

## 應用詞向量模型於日文單詞可讀性評估之研究

### Japanese Word Readability Assessment using Word Embeddings

楊正仁<sup>1</sup>, 林淑璋<sup>2</sup>, 薛芸如<sup>2</sup>, 鄭靖潔<sup>1</sup>, 王纘綦<sup>1</sup>, 黃依賢<sup>1</sup>

Cheng-Zen Yang<sup>1</sup>, Shu-Chang Lin<sup>2</sup>, Yun-Ju Hsueh<sup>2</sup>, Ching-Chieh Cheng<sup>1</sup>, Yih-Chen Wang<sup>1</sup>,  
I-Shyan Hwang<sup>1</sup>

<sup>1</sup>元智大學資訊工程學系

<sup>2</sup>元智大學應用外語學系

<sup>1</sup>Department of Computer Science & Engineering

<sup>2</sup>Department of Foreign Languages and Applied Linguistics

Yuan Ze University

Email: {czyang,shuchang,connie,ishwang}@saturn.yzu.edu.tw

{s1051530, s1051531}@mail.yzu.edu.tw

#### 摘要

在文本分析上，文本可讀性多年來都是一項重要的研究議題。然而過往研究大多偏向討論文章的可讀性，鮮少有研究討論單詞的可讀性。在單詞可讀性自動化評估問題上，最大的挑戰來自於單詞沒有豐富的可讀性特徵，不像一般文本有語句長度、音節多寡等特徵來辨別文本可讀性，因此難以利用機器學習技術來協助人工評估。然而由於詞向量技術的進步，在本論文中，我們提出一個新的單詞可讀性評估方式，使用詞向量技術將單詞取出對應的詞向量特徵，然後利用詞向量的語意相似關係，以最近鄰居法建構一個單詞可讀性評估分類模型WR-kNN。由於單詞可讀性語料的貧瘠，本研究以目前可獲得的日文單詞可讀性來進行探討。實驗結果顯示，WR-kNN的單詞可讀性分類準確率最多可以達到42.50%，比基於單詞上下文周邊詞的計算方法可以得到2.51倍的改善。

#### Abstract

In text analysis, text readability has been an important research topic for many years. However, most studies focus on the document readability rather than the word readability. To decide the readability levels of words, linguists need to spend a large amount of human effort and assessment time. The most challenging problem faced in the task of automatic word readability assessment is the lack of research on readability for each word. In this study, we propose a novel assessment model for word readability called WR-kNN based on the word embedding technology by calculating the word vectors

and classifying them with a k-nearest neighbor model. Since the resource of word readability corpus is rare, this study uses a Japanese corpus to discuss. The experimental results show that WR-kNN can predict the word readability with 42.50% accuracy. Compared with a co-occurrence-based approach, WR-kNN can achieve an improvement by a factor of 2.51.

關鍵字：單詞可讀性，詞向量，自然語言處理，效能評估

Keyword: Word Readability, Word Embedding, Natural Language Processing, performance evaluation

## 一、緒論

在文本分析上，可讀性（Readability）一直是許多研究所探討的對象[1,2]。Geroge Klare [1] 指出，可讀性的重要是因為它能夠用來 (1)表示文件手寫或排版的易讀性（Legibility）；(2)表示由於閱讀興趣或寫作生動而易於閱讀的程度；(3)表示由於寫作風格而易於理解或理解的程度。可讀性高的文章就容易被讀者理解，因此文本可讀性的評估對於出版或教育等皆具有重要的意義。自 20 世紀初開始，就有許多研究對文本可讀性進行探討，並且紛紛提出不同的評估模型，用來分析文章的可讀性，例如 [3,4,5,6,14,15,16,17,18,19]。

在這些文本可讀性的研究中，考慮了許多文本以及語言特性，例如語句長度、英文音節（Syllable）個數、難詞的詞數，中文不同筆劃的詞數、日文和語詞數、外來語等特徵。雖然有些研究會對單詞的難易度加以考慮，然而過往對於單詞可讀性的研究卻相當缺乏。一些文本可讀性研究曾考慮單詞可讀性的難易度，但是它們只有單純考慮單詞在語料庫當中出現的頻率(The frequency of words)或是考慮音節個數等單字特徵。例如在詞頻上，Lively與Pressey是以Thorndike編著的常用詞彙表來進行研究[3]，劉憶年等人也考慮文章中難詞（不在常用詞表的詞）出現次數[16]。以上這些研究都沒有再深入探討單詞可讀性。從這些研究中也顯示單詞可讀性對於文本可讀性的判斷會有影響。如果單詞可讀性語料具有一定規模，這些資訊將可以有效被運用來進行文本可讀性的評估。

雖然單詞可讀性對於文本可讀性評估有其重要性，但目前缺乏相關研究，所能獲得單詞可讀性語料規模也有限。就目前所搜尋的相關文獻所及，我們只看到Sunakawa等人發表的「日本語學習詞書」(Japanese Language Learners' Dictionary)[7]曾標註日文單詞的

テキスト詳細

結果保存(CSV:Shift-JIS) 結果保存(CSV:UTF-8)

総文数:4 文の平均語数:43.25

色の付いた語をクリックすると辞書引きを行います。

内容語レベル:  初級前半  初級後半  中級前半  中級後半  上級前半  上級後半

機能語レベル:  初級  初中級  中級

- 1 岐阜県は15日、県農業大学校(可児市)で飼育している実習用のメスの親豚から豚(とん)コレラの陽性反応が出たと発表した。
- 2 岐阜県内の飼育施設での豚コレラ確認は5例目、県立施設での確認は今年5日の県畜産研究所(美濃加茂市)に次いで2例目。
- 3 県によると、感染した豚から検出されたウイルスを農業・食品産業技術総合研究機構で精密検査したところ、これまでに県内で感染が確認されたウイルスと同一のものと確認された。
- 4 畜産研究所の感染が判明したため7日に検査したが、この時は同大学校で飼育していた13頭全てが...

圖 1、日本毎日新聞社有關豬瘟的新聞文章，經由 Jreadability 分析。

可讀性。他們將日文單詞可讀性分成6個級別：初級前半，初級後半，中級前半，中級後半，上級前半，上級後半。Jreadability (<http://jreadability.net/>) [13]日文文本可讀性分析網站便使用這些語料來標註日文單詞的可讀性。

雖然「日本語學習詞書」提供了可讀性語料，但是其中的單詞數量仍然有限。如圖 1 所顯示，一篇有關豬瘟的新聞「豚コレラ、岐阜で5例目確認 県農業大学校」(<https://mainichi.jp/articles/20181216/k00/00m/040/012000c>)經過Jreadability分析，可以發現許多字詞依然沒有可讀性級別。在圖1的例子中，總共173個單詞的84個相異詞型 (Word Type) 當中，有26個詞型沒有可讀性級別。我們在分析其他文章的時候，發現甚至可以高達35%以上的單詞沒有可讀性級別。

在單詞可讀性自動化評估問題上，最大的挑戰來自於過往缺乏單詞可讀性特徵的研究，以至於缺乏可讀性特徵。不像一般文本有語句長度、音節多寡等特徵來辨別文本可讀性，因此過往是以人工研讀文本的方式來決定單詞可讀性，沒有機器學習的人工智慧方式來判斷。由於單詞數量龐大，人工研讀文本會耗費大量人力與時間。更因為目前網路時代造成資訊爆炸，每年都有不少新的詞彙產生。這些數量可觀的新詞彙將使單詞可

讀性評估更具挑戰。

最近在單詞特徵技術上，Mikolov等人所提出的詞向量（亦稱之為詞嵌入，Word Embedding）技術[8]，為單詞的特徵表示方式開啟新的里程碑。由於詞向量技術的發展，在本研究中我們提出一個新的單詞可讀性評估方式，使用詞向量技術從所蒐集的語料當中計算出單詞的詞向量特徵。利用詞向量特徵表現出單詞在文本當中的語意特性，在語意空間中找出語意關係相似的詞向量，以最近鄰居法(k-Nearest Neighbor, kNN) [9]建構一個單詞可讀性評估分類模型稱之為WR-kNN (Word Readability kNN)，從這些相似的詞向量的可讀性等級來預測單詞的可讀性。

由於過往缺乏單詞可讀性的研究，本研究的主要貢獻在於首先針對單詞可讀性，以詞向量技術來建構機器學習模型，進行單詞可讀性預測。這個機器學習模型將可以協助語言學專家判讀單詞的可讀性級別，同時可以進而運用在文本可讀性判讀的相關技術中，提昇文本可讀性的評估。

然而由於單詞可讀性語料稀少，因此本研究以Jreadability與「日本語學習詞書」標註的日文單詞可讀性來探討。我們蒐集了福娘童話集(<http://hukumusume.com/douwa/>)、青空文庫(<https://www.aozora.gr.jp/>)、及IEICE日文期刊論文摘要(<https://www.ieice.org/>)進行詞向量計算。實驗結果顯示，WR-kNN的單詞可讀性分類準確率最多可以達到42.50%，比基於單詞上下文周邊詞的計算方法可以得到2.51倍的改善。

本論文其餘內容安排如下：第二節將介紹最近的可讀性相關研究。第三節將說明基於詞向量的最近鄰居法單詞可讀性評估模型WR-kNN的設計。第四節將說明實驗環境設定、蒐集的文本語料、以及實驗的結果。第五節是本論文的結論。

## 二、 相關研究

目前關於可讀性的研究，幾乎都集中在文本可讀性的探討，例如在經驗法則的設計上，1921年時Kitson就使用單詞的音節數與句子長度來討論英文新聞文章的可讀性[1]。Lively與Pressey在1923年對英文文章提出一個可讀性評估公式來計算詞彙負擔

(Vocabulary Burden) [3]。1948年Flesch 發表著名的Flesch Reading Ease公式[4]。在中文文章的可讀性評估中，楊孝滌[14]與陳世敏[15]分別提出不同的迴歸方程式來計算。在日文文章的可讀性評估研究中，Tateisi等人在1988年使用Principal Component Analysis (PCA)分析文章的10種特性，提出一個可讀性評估公式[5]。2008年時Sato等人以教科書為語料庫，提出一個機率式語言模型來評估文章的可讀性[6]。然而這些研究並沒有涉及單詞可讀性的討論。

隨著機器學習技術以及詞向量技術的發展，劉憶年等人對於中小學國語文教材，利用詞向量技術結合逐步迴歸以及支援向量機，建構中文文本可讀性分析模型[16]。曾厚強等人並提出快速文本(fastText)與卷積神經網路(Convolutional Neural Network, CNN)來建構中文文章的可讀性模型[17,18]。曾厚強等人並於2019年提出以BERT (Bidirectional Encoder Representation from Transformers)技術[10]，利用WECA[11]來進行中文文本可讀性評估的研究[12]。即便如此，單詞可讀性也沒有機器學習模型的研發。

在單詞可讀性的研究上，Sunakawa等人在「日本語學習詞書」(Japanese Language Learners' Dictionary)[7]曾對日文單詞進行探討。他們邀請五位日語專業教師以人工方式

表 1、「日本語學習詞書」的單詞可讀性級別屬性範例。單詞的可讀性級別（語彙難易度）分為 6 種等級。取自[7]。

語彙 ID	標準的表記	読み	語彙難易度	品詞	語種	旧試験語彙級	意味分類	アクセント情報
10	アート	アート	中級前半	名詞-普通 名詞-一般	外来語		体-活動-芸術・美術	1
40	アイスコーヒー	アイスコーヒー	初級前半	名詞-普通 名詞-一般 and 名詞-普通 名詞-一般	外来語		体-生産物-食料-飲料・たばこ	6
109	明かり	アカリ	中級前半	名詞-普通 名詞-一般	和語	2級	体-生産物-機械-灯火   体-自然-自然-光	0
222	足掛かり	アシガカリ	上級後半	名詞-普通 名詞-一般	和語		体-關係-空間-点	3
294	厚かましい	アツカマシイ	中級後半	形容詞-一般	和語	2級	相-活動-心-自信・誇り・恥・反省	5
262	温まる	アタタマル	中級前半	動詞-一般	和語	2級	用-自然-物質-熱	4

檢查18010個單詞，他們邀請五位日語專業教師以人工方式檢查18010個單詞，將它們分成6個可讀性級別：初級前半，初級後半，中級前半，中級後半，上級前半，上級後半，如表1所示。基於「日本語學習詞書」的開發，這些資料後來彙整為日本語教育語彙表 (<http://jisho.jpn.org/>)，運用在由Hasebe等人所研發的日語可讀性評估工具Jreadability [13]文本可讀性分析網站之中。就目前文獻回顧所及，目前還沒有國內外研究針對單詞可讀性的機器學習模型進行討論。因此在本研究中，我們將探討單詞可讀性評估分類模型的建構。

### 三、 研究方法

在本論文中我們提出一個基於詞向量技術與kNN的單詞可讀性分類模型WR-kNN。這一節將說明WR-kNN的設計。首先將說明WR-kNN的處理流程以及各個處理部份的方法。接下來說明實驗中用來作為效能比較baseline的基於單詞上下文分類的作法。

#### (一) 單詞可讀性分類模型

針對一個單詞  $t_i$ ，以及一個具有  $m$  個可讀性級別的分類集合  $L = \{l_1, l_2, \dots, l_m\}$ ，本研究定義單詞可讀性評估為一個分類問題，要找到  $t_i$  的可讀性級別  $l_i$  如下：

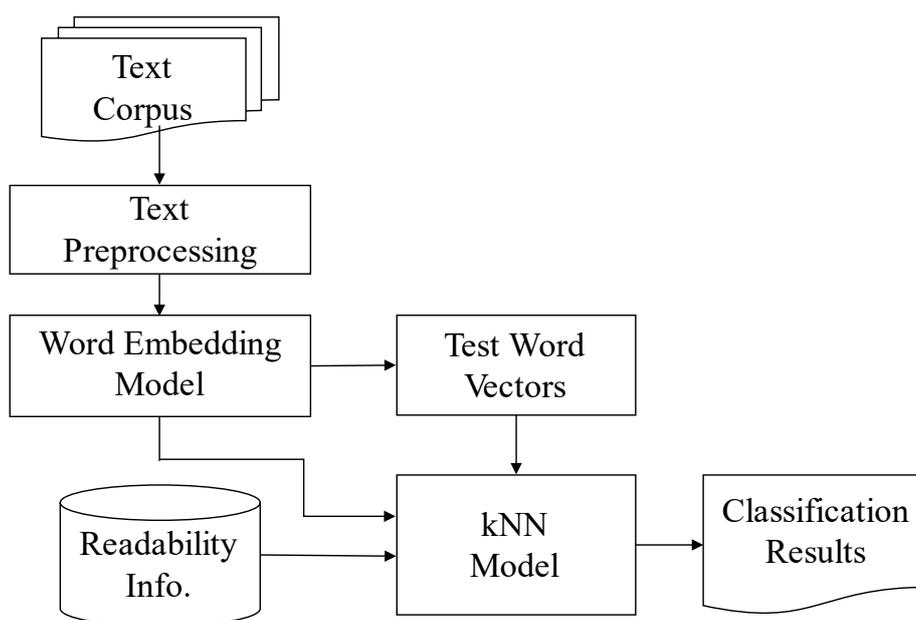


圖2、單詞可讀性分類模型。

$$l_i = \underset{L}{\operatorname{argmax}} P(l_x | t_i)$$

其中 $P(l_x | t_i)$ 是 $t_i$ 的可讀性級別為 $l_x$ 的機率。

圖2是本研究所建構的單詞分類模型架構。以下將說明文本前處理方式、詞向量生成方式、以及kNN分類器處理方式。

## (二) 文本前處理

對於日文文本，首先將文本中的一些標點符號去除，並將一些文本中的註解消去，以免影響文本內容的品質。接著進行斷詞，我們使用 MeCab (<http://taku910.github.io/mecab/>)與Unidic字典對日語文本進行斷詞以及相關詞性分析。我們取得MeCab斷詞結果以及相關單詞資訊，包括：表層形、詞性、詞性細分類1、讀音等資訊。圖3是一個文本前處理的範例，顯示句子「居留地女の間では」經過斷詞後的結果。

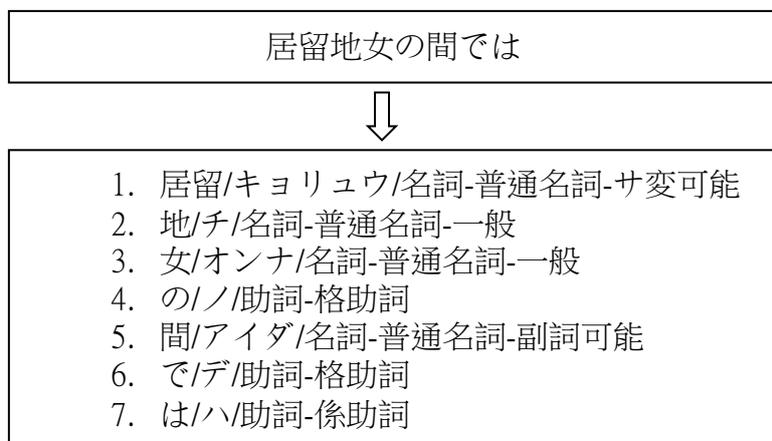


圖3、日文斷詞範例。

## (三) 詞向量計算

透過文本中的上下文關係，單詞可以經由N-gram語言模型(Language Model)計算出單詞的詞向量。在本研究中，我們使用Mikolov等人提出的Word2vec模型[8]來計算出單詞的詞向量。

Word2vec詞向量模型有兩種不同的架構，分別是Continuous Bag-of-Words model (CBOW)與Skip-Gram model (SG)。在CBOW模型中，對於目標字詞 $w(t)$ 的詞向量，會將

表2、和「居留」語意最相似的5個單詞。

單詞	相似度	可讀性級別
浴室/ヨクシツ/名詞-普通名詞-一般	0.7632547616958618	中級後半
いかり/イカリ/名詞-普通名詞-一般	0.7366321682929993	上級後半
桜/サクラ/名詞-普通名詞-一般	0.7297971248626709	初級後半
停車/テイシャ/名詞-普通名詞-サ変可能	0.7289541363716125	中級後半
倚つ/ヨル/動詞-一般	0.7197038531303406	中級後半

上下文其他字詞的詞向量經過淺層神經網路加以計算，得出 $w(t)$ 的詞向量。在SG模式中，則剛好相反，利用 $w(t)$ 詞向量，來計算 $w(t)$ 上下文字詞的機率分佈而得出詞向量。經由詞向量計算，每個詞都映射到 $D$ 維度的語意空間。在這個語意空間中，呈現了單詞的語意特性。

在本研究中，單詞都經過Word2vec的計算得出它的詞向量。利用詞向量的特性，可以找出語意相似的單詞。表2是一個利用詞向量來找出與「居留」語意相似單詞的例子。在這個例子中，我們透過Cosine similarity的計算，從目前實驗的文本中找出與語意「居留」最相似的五個單詞。表中也列出這些單詞的可讀性級別。

#### (四) kNN分類

在文本的詞向量空間當中，單詞之間的語意相似度(Semantic Similarity)會嵌入在詞向量資訊中。對於語意相似度高的單詞，其可讀性級別可能大部分相近。因此基於這個

---

#### Algorithm 1 kNN classification

---

**Input:**  $t_i$ : target word;  $W$ : vocabulary with readability ;  $k$ : # of nearest neighbor;

**Output:**  $l_i$ : readability level of  $t_i$ ;

1: **for**  $(t_j, l_j) \in W$  **do**

2:   calculate the similarity  $Sim(t_i, t_j)$ ;

3: **end for**

4: Sort  $W$  according to the similarity;

5: Count the number of occurrences of each class  $l_j$  among the  $k$  nearest neighbors;

6: Assign to  $t_i$  the level  $l_i$  which is the most frequent class;

---

圖4、kNN分類演算法。

構想，當Word2vec計算出單詞的詞向量後，我們利用kNN方式來進行單詞可讀性級別的分類。圖4是kNN演算法的演算法。

在本研究中，為了與kNN預測模型進行效能比較，我們基於單詞的共現關係（Co-occurrence），利用所有在  $t_i$  上下文跨距為  $k$  的視窗內的有可讀性級別周邊詞，以 Majority Voting的方式來決定 $t_i$ 的可讀性級別。例如圖5中，如果 $k=5$ ，我們要預測「夫婦は貧乏で食べる物がないので、他の家が捨てたイモの尻尾ばかりを食べています。」中「家」的可讀性級別，就以「家」為目標詞，用前後 $k$ 個有可讀性級別的周邊詞來進行投票。如果「家」在文本中總共出現 $m$ 次，這些 $2km$ 個有可讀性級別的周邊詞都一起進行投票。

單詞	夫婦	は	貧乏	で	食べる	物	が	ない	の
級別	N/A	N/A	4	N/A	1	3	N/A	1	N/A
單詞	で	他	の	家	が	捨て	た	イモ	...
級別	N/A	3	N/A	目標詞	N/A	N/A	N/A	3	...

圖5、考慮共現關係來決定目標詞的可讀性級別。

#### 四、實驗結果及分析

##### （一）實驗環境設定

在本研究中，我們所探討的日語的單詞可讀性級共分為6級：「初級前半」、「初級後半」、「中級前半」、「中級後半」、「上級前半」及「上級後半」。日文文章的可

表3、實驗的日文文章數量。

文本來源	數量
福娘童話集-日本童話	368
福娘童話集-世界童話	372
青空文庫	273
IEICE論文摘要	214
總計	1227

表4、日文文章可讀性級別的分佈。

文本來源	初級 前半	初級 後半	中級 前半	中級 後半	上級 前半	上級 後半
福娘童話集	23	504	208	5	0	0
青空文庫	1	14	128	74	47	9
IEICE論文摘要	0	0	2	25	96	91
總計	24	518	338	104	143	100

讀性級別也是分成相同的6級。為了進行效能評估，我們從福娘童話集、青空文庫以及IEICE的論文摘要分別蒐集日文文章，希望能涵蓋不同可讀性級別的文章。表3是本研究所蒐集的日文文章的數量，表4是文章可讀性級別分佈。

## (二) 實驗結果

在本研究中，我們提出來基於詞向量的kNN模型稱之為WR-kNN。在實驗中，由於Word2vec的CBOW模型與Skip-gram模型，因此實驗分別量測WR-kNN-C (使用CBOW模型)與WR-kNN-SG (使用Skip-gram模型)的效能。由於詞向量在語意空間維度  $D$  的多寡會影響詞向量蘊含語意的豐富度，也會影響整體的效能表現，因此在實驗中，我們分別討論了50、100、200、300這4種維度，來觀察不同維度的影響。所進行比較的共現周邊詞的計算方式稱為kCO。在kCO中，我們考慮周邊詞所選的上下文跨距範圍是否受到語句邊界的限制。kCO-S表示受到語句邊界限制。kCO-D則表示不受到語句邊界限制，只受到文章本身邊界的限制。

在測試的時候，我們按照文本的可讀性級別，在各個級別中隨機選取100分文本，希望能保持各個級別文本數量的平衡。但是因為「初級前半」的文章數量很少，因此我們總共取出524篇文章。我們再從這524篇文章中按照單詞的可讀性級別，在各個級別中取

表5、測試資料集中的文章與單詞數量。

	初級 前半	初級 後半	中級 前半	中級 後半	上級 前半	上級 後半
文章數量	24	100	100	100	100	100
單詞數量	72	110	252	556	494	95

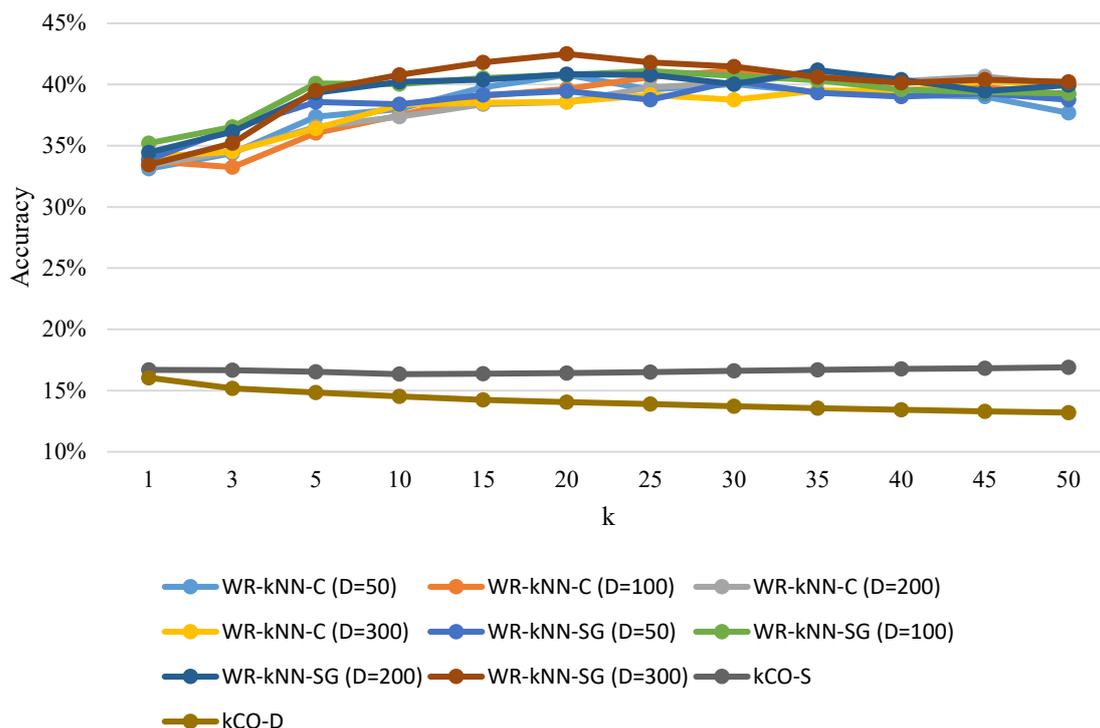


圖6、單詞可讀性準確率預測效能。

出前10%具有可讀性級別標註的高頻詞進行測試。從中總共取了1579個高頻單詞進行測試。表5是這些用來測試的文本與單詞在不同級別中的數量。

圖6是進行測試的準確率結果。從圖中可以看到WR-kNN的表現都比kCO為佳。WR-kNN的準確率表現在33.12%~42.50%，而kCO的準確率表現在13.20%~16.90%。其中WR-kNN-SG在詞向量維度 $D=300$ 、 $k=20$ 的時候有最佳準確率42.50%的表現。整體來說，使用Skip-gram模型的詞向量要比使用CBOW模型的詞向量有好的準確率表現。

我們另外計算了可讀性級別誤差一個級別也在可容許範圍內的準確率( $\pm 1$  Accuracy)。這是由於在一些實際情況中，即使預測的結果有一個級別的誤差，也還是能夠對專家的辨識有所幫助。圖7是 $\pm 1$ 準確率的效能表現。從圖中可以看到，kCO的表現隨著 $k$ 值的增加而逐步下降，但是WR-kNN卻有逐步上升的趨勢。WR-kNN的 $\pm 1$ 準確率表現在78.21%~87.78%，而kCO的 $\pm 1$ 準確率表現在30.41%~40.82%。其中WR-kNN-C在詞向量維度 $D=200$ 、 $k=45$ 的時候有最佳準確率42.50%的表現。整體來說，使用CBOW模型的詞向量要比使用Skip-gram模型的詞向量有好的 $\pm 1$ 準確率表現。

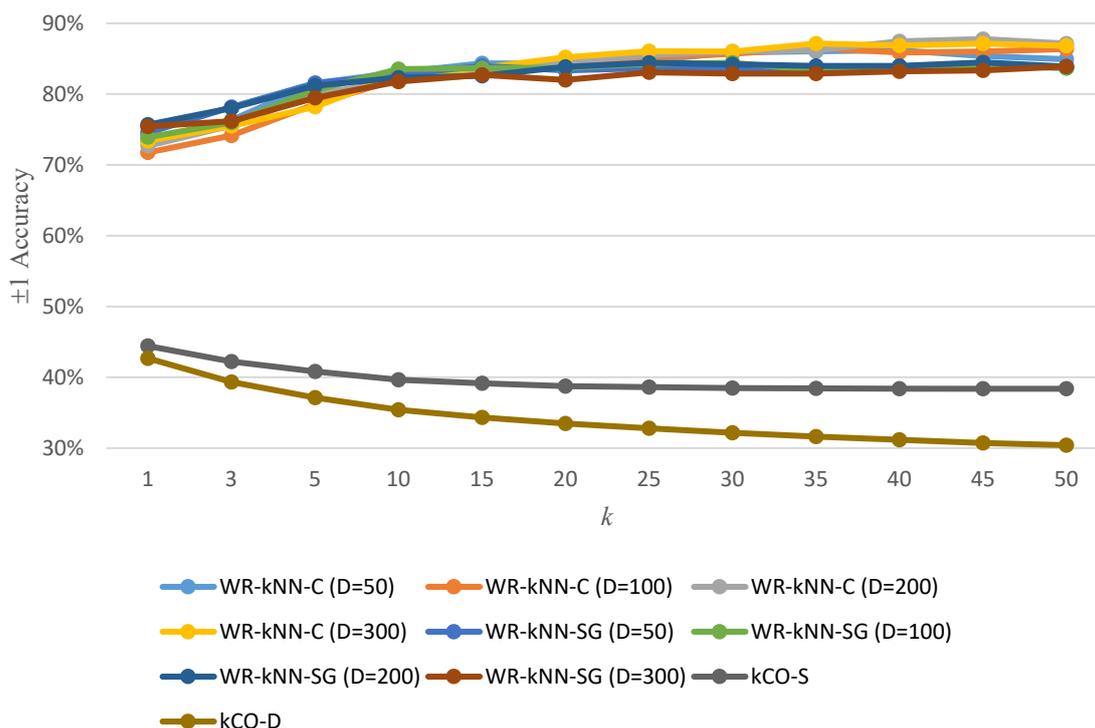


圖7、單詞可讀性±1準確率預測效能。

## 五、 結論

文本可讀性多年來都是一項重要的研究議題。然而過往研究大多偏向討論文章的可讀性，鮮少有研究討論單詞的可讀性。在本研究中，我們利用詞向量的特性，使用kNN分類器提出一個新的單詞可讀性級別評估的機器學習模型WR-kNN。經由與共現周邊詞的計算方式進行比較，實驗結果顯示，WR-kNN的單詞可讀性分類準確率最多可以達到42.50%，比基於單詞上下文周邊詞的計算方法可以得到2.51倍的改善。

在未來研究規劃中，我們預計將探討其他分類器的設計，希望能提昇評估的準確率效能。我們同時繼續蒐集更多文本來進行訓練與測試，希望能夠透過大量文本的訓練，能更準去表現出單詞之間的語意關係。

## 六、 致謝

本研究感謝科技部計畫編號 MOST 109-2221-E-155-028 部份支持。

## 七、 參考文獻

- [1] G. R. Klare, *The Measurement of Readability*: Iowa State University Press, 1963.
- [2] G. R. Klare, “The measurement of readability: useful information for communicators,” *ACM Journal of Computer Documentation*, vol. 24, no. 3, pp. 107-121, 2000.
- [3] 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.
- [4] R. F. Flesch, “A New Readability Yardstick,” *Journal of Applied Psychology*, vol. 32, no. 3, pp. 221-233, 1948.
- [5] Y. Tateisi, Y. Ono, and H. Yamada, “A Computer Readability Formula of Japanese Texts for Machine,” in Proceedings of the 12th International Conference on Computational Linguistics (COLING '88), 1988, pp. 649-654.
- [6] S. Sato, S. Matsuyoshi, and Y. Kondoh, “Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus,” in Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08), Marrakech, Morocco, 2008.
- [7] Y. Sunakawa, J. Lee, and M. Takahara, “The Construction of a Database to Support the Compilation of Japanese Learners’ Dictionaries,” *Acta Linguistica Asiatica*, vol. 2, no. 2, 10/23, 2012.
- [8] T. Mikolov, W.-T. Yih, and G. Zweig, “Linguistic Regularities in Continuous Space Word Representations,” in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2013), 2013, pp. 746-751.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2nd edition, 2009.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *CoRR*, vol. arXiv:1810.04805, 2018.
- [11] Y.-T. Sung, T.-H. Chang, W.-C. Lin, K.-S. Hsieh, and K.-E. Chang, “CRIE: An automated analyzer for Chinese texts,” *Behavior Research Methods*, vol. 48, no. 4, pp. 1238-1251, 2016/12/01, 2016.
- [12] H.-C. Tseng, H.-C. Chen, K.-E. Chang, Y.-T. Sung, and B. Chen, “An Innovative BERT-Based Readability Model,” in Proceedings of the 2nd International Conference on Innovative Technologies and Learning (ICITL 2019), pp. 301-308, 2019.
- [13] Y. Hasebe, and J.-H. Lee, “Introducing a Readability Evaluation System for Japanese Language Education,” in Proceedings of the 6th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J 2015), pp. 19-22, 2015.

- [14] 楊孝滌, “中文可讀性公式,” *新聞學研究*, 8 卷, 77-101, 1971.
- [15] 陳世敏, “中文可讀性公式試擬,” *新聞學研究*, 8 卷, 181-226, 1971.
- [16] 劉憶年、陳冠宇、曾厚強、陳柏琳, “可讀性預測於中小學國語文教科書及優良課外讀物之研究,” in *Proceedings of the 27th Conference on Computational Linguistics and Speech Processing (ROCLING 2015)*, 2015, 71-86.
- [17] 曾厚強、宋曜廷、陳柏琳, “探究不同領域文件之可讀性分析,” in *Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (ROCLING 2017)*, 2017, 116-118.
- [18] 曾厚強、陳柏琳、宋曜廷, “探索結合快速文本及卷積神經網路於可讀性模型之建立,” in *Proceedings of the 30th Conference on Computational Linguistics and Speech Processing (ROCLING 2018)*, 2018, 116-125.
- [19] 荊溪昱, “中文國文教材的適讀性研究：適讀年級值的推估,” *教育研究資訊*, 3 卷, 3 期, 113-127, 1995.