Computational Linguistics and Chinese Language Processing
Vol. 23, No. 2, December 2018, pp. 85-98

以深層類神經網路標記中文階層式多標籤語意概念

Hierarchical Multi-Label Chinese Word Semantic Labeling using Deep Neural Network

周瑋傑*、王逸如*

Wei-Chieh Chou and Yih-Ru Wang

摘要

傳統上對超過 100 個階層式標籤分類可以使用扁平 (flatten) 標籤做分類，但如此會喪失架構樹 (taxonomy) 的階層資訊。本研究旨在對廣義知網中文詞彙做概念分類與標記，提出考慮廣義知網架構樹階層關係之深層類神經網路訓練方法，其輸入為詞彙樣本點的詞向量，詞向量方面本研究亦提出考慮上下文前後關係之 2-Bag Word2Vec，而各階層的訓練結果有不同的重要性，所以在模型的最後使用最小分類誤差法，賦予各階層在測試階段時不同的權重。實驗結果顯示階層式 (hierarchical) 分類預測正確率會比扁平分類還高。

關鍵詞：詞向量、類神經網路、最小分類誤差、廣義知網、階層式分類、多標籤分類

Abstract

Traditionally, classifying over 100 hierarchical multi-labels could use flatten classification, but it will lose the taxonomy structure information. This paper aimed to classify the concept of word in E-HowNet and proposed a deep neural network training method with hierarchical relationship in E-HowNet taxonomy. The input of neural network is word embedding. About word embedding, this paper proposed order-aware 2-Bag Word2Vec. Experiment results shown hierarchical classification will achieved higher accuracy than flatten classification.

* 國立交通大學電機工程學系
Department of Electrical Engineering, National Chiao Tung University
E-mail: m0450743.eed04g@g2.nctu.edu.tw; yrwang@mail.nctu.edu.tw

Keywords: Word2Vec, Neural Network, Minimum Classification Error, E-HowNet, Hierarchical Classification, Multi-label Classification.

1. 緒論 (Introduction)

文字是資訊傳遞的重要媒介，其使用上豐富多變，中文語言中有所謂同義詞(synonym)，例如: 星期一與禮拜一，其為同義並且可相互替換字型的詞。另外也有部分詞彙為相同概念(concept)但其義不同的詞，例如: 戰鬥機與轟炸機，而這兩者都有飛行器以及戰鬥的概念，但其兩者意旨不同實體。我們希望可以在搜尋引擎中鍵入一個詞彙就可以搜尋到相似概念詞彙之搜尋結果，因此如何對中文詞彙進行語意概念標記將是本研究之重點。

在概念的範疇分析(ontology of concept)中，本研究使用廣義知網(Extended-HowNet, E-HowNet[1])的詞彙概念做為樣本點，廣義知網是 2003 年中央研究院資訊所詞庫小組將詞庫小組詞典(CKIP Chinese Lexical Knowledge Base)的詞條與董振東先生創建之知網(HowNet)的語意定義機制做連結、擴充以及修改(Huang, Chung & Chen, 2008)，以新的語義義原(sememe)通過義原的組合來標記各種單純或複雜的概念，以及各個概念與概念之間，概念的屬性與屬性之間的關係(Su, Li & Li, 2002) (Liu & Li, 2002)。

當一個樣本點對應到多個類別時可稱為多標籤(multi-label)，而一個樣本點對應到的類別數僅有一個時是為多類別(multi-class)。廣義知網中每一個詞彙樣本點皆會在架構樹(taxonomy)中對應至一個階層式標籤，其標籤為人工標記，圖 1 為某詞彙之概念分類為{mental|精神}在廣義知網上截至該節點之階層式關係圖，該詞彙之階層式多標籤資訊為物體 -> 萬物 -> 抽象物 -> 精神 -> null (null 為停止節點，原先架構並無該資訊，停止節點使用方式將在後續章節介紹)。

扁平(flatten)分類是不考慮架構樹階層式資訊，其在訓練上簡單容易，但資料本身的結構化標籤關係就未被考慮，故本研究以階層式(hierarchical)分類為基礎，提出考慮廣義知網架構樹上下位階層式標籤資訊並以類神經網路建構之階層式多標籤分類模型，網路之輸入為詞彙的詞向量，本研究另將 Mikolov 提出之 Word2Vec[2] (Mikolov, Chen, Corrado & Dean, 2013) (Mikolov, Sutskever, Chen, Corrado & Dean, 2013)模型做修改，提出考慮目標詞上下文前後關係之 2-Bag Word2Vec 架構，希望以不同的詞向量模型評測詞向量間的效能。若詞向量可以將語意及語法隱含在其中，那麼詞向量也可以運用在詞彙的概念分類上。

[1] http://ehownet.iis.sinica.edu.tw/index.php

[2] https://code.google.com/archive/p/word2vec/

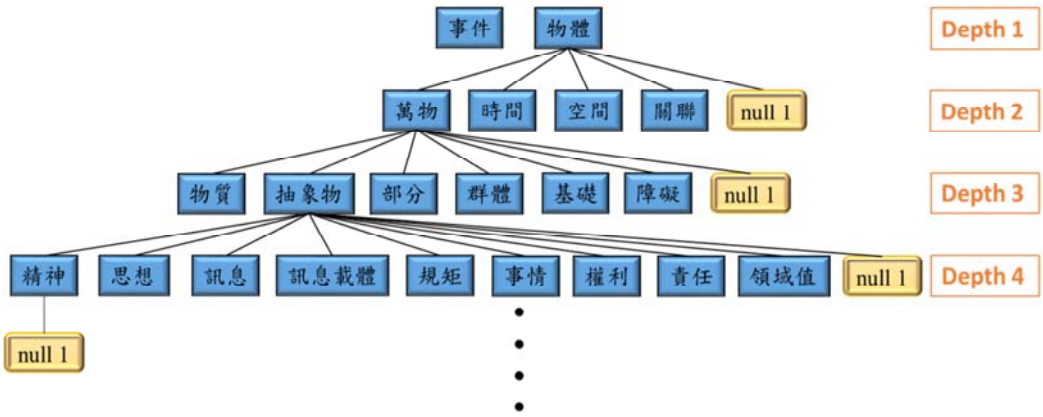

圖 1. 廣義知網樹狀階層式架構 — 截至{mental\|精神}之節點為
[Figure 1. Taxonomy of E-Hownet --- take {mental\|精神} as example]

2. 詞向量模型 (Word Vector Model)

本章節我們將會介紹本研究使用的詞向量模型，此處將介紹 Mikolov 所提出的 Word2Vec 模型，其中包含 Continuous Bag-Of-Words (CBOW)和 Skip-gram，本研究亦提出 2-Bag Word2Vec，後者與前者不同的是後者為考慮目標詞前後文關係之詞向量模型。

在 CBOW (圖 2 左側) 和 Skip-gram (圖 2 右側) 的架構中，投影層至輸出層的預測矩陣(prediction matrix) 皆相同為 $M' \in \mathbb{R}^{(|V|) \times d}$，在不同位置的詞使用共同的預測矩陣，也就是說這兩種模型在學習上並沒有加入詞在句子中的位置資訊，其會導致詞向量訓練結果不包含相鄰詞的順序關係。

本研究將對 Mikolov 所提出的 Word2Vec 模型修改，提出 2-Bag Word2Vec，該模型下包含 2-Bag CBOW 以及 2-Bag skip-gram。以圖 3 左側之 2-Bag CBOW 為例，2-Bag CBOW 中輸入為 (2 × d) 維的向量，其中投影層 (projection layer) 加大，目標詞之前與之後的投影層分開成兩個輸入相鄰詞[$e(W_{-c},\ldots, W_{-1})$, $e(W_1, , \ldots, W_c)$]，投影矩陣變為 $M' \in \mathbb{R}^{(|V|) \times 2d}$，意即加大投影矩陣分開保留字詞前後的關係。圖 3 右側之 2-Bag skip-gram 則使用兩個預測矩陣M'_a、M'_b，大小各為$M' \in \mathbb{R}^{(|V|) \times d}$，也是考慮了相鄰詞的前後順序。

僅考慮目標詞的前後關係而非完整考慮目標詞之順序的原因在於，系統參數量(weight) 增加，M′的大小變為兩倍，訓練語料要夠大才能得到好處，故沒有考慮各個輸入詞彙的順序，以觀察在系統參數量較少的情況下，詞向量訓練的成效如何。

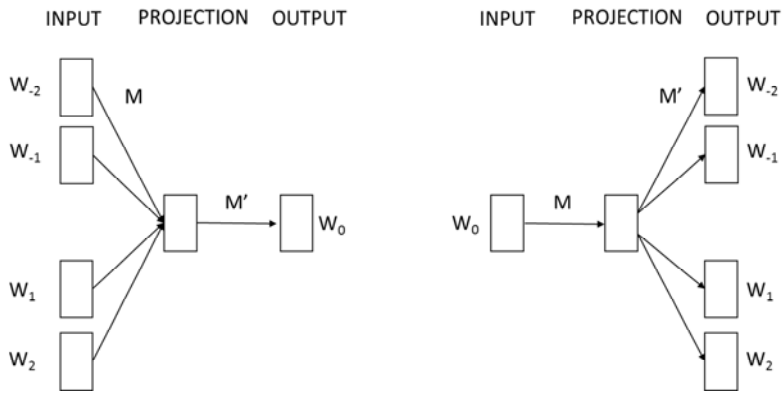

圖 2. Word2Vec 模型，左側為 CBOW，右側為 Skip-gram
[Figure 2. Word2Vec model. The left side of the figure is CBOW. The right side of the figure is Skip-gram.]

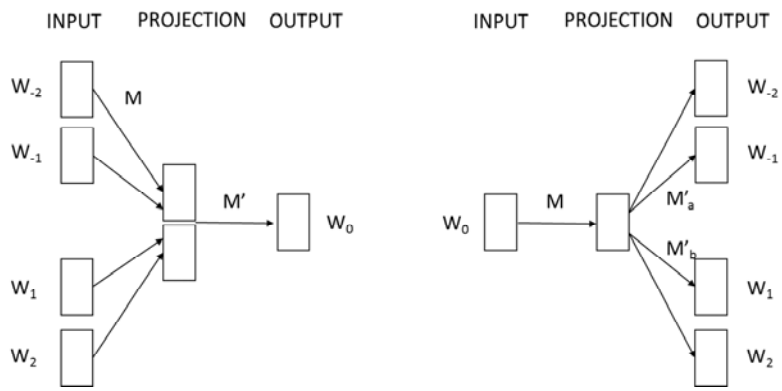

圖 3.本研究提出之 2-Bag 模型，左側為 2-Bag CBOW，右側為 2-Bag Skip-gram
[Figure 3. 2-Bag model proposed by this research. The left side of the figure is 2-Bag CBOW. The right side of the figure is 2-Bag Skip-gram

3. 以類神經網路建構之階層式多標籤分類 (Hierarchical Multi-label Classification using Neural Network

3.1 模型 (Model)

相比扁平分類，階層式分類為考慮資料標籤之階層式架構。階層式多標籤分類模型以架構樹之階層式標籤使用多個類神經網路建立，使用類神經網路可讓模型建構上變得較有彈性，圖 4 為本研究所提出之階層式模型架構，在不同深度 (depth) 之各階層分類上皆使用一個類神經網路，而各階層皆有一個輸出，各輸出分別對應至各階層的標籤，其為一個多輸出 (Multi-Output) 模型。階層式標籤可能有長有短，在各階層之類神經網路輸出層中加入一個停止節點 (圖 1 的 null 標籤) ，所以一個長度較短的標籤從模型由上而

下 (top-down) 遇到停止節點後，其後在輸出層皆為停止標籤 (null) 直到模型的最底。

訓練階段一開始，一個神經網路負責架構樹資料的第一層資訊 (depth 1，較靠近 root 節點的階層)，此網路有一個隱藏層以及一個輸出層 (Depth 1 的類別)，網路權重更新的方式為倒傳遞演算法 (Back-propagation)，每一個神經網路僅負責預測一個架構樹階層中的類別。當第一階層的神經網路訓練完畢後 (圖 4 上方對應到架構樹第一階層)，第二階層會另外有一個類神經網路負責訓練，差別在於第一階層神經網路的輸出與整個網路一開始的輸入特徵相接 (concatenate) 後做為第二階層神經網路的輸入 (圖 4 下方第二階層之輸入)，此訓練程序會不斷重複直到最後一階層之神經網路被訓練，如此將各階層神經網路相接形成一個深層神經網路。

測試階段將測試資料餵入第一個神經網路 (第一階層之神經網路)，而後第一階層的輸出做為第二階層的輸入，此過程將不斷重複直到抵達最後一階層，在各階層神經網路的輸出後使用 softmax，最後執行下一節要介紹的修正矛盾現象。

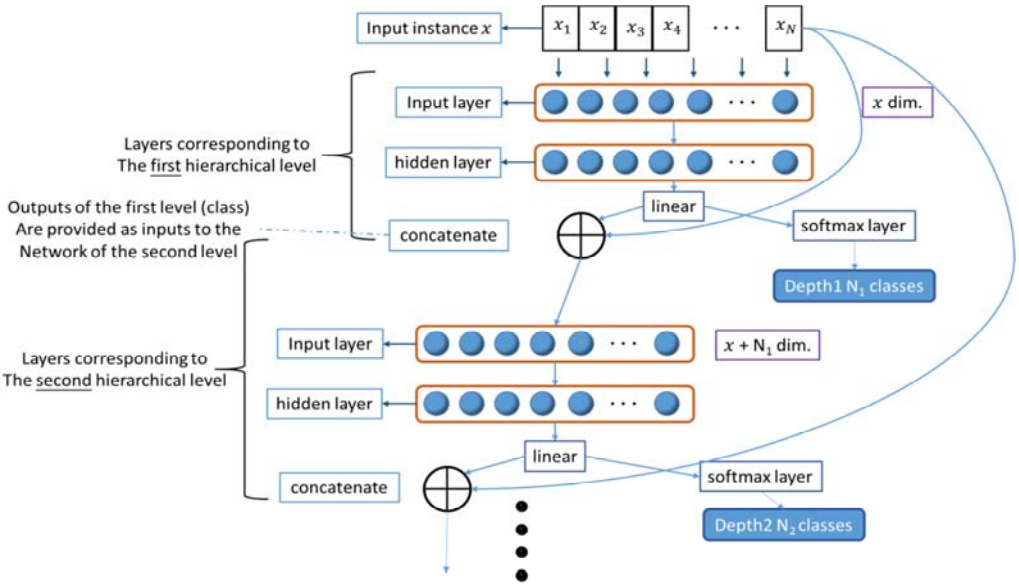

圖 4.以深層類神經網路架構之階層式多標籤分類模型
[Figure 4. The structure of Hierarchical Multi-Label Classification using neural network]

3.2 階層式標籤矛盾現象 (Correcting Inconsistencies)

當階層式多標籤類神經網路完成其訓練之後，在測試階段，還需做一個後處理來校正上下位標籤關係矛盾情況，例如: 一個下位階層的節點 (subclass) 被預測，但其所屬上位節點 (superclass) 卻沒有。這些矛盾情況的發生原因在於每個階層的神經網路都是各別輸出結果的，也就是說在測試階段，一個詞彙的預測結果是與其他階層的結果沒有相依性的，所以會造成上下位階層前後矛盾 (inconsistent)，意即出現架構樹上不存在的階層式標籤資訊，使用後處理可保證在最後階段不會存在矛盾情況。

$$p(X) = \sum_{L=1}^{N} \log\left(\sigma(X_{Lk})\right) \quad (1)$$

針對各階層內的多類別分類上本研究在輸出層後使用 Softmax，式(1)為考慮架構樹路徑資訊後對單筆路徑計算 softmax 機率和的方法，其中 L 為該階層架構下的某一層，k 為該階層中某一分類的機率，在此依照資料集之架構樹的資訊對架構樹的多標籤路徑資訊做加總機率和後的結果稱為分數，最後總共有 N 個分 數 (架構樹共有 N 條路徑)，再依照分數的大小作排序，求其正確率。

3.3 效能評估方式 (Measurement)

實驗所使用的評估方式為正確率 (accuracy)，其計算為標記正確的詞彙數與總詞數的比值，而在考慮架構樹後的多標籤深層類神經網路中我們將探討考慮架構樹階層時的正確率，故我們將正確率再細分為總體正確率與各層正確率，分別如下:

總體正確率: 該層類別是否正確需考慮上層所有測試路徑是否正確

各層正確率: 僅考慮該層類別是否正確

表 1. word α 正確率計算示意
[Table 1. word α test path]

	type	Depth 1	Depth 2	Depth 3	Depth 4	Depth 5
word α	TEST	classA	classF	classC	null	null
	TRUTH	classA	classB	classC	null	null

以表 1 為例，如 word α 的 TRUTH 所示，word α 落在架構樹的節點 C，在測試階段 word α 在架構樹的預測結果也落在節點 C，但其預測之多標籤路徑 A -> F -> C (圖 5 橘色路徑，正確路徑為綠色) 為一個不存在的路徑，計算正確率時以 Depth 3 為例，因為 word α 在 Depth 2 就錯誤，所以在計算總體正確率時僅有 Depth 1 為正確，而 Depth 2 和 Depth 3 皆為錯誤 ; 而在各層正確率方面 Depth 1、3 都計為正確，而 Depth 2 計為錯誤。

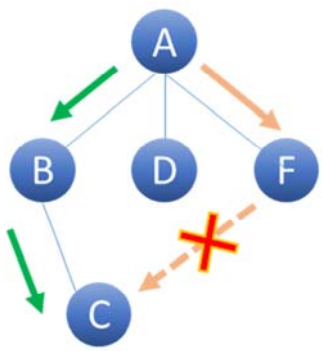

圖 5. word α之預測與實際路徑
[Figure 5. word α test and truth path]

4. 實驗與分析 (Experiment and Analysis)

4.1 實驗資料與實驗設定 (Experiment Setup)

在建立詞向量方面使用的語料包含 (1) LDC Chinese Gigaword Second Edition[3]、(2) Sinica Balanced Corpus ver. 4.0、(3) CIRB0303[4] (Chinese Information Retrieval Benchmark, version 3.03)、(4) Taiwan Panorama Magazine[5]、(5) TCC300[6] 和 (6) Wikipedia (ZH_TW version)，以上各語料庫經交通大學王逸如老師開發之斷詞器(Wang, Wu, Liao & Chang, 2013)斷詞後共約 4.4 億詞。為了使詞向量更精確，在建立詞向量之前會先經過文字正規化、同義詞替換、少數人名和數字合併。訓練詞向量時，目標詞前後窗口為 7 (c = 7)，維度為 200 維，詞頻篩選剔除小於 25 的詞彙，最後共建立出 19 萬個詞向量。

廣義知網詞彙概念樣本點方面，廣義知網未對所有的詞彙標上概念展開式，例如「新歌」在詞向量有被建立，但是廣義知網卻未標上其概念展開式，扣除未標上概念展開式以及詞向量未被建立的詞彙，最後約使用 5 萬筆資料當作概念樣本點，其中 90%為訓練資料集，其餘 10%為測試資料集。

某些概念的詞彙樣本點過少，故本研究也將廣義知網的少數概念類別向上一層合併，合併完成後為 403 個概念類別。

在未進行概念類別合併之前，從廣義知網的詞彙統計結果發現，{人}的類別所佔的數目最多佔了 3404 個，而第二名的{事物}為 1511 個 ；如果看到最少的資料，其中數量小於 10 的類別有 521 個，小於 5 的有 321 個，小於 1 的有 107 個。因此針對不平衡資料做處理，此處使用隨機資料增生，將所有個別數量小於 30 的類別增生至 30 類，如此原先資料從 50924 筆增加至 52906 (+1982) (+3.9%)

4.2 扁平分類 (Flatten Classification)

扁平分類為不考慮廣義知網架構樹上下位階層關係之分類方法，此處將使用最近鄰居法 (K-Nearest Neighbors Algorithm, KNN) 與類神經網路(Neural Network)。

在進行最近鄰居法時，必須先決定 K 值，然而最佳 K 值取決於樣本點等等因素，隨著樣本點的改變其值亦會改變，雖然我們可以透過一些最佳化演算法去求得最佳的 K 值，但我們未採用任何演算法求最佳 K 值，而是希望透過實驗結果決定。如果在投票時平票 (tied-vote)，則考慮相似度最大者為第一名 (Top 1)，如不依照相似度排序則為第一名最大值 (Top 1 MAX)。此處最近鄰居法使用餘弦距離 (cosine distance)。

---

[3] https://catalog.ldc.upenn.edu/LDC2005T14

[4] http://www.aclclp.org.tw/use_cir.php

[5] https://www.taiwan-panorama.com/en

[6] http://www.aclclp.org.tw/use_mat.php#tcc300edu

表 2. 最近鄰居法標記正確率
[Table 2. Accuracy of KNN]

	CBOW200	Skip-gram200	2Bag-CBOW200	2Bag-Skip-gram200
Top 1	**55.8%**	53%	55.2%	52.7%
Top 1 Max	63.9%	62%	62.6%	61.1%
Top 3	74.4%	73.4%	74.3%	72.5%
Top 3 Max	80.2%	79.4%	80.1%	78.7%
K value	8	10	8	9

由表 2 發現 Mikolov 所提出的 Word2Vec 模型之效能比本研究所提出之 2-Bag 模型要來得好，而採用以鄰近詞預測目標詞的 CBOW 和 2-Bag CBOW 正確率皆較高。

在最近鄰居法方面的實驗結果驗證在詞向量模型方面 Word2Vec 的 CBOW 和 Skip-gram 有較好的效能，故在類神經網路方面選用 Word2Vec 的詞向量模型架構做後續詞彙標記模型的輸入。

表 3. 類神經網路法標記正確率
[Table 3. Accuracy of Neural Network]

word vector	Top1 Acc.	Top1 Acc. + POS	Top3 Acc. + POS
CBOW 200	54.1%	58.4%	75.3%
Skip-gram 200	53.7%	56.7%	74.6%
CBOW 200 + Skip-gram 200	**55.4%**	**59.5%**	**76.4%**

此處探討加入 POS (part of speech) 後對正確率的影響，POS 經 one-hot encoding 為 11 維向量，該 POS 向量與 CBOW 或 Skip-gram 向量相接後各為 211 維向量，而 CBOW 200 + Skip-gram 200 + POS 為 411 維向量，由表 3 得知，加入 POS 資訊後在各詞向量模型上可以提高正確率，而其中詞向量又以 CBOW 200 + Skip-gram 200 + POS 的正確率最佳，儘管 CBOW 200 + Skip-gram 200 中兩個詞向量相接上並未做任何優化處理，但也得到了較好的效果，使用 CBOW 200 + Skip-gram 200 + POS 的特徵達到了 59.5%的正確率。

4.3 階層式多標籤分類 (Hierarchical Multi-Label Classification)

從扁平式分類章節發現 CBOW + Skip-gram + POS 之 411 維特徵之正確率表現較好，故階層式多標籤分類模型的輸入也使用相同設定，其中 CBOW 和 Skip-gram 各 200 維，另外加入 POS one-hot 之 11 維資訊。

在訓練階段，將各詞彙抽取 head concept 後，依照架構樹的資訊從樹葉節點 (leaf node) 往 root 節點尋找每個詞彙的路徑資料，各標籤的路徑長短不一，資料未達到 N 層的情況下將在下位階層的節點補上 null。

表 4. 階層式多標籤分類模型之測試階段正確率 (更正標籤矛盾情況)
[Table 4. Accuracy of Hierarchical Multi-Label Classification (Correcting inconsistencies)

	Overall Accuracy		Layer Accuracy		
Depth N	**Acc. Top1**	**Acc. Top3**	**Acc. Top1**	**Acc. Top3**	**類別數**
Depth 1	98.3%	99.1%	98.3%	99.1%	3
Depth 2	90.7%	95.0%	90.7%	95.0%	7
Depth 3	81.9%	89.8%	81.9%	89.8%	18
Depth 4	77.1%	87.0%	77.3%	87.2%	83
Depth 5	70.9%	83.1%	75.1%	86.3%	85
Depth 6	65.8%	79.4%	72.4%	85.2%	127
Depth 7	63.4%	77.7%	79.8%	88.9%	78
Depth 8	62.2%	76.9%	87.1%	92.9%	23
Depth 9	61.7%	76.5%	91.8%	95.6%	31
Depth 10	61.3%	76.2%	95.6%	97.6%	38
Depth 11	**61.0%**	76.1%	98.2%	99.1%	10

在測試階段採用 3.2 節更正標籤矛盾情況方法，測試階段正確率顯示於表 4 中，觀察正確率發現在 Depth 1 到 Depth 2 時僅僅是增加 4 個類別 (3 類->7 類)，總體正確率就下降了將近 8%。

4.3.1 賦予各階層不同權重 (Given Different Layer Different Weight)

在階層式多標籤類神經網路測試結果中發現某些階層的正確率偏低，由表 5 中觀察到 Depth 4 到 Depth 7 之正確率都不足 80%，低於其他階層的正確率。可能原因為該層類別數較多，導致該層分類困難。此處可以給予每一層不同的 weight 來較相信或較不相信才自某幾層的資訊，幫助辨認結果。

在此可以將各階層的權重視為一個參數集來最大化總體正確率，在此應用最小分類誤差法，以分類的方式決定每一個階層的權重。最小分類誤差之決策式如下:

$$d_k(X) = -g_i(X;w) + \log\left[\frac{1}{M-1}\sum_{j,j\neq c} exp\big(g_j(X;w)\eta\big)\right]^{1/\eta} \quad (2)$$

其中η為一個正整數，$d_k(X) > 0$時為分類錯誤，而$d_k(X) \leq 0$為分類正確， logarithm 與 exponential 互為反函數，其目的為避免方程式$g_j(X;w)$進行η次方時產生計算機 underflow 問題。

此處可以看成是正確類別和所有錯誤類別的競爭學習過程，當 η 趨近∞ 時，(2)中

括弧內的方程式會變為$\max_{j,j\neq i} g_j(X;w)$，意即僅找所有錯誤類別中錯誤分數最大的結果來訓練以加快訓練速度，如此(2)會變為:

$$d_k(X) = -g_i(X;w) + g_j(X;w) \quad (3)$$

(2)中$g_j(X;w)$為類別條件似然函數(class conditional likelihood functions)，可將 softmax 後結果 X 視為已知，而欲找到一組最佳權重 w 來最大化似然函數，(2)帶入各階層 softmax 後結果以及 weight，可將方程式寫為:

$$d_k(X) = -\sum_{L=1}^{11} X_{LC} W_L + \log\left[\frac{1}{M-1}\sum_{j,j\neq c} exp\left(\left(\sum_{L=1}^{11} X_{Lj} W_L\right)\eta\right)\right]^{1/\eta} \quad (4)$$

其中C ∈廣義知網架構樹的正確路徑，該方法將所有錯誤以及正確路徑做競爭學習。

而在尋找最佳的參數叢集 X 時，也可將個別的錯誤資料是其重要程度，改變 η 和 M 將所有錯誤競爭類別加入考慮，而此處把(3)嵌入 zero-one function，定義損失函數 (loss function)，此處以 sigmoid function (5)當作考量。

$$\ell(d) = \frac{1}{1+exp(-\gamma d+\theta)} \quad (5)$$

損失函數的 θ 通常是 0，$\gamma \geq 0$，改變 θ 和 γ 可改變 loss 被調整的範圍。

表 5. 各別階層正確率中，部分階層 (框選處) 正確率較低
[Table 5. Accuracy of each layer. Accuracy of some layer (circled) are lower than the others.]

	Accuracy of each layer	
Depth N	**Accuracy**	**Classes per layer**
Depth 1	98.3%	3
Depth 2	90.7%	7
Depth 3	81.9%	18
Depth 4	77.3%	83
Depth 5	75.1%	85
Depth 6	72.4%	127
Depth 7	79.8%	78
Depth 8	87.1%	23
Depth 9	91.8%	31
Depth 10	95.6%	38
Depth 11	98.2%	10

在 Minimum Classification Error 訓練階段以梯度下降法做參數的更新:

$$W_L(t+1) = W_L(t) - \varepsilon \frac{\partial \ell(X;W)}{\partial W_L} \tag{6}$$

其中 ε 為學習率，而偏微分項次可以透過鏈鎖律(chain rule)求得

$$\frac{\partial \ell(X;W)}{\partial W_L} = \frac{\partial \ell(X;W)}{\partial d}\frac{\partial d}{\partial W_L} \tag{7}$$

其中 $\frac{\partial \ell(X;W)}{\partial \mathrm{d}}$ 為對 sigmoid 微分的結果，可寫為:

$$\frac{\partial \ell(X;W)}{\partial \mathrm{d}} = \gamma \ell(d)\big(1-\ell(d)\big) \tag{8}$$

其中$\frac{\partial d}{\partial W_L}$

$$= -X_{LC} + \frac{1}{\eta}\Big[\frac{1}{M-1}\sum_{j,j\neq c} exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]\Big]^{-1}\frac{\partial d}{\partial W_L}\Big[\frac{1}{M-1}\sum_{j,j\neq c} exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]\Big]$$

$$= -X_{LC} + \frac{1}{\eta}\Big[\frac{1}{M-1}\sum_{j,j\neq c} exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]\Big]^{-1}\Big[\frac{1}{M-1}\sum_{j,j\neq c} \eta X_{Lj}\, exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]\Big]$$

$$= -X_{LC} + \frac{\sum_{j,j\neq c}\big[exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]X_j\big]}{\sum_{j,j\neq c} exp\big[\big(\sum_{L=1}^{11} X_{Lj}W_L\big)\eta\big]}$$

訓練完成的各階層權重如圖 6 所示，觀察圖表發現最小分類誤差法訓練出的權重在第八層時為最小，而在階層中正確率最低者為第六層，與預期上權重最低應該在第六層有所不符，但權重分布與階層正確率一樣為山谷型走勢。

原先未使用最小分類誤差法時各階層的權重皆為1。在最小分類誤差訓練法完成後，在測試階段計算各路徑分數時可以考慮各階層的權重的不同再進行 3.2 更正矛盾情況時計入不同的 weight，此處將(1)稍作修改如下所示:

$$\mathrm{p}(X) = \sum_{L=1}^{N} W_L \log\big(\sigma(X_{Lk})\big) \tag{9}$$

W_L為第 L 層之權重，此處在計算各階層分數並加總時考慮權重W_L。

計入權重後的測試階段預測結果如表 6 所示，在加入 weight 後其總體正確率之第 11 層較未修正 weight 的情況多了 0.3%正確率，為 61.3%，而在 Top 3 情況其總體正確率上升 0.8%。

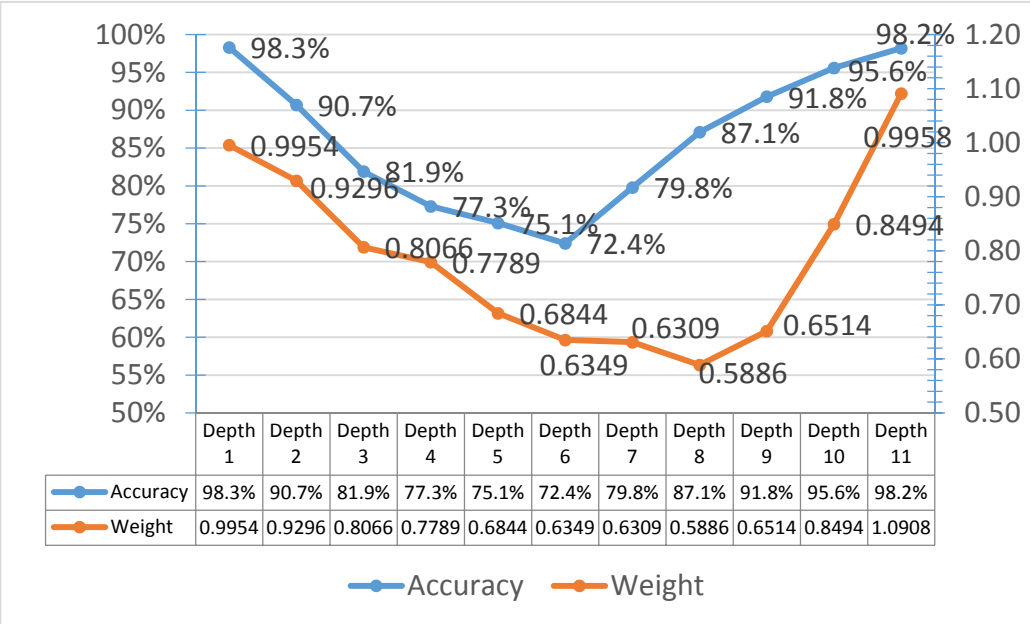

圖 6. 各階層正確率和 weight 走勢
[Figure 6. Weight and accuracy of each layer]

表 6. 調適各階層權重後的正確率
[Table 6. Accuracy of each layer after adjusting the weight]

	Accuracy Top1		Accuracy Top3		
Depth N	**original**	**add weight**	**original**	**add weight**	**classes**
Depth 1	98.3%	98.4%	99.1%	99.2%	3
Depth 2	90.7%	90.8%	95.0%	95.6%	7
Depth 3	81.9%	81.9%	89.8%	90.5%	18
Depth 4	77.1%	77.1%	87.0%	87.6%	83
Depth 5	70.9%	70.9%	83.1%	84.1%	85
Depth 6	65.8%	65.8%	79.4%	80.3%	127
Depth 7	63.4%	63.5%	77.7%	78.4%	78
Depth 8	62.2%	62.3%	76.9%	77.7%	23
Depth 9	61.7%	61.8%	76.5%	77.2%	31
Depth 10	61.3%	61.5%	76.2%	76.9%	38
Depth 11	61.0%	61.3%	76.1%	76.9%	10

4.4 標記結果討論 (Discussion)

由於詞向量是根據前後文相鄰的關係所訓練的，所以兩個同義的詞彙在用法上類似其餘弦相似度就會高，但也可能會有兩詞彙互為反義但其兩者因前後文類似所以其餘弦相似度高造成標記錯誤的情況發生，例如「瘦」與「胖」。

自動標記之第一名類別和廣義知網所標記的類別有所不同，例如: 戰車_N 在廣義知網中的類別為{車}，而自動標記第一名為{武器}，但在常識上似乎也不能直接算是錯誤類別，此處參考教育部線上辭典或維基百科對於戰車的解釋: 戰車 1. 作戰用的車輛 2. 全裝甲結構之全履帶車輛，有砲塔、自動武器、通信等裝置。

但有時廣義知網內的類別也並非完全客觀，看到下列例子: 魚丸_N 在廣義知網中的類別為{身體部件}，而自動標記第一名為{食品}，而要判斷標記之類別是否正確，不妨先了解其真實的語意再做判斷，以下列出依照教育部線上辭典或維基百科對於此詞彙的解釋：魚丸，魚肉調蛋清製成的丸子。此例子中自動標記結果比起廣義知網更貼近真實詞義，故自動標記之類別有時反而比廣義知網所標記的類別更加接近真正的語意，但這些標記錯誤的詞彙最後仍須人工檢查，且其為錯誤或正確之標記常常是見仁見智。

5. 結論與未來展望 (Conclusion and Future Work)

本研究提出考慮架構樹 (taxonomy) 之階層式多標籤資訊後以類神經網路建立的階層式多標籤分類模型，其應用於廣義知網的詞彙語意概念標記，神經網路的輸入為詞向量，在詞向量方面本研究亦提出 2-Bag model，其為將詞向量之投影層至輸出層的 weight 數量增加且考慮目標詞前後關係之模型，唯因系統參數量增加的情況下，訓練語料 4.4 億詞過少(Wang, Dyer, Black & Trancoso, 2015)，因而無法有效地訓練 2-Bag model。

實驗階段比較了階層式 (hierarchical) 與扁平式 (flatten) 分類，其兩者同樣以類神經網路建立分類模型，不同的是階層式架構之類神經網路是深層且階層的，扁平式分類的輸出層節點數與資料集中的類別數量相同。從實驗結果來看，階層式分類之正確率會比扁平分類還高。而在測試階段也採用最小誤差分類法 (minimum classification error)，讓機器自行學習各階層的重要性，賦予不同層權重，改善最後測試階段之正確率。

本研究輸入詞彙有消歧義情況，未來可以在本研究中加入中文消歧模型，增加模型辨認率。另外階層式多標籤分類法也可應用在其他同樣具有階層式多標籤資訊的資料庫，例如檔案目錄系統、生物資訊系統。

致謝 (Acknowledgements)

This work was supported by the Ministry of Science and Technology, Taiwan with contract MOST-105-2221-E-009-142-MY2.

參考文獻 (References)

Huang, S.-L., Chung, Y.-S., & Chen, K.-J. (2008). E-HowNet: the expansion of HowNet. In *Proceedings of the First National HowNet Workshop*, 10-22.

Liu, Q. & Li, S.-j. (2002). Word Similarity Computing Based on How-net. *International Journal of Computational Linguistics and Chinese Language Processing*, *7*(2), 59-76. [In Chinese]

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*, Retrived from arXiv:1301.3781v1

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of NIPS 2013*, 3111-3119.

Su, W.-F., Li, S.-Z., & Li, T.-Q. (2002). A Module of Automatic Chinese Documents Classification Based on Concept. *Computer Engineering and Applications*, *2002*(6).

Wang, L., Dyer, C., Black, A. W., & Trancoso, I. (2015). Two/Too Simple Adaptations of word2vec for Syntax Problems. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1299-1304. doi: 10.3115/v1/N15-1142

Wang, Y.-R., Wu, Y.-K., Liao, Y.-F., & Chang, L.-C. (2013). Conditional random field-based parser and language model for traditional Chinese spelling checker. In *Proceedings of the 7th SIGHAN Workshop on Chinese Language Processing (SIGHAN-7)*, 69-73.