung, and Y. T. Lee, “A Chinese word segmentation and POS tagging system for readability research.” in *Proceedings of the Annual Meeting of the Society for Computers in Psychology*, 2012.
- [44] "Keras", Available at: <https://github.com/fchollet/keras>.
- [45] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27, 2011.
- [46] Q. V. Le and T. Mikolov, “Distributed representations of sentences and documents,” *arXiv preprint arXiv:1405.4053*, 2014.