

基於字元階層之語音合成用文脈訊息擷取

Character-Level Linguistic Features Extraction for Text-to-Speech System

陳冠宏*、廖書漢*、廖元甫*、王逸如⁺

Kuan-Hung Chen, Shu-Han Liao, Yuan-Fu Liao and Yih-Ru Wang

摘要

優良的語言文脈訊息是語音合成的關鍵部分，傳統的文脈訊息都是依賴於自然語言處理(Natural Language Processing, NLP)，即使用 parser 分析文字。但是 parser 設計困難無法專門為語音合成設計；所以我們想直接以字元為處理單元建立一個 end-to-end 的語音合成系統，在這想法下我們改用字元層級(character-level)的 word2vec 與遞迴類神經網路，直接將輸入字元序列轉換成隱藏特徵向量當做語言合成的文脈訊息。最後我們利用一中英夾雜語音合成系統測試此想法，語音合成的實驗的結果表明，我們提出的方式的確比傳統使用 parser 的方式有更好的性能。

Abstract

High quality linguistic features is the key to the success of speech synthesis. Traditional linguistic feature extraction methods are usually relied on a word-level natural language processing (NLP) parser. Since, a good parser requires a lot of feature engineering to build, it is usually a genral-purpose one and often not specially designed for speech synthesis. To avoid these difficulties, we propose to replace the conventional NLP parser by a character embedding and a chacter-level

* 國立台北科技大學電子工程系

Department of Electronic Engineering, National Taipei University of Technology
E-mail: { s970428, sam8105111 } @gmail.com; yfliao@mail.ntut.edu.tw

⁺ 國立交通大學電機工程系

College of Electricl and Computer Engineering, National Chiao-Tung University
E-mail: yrwang@mail.nctu.edu.tw

recurrent neural network language model (RNNLM) module to directly convert input character sequences, character-by-character, into latent linguistic feature vectors. Experimental results on Chinese-English speech synthesis system showed that the proposed approach achieved comparable performance with transitional NLP parser-based methods.

關鍵詞：語音合成、文脈訊息、文字向量、遞迴類神經網路語言模型

Keywords: Speech Synthesis、Linguistic Features、Word2vec、RNNLM

1. 簡介

在語音合成系統中分為兩大模組(如圖 1)，分別是文本分析(自然語言處理)與聲音合成(語音訊號處理)。其中前端的文本分析通常會做文字正規化、斷詞(word segmentation)、part of speech(POS)標註與相關文法分析，甚至是藉由韻律預測從文本提取文脈訊息特徵。例如在中文語音合成中經常採用條件隨機場(Conditional Random Fields, CRF)做斷詞和 POS 標註。另一方面在後端聲音合成一般會透過決策樹依據文脈訊息選擇最適合的聲音或韻律特徵，將選出來的聲學參數給語音編碼器合成語音波形產出聲音。因此若是我們想在語音合成系統中獲得自然、流暢的合成聲音，需要能萃取有效且有用的文脈訊息。

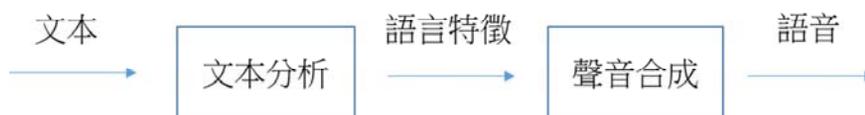


圖 1. 傳統 TTS 系統架構圖
[Figure 1. The traditional two-stage TTS approach]

在傳統方式上，文本分析主要使用 NLP 的 parser(The Stanford Natural Language Processing Group, 2015)。使用 parser 的好處是經過語言學專家所設計的斷詞能夠比較完整的表現出每段語句內的詞性狀態。但是要建立傳統斷詞很難做，需要大量專業人士標註的資料庫以及專家設計好的斷詞特徵參數，所以我們通常只能使用現成的，無法專門根據合成的需求來設計。再加上標註的過程中不同的標註人員可能會對同一句話產生不一致的標註方式，這導致機器不好學習。所以為了解決這些問題，我們想要讓機器能夠自動去學習每一個字元的屬性與字元串間的關係。就像我們人在閱讀文章時不會先對文章做斷詞及 POS 標註，我們也是一個字一個字讀過去，然後透過前後文的拆解去理解字義、詞性。所以我們能不能想一個方法跟人在閱讀的時候一樣，利用以字為單元這種比較簡單的方式，避免使用 parser 做斷詞和 POS 標註的問題，以改善原有系統求取語言特徵的缺點。

近年來，類神經網路(DNNs) (Licstar, 2013)(在 NLP 方面有越來越多的研究，提出很多字元層級的語言模型，因為使用 DNNs 可以建立更好的模型，並且學習大量無須標註的文本資料，所以基於使用類神經網路的方法能夠有效的減少 POS 標註這類的語言特徵參數設計工程(feature engineering)；例如(Greff, Srivastava, Koutník, Steunebrink, &

Schmidhuber, 2015)使用以字元為層級的語言模型,在相同的性能下,會比 n-gram 模型小。而且對於中文的語音合成來說,以字元層級為處理單元的方式更具有意義,因為中文是沒有空格的連續字元串,也沒有詞的分隔符號,所以中文的詞在定義上是很模糊的,在這情況下許多以字元層級的處理方法方法被提出,像是(Zheng, Chen & Xu, 2013)中文斷詞與 POS 標註的深層學習網路以及(Ding, Xie, Yan, & Zhang, 2015)採用雙向長短期記憶遞迴類神經網路(BLSTM)直接從中文預測韻律邊界的標註,這些研究都證明了使用 DNNs 是能夠實現比傳統 CRF 有類似甚至更優良的效能。

在這麼多使用 DNNs 獲得不錯成果的研究下,我們想說看可不可以使用大量沒有經過標註的語料來訓練 DNN,這樣不僅可以免去標註工作也無需專家設計的斷詞或 POS 特徵參數,而要達成使用無標註語料的目的,我們需要靠字元層級(character-level)的文字處理的幫助,因此我們捨棄以詞為單元,將輸入簡化成以字元為單位,不經過 parser 做斷詞、求詞性等前處理,而是一個字元一個字元逐次輸入網路中;希望能將字元轉到更高維的向量空間,在向量空間進行分析,能有效的從語料中,自動學習字元之間隱藏的相對關係。所以我們將透過建立語料的字向量空間並進行分析,將文本字元分類,產生字元語意、文法腳色資訊等文脈訊息;另一方面則使用遞迴神經網路進行字元串訓練,分析目前輸入字元在整句話中的狀態,並猜測下一狀態可能為何,最後並擷取隱藏層在各個神經元的狀態輸出做為字元的時間順序資訊當作字元時序狀態的文脈訊息。如此一來我們就能利用大量未標註的文字語料,自動擷取文脈訊息,並探討更多種文脈訊息的可能性。

2. 傳統文脈訊息擷取方法

一般文本分析的文脈訊息是輸入文本經由 parser 做斷詞、抓 POS、位置再加上切割資訊、聲調資訊等合起來的,所以要得到好的文脈訊息這些參數需要精確和有用,而 parser 在文本分析中扮演舉足輕重的角色。

在語言學上,詞是能夠獨立運用而且含有語義內容的最小語言單位。在英文文本中,每個單字(word)即是一個詞,具有完整意義,而且每個單字間都以空白做區隔,但是在中文文本裡,詞與詞之間是不會有空白做為區隔的。因此在中文的 NLP 中,為了讓電腦能夠分辨文本中的詞義,就必須先正確的將詞區隔開來,才能進一步發展出相關演算法。例如機器翻譯、資訊檢索與擷取、語言分析、語音辨識和合成,為此發展出斷詞器來使用,而斷詞器主流使用 CRF 處理輸入文本,透過訓練 CRF 對輸入句子做猜測判斷兩個字中間是否為斷點,訓練方法要依照專家設計斷詞參數去學習怎麼預測詞斷點。但詞斷點要參照已經標註好的資料庫來學習,所以對於未知詞的錯誤率無法有效降低。如何改善中文斷詞的條件機率模型相關資料可以參照(黃昭銘, 2010),當中有提到使用一個線性的 CRF 來達成更準確的中文斷詞,如果分割出的詞與相對應標準語料庫的詞不同時,透過該詞和前後斷詞的重組,可求出更適當的斷詞。

詞性標註(Part Of Speech Tagging)在 NLP 中也是一大課題。一般 parser 就包含了斷詞與 POS Tagging 兩部分，順序上是先斷詞再做詞性標註，而詞性標註就是透過適當的方式對經過斷詞處理後的每個詞給予一個合適的詞性，也就是要確定這個詞是名詞、動詞或是副詞等等，關於詞性標註可以參閱(Brill, 1992)基於 POS 標註的簡單規則。但是詞性標註的困難點在於詞性不定的問題上，這種現象是自然語言中有很多詞語本身包含很多詞性，擺在不同地方就會變換詞性。對於人在閱讀而言，這種詞性歧義現象比較容易排除，但是對於機器而言則不容易區分。傳統使用的詞性標註規則是由語言學家根據語言規律進行人工標註完成的，有了標註完的資料庫，在使用上利用 CRF 從已知的序列求對應序列的方式，來猜測輸入語料的詞性。詳細內容可以參閱相關研究，例如(唐大任，2002)中文 parser 之研究，內文探討了設計 parser 時利用的斷詞與構詞規則，和標記詞類方式的一些問題。

在 NLP 中使用 parser 是非常普遍的，對英文而言使用 parser 能得到不錯的結果，但是在中文使用起來卻不盡理想，因為中文結構在歧義性上有太多變化，需要搭配上上下文來觀察，無法單純靠人工標註及專家設計所有可能，所以我們需要一套新方法來克服這些問題。

3. 字元階層文脈訊息擷取方法與語音合成系統

在前一章中我們已經知道傳統架構擷取語言特徵會遇到的問題點，所以我們希望能使用大量無標註資料庫與非監督式學習(unsupervised learning)來訓練深層類神經網路，達到無需人工介入讓機器自動學習文本取代 parser。在這基礎下我們使用以字元為層級的訓練方式來達到我們零標註的目的。在字元屬性求取上我們採用 word2vec 來對文本進行字元的語意與文法角色分類，字元之間的時序順序關係我們利用遞迴式神經網路，擷取字元在句子中的狀態來當作時序關係，最後得到字元屬性與字元時序狀態的文脈訊息。

3.1 新系統架構

為了達成以字元層級為輸入的目的，我們設計另一個新的架構，將 parser 部分做替換，主要是對字元語意、文法角色等資訊利用 word2vec 來產生。字元時間前後資訊則使用 RNNLM 擷取，最後形成一個新的文脈訊息全部是由機器自行學習未標註語料所產生出來的，這樣能達成我們想使用字元層級的目的也能避開傳統 parser 的問題。以下為新系統架構以及新方法的詳細說明。

首先，圖 2 為我們的新語音合成系統架構，將語料經過正規化，文本轉拼音方面，由於我們的中文拼音字典並沒有歸納出哪些字應該何時讀破音字的資訊，在此選用第一組拼音來對應，再將拼音轉音素。我們更動的地方是將 parser 拿掉，以 word2vec 與 RNNLM 做替換，主要是用 word2vec 與 RNNLM 求取字元語意、文法角色資訊與字元時間前後資訊當作文脈訊息，以合成出新的語音。

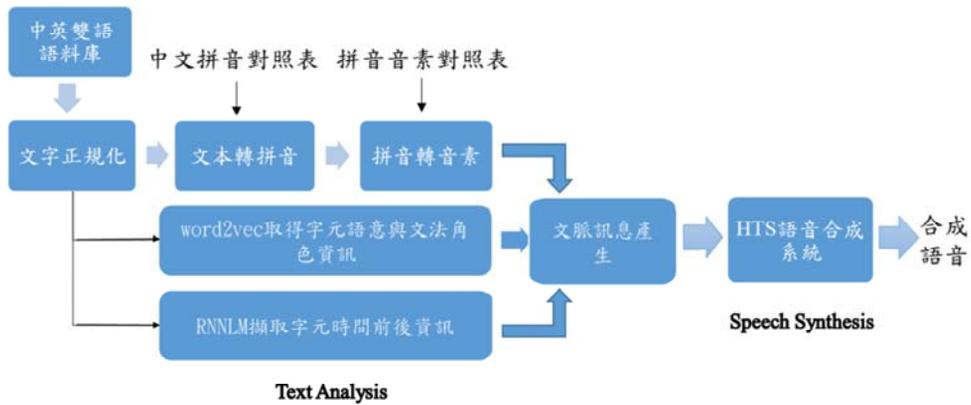


圖 2. 新語音合成系統架構圖
 [Figure 2. The proposed character-based TTS approach]

3.2 字元語意與文法屬性之文脈訊息擷取

文字探勘以及 NLP 在數據分析世界中一直是非常重要的部分。其中 word2vec 被頻繁的討論以及使用，因為使用它能夠將輸入的詞轉到向量空間上並進行演算，分析後可以發現在向量空間中，相聚在一起的詞向量轉換回文字後會是相同屬性的詞彙。也就是說它有能將字詞語意或文法角色做分類的能力，而且它無需給定標註過的資料庫就能對語料直接進行訓練，這非常符合我們想要避開人工標註的初衷。

Word2vec 是 Google 公司在 2013 年開放的一款用於訓練詞向量的軟體工具。它根據給定的語料庫，通過優化後的訓練模型快速有效的將一個詞語表達成向量形式，其核心架構包括 CBOW(Continuous Bag-Of-Words Model)和 Skip-gram。圖 3 為 word2vec 的核心架構圖。

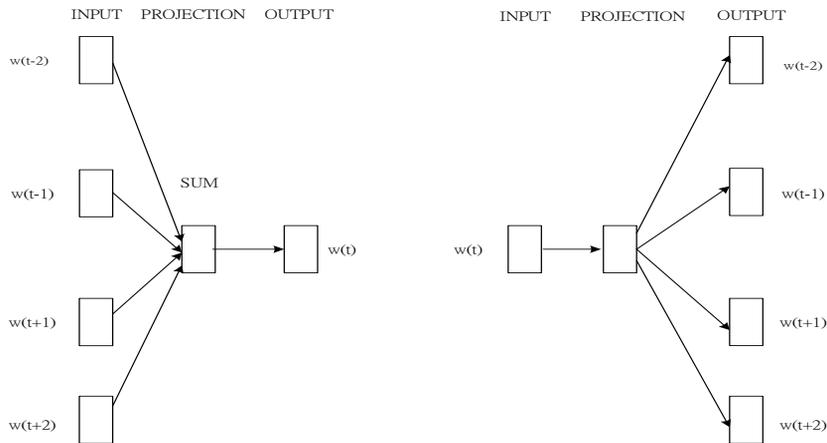


圖 3. word2vec 架構圖
 [Figure 3. The continuous bag of word (CBOW) and Skip-gram word2vec models]

3.3 字元時序狀態之文脈訊息擷取

在我們提出的字元概念下，我們希望能從文本中獲得字元在當前語句中的狀態，讓我們學習文章的脈絡，可以從這句預測下句可能為何。為了完成這功能，遞迴神經網路(RNN)可能是個不錯的選擇，例如(Mikolov & Zweig, 2012)基於 RNNLM 上下文相關性的研究也指出使用遞迴類神經網路模型進行訓練能從隱藏層中連續的輸出向量獲得字詞在句子中的狀態。

本文是以 Mikolov 改良的 RNNLM 來進行，遞迴式類神經網路包含輸入層 (input layer)、隱藏層(hidden layer)、輸出層(output layer)和類別層(class layer)。而 U 、 V 、 W 和 C 為各層的權重。 $w(t)$ 為輸入， t 依時間排序為 1 到 N ，也是 RNN 的權重， $s(t)$ 為隱藏層輸出也就是神經元(neurons)的值也是它的 state， $y(t)$ 為輸出須與輸入同維度。而 $c(t)$ 為類別層，Mikolov 提出輸出層分解可以降低語言模型中的運算複雜度，使訓練效率提高。圖 4 為 Mikolov 改良的 RNNLM 架構。

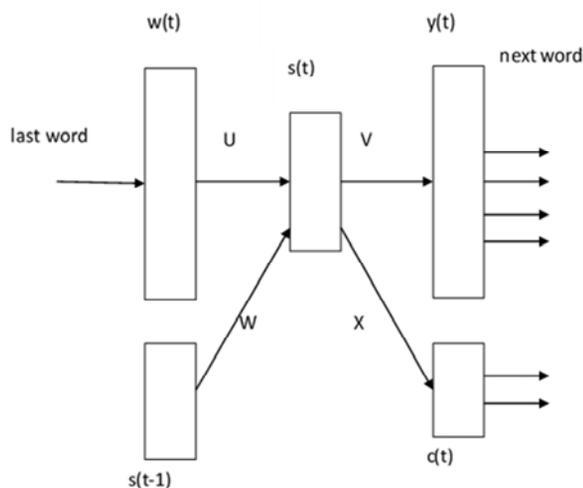


圖 4. Mikolov 改良的 RNNLM 架構

[Figure 4. The block diagram of the RNNLM model.]

遞迴神經網路最大的優勢在於，可以真正充分的利用所有上文訊息來預測下一個詞，不像其它神經網路只能一次看 n 個字，只能從前 n 個字來預測下一個詞，簡單來說 RNN 就是一個有隱藏層的自我相連網路，隱藏層同時接收來自 t 時刻的輸入和 $t-1$ 時刻的隱藏層輸出做為輸入，這使得 RNN 具有短期記憶能力，能夠學習到較長時間的文章脈絡，建立起語言模型。

表 2 為我們在一前置實驗中，利用 Chinese Gigaword 語料庫與 RNNLM 學習字元時序關係的結果，我們可以發現 RNNLM 模型產生出來的句子不管在語意，時間順序，或是流暢度而言都蠻貼近一般新聞的文字。

表2. 使用RNNLM 產生之文句

[Table 2. Experimental results on RNNLM-based sentences generation.]

時序方向	範例
Forward	在民主黨的表現，他們一致認為，一切不可能會對我們的壓力。
	警方今天在高雄縣警察局長劉松藩訪問時表示，這次的民進黨立法院黨團將在明天召開記者會，會中表示，民進黨決定將全力支持。
Backward	. 遇待國惠最的國美開離定決已，判談的國美與國合聯在國美兼理總副國美，示表統總李
	收作點3818以，點4十7百1漲大數指價股權加量行發，點95.591以貨期月期10.959金基太亞銷沖天今元美兌幣台新

因此應用 RNNLM 的特性來訓練文本，可以學習前面以及當前看過的句子來猜測下一的字或是下一句會是甚麼，也因為遞迴神經網路能夠這樣分析句子前後時間關係，所以能夠知道整個語句的脈絡，而這剛好與傳統文本分析的時間順序資訊相似，因此我們將其在我們的新架構上看看能否真的有用。

4. 實驗結果與分析

本實驗中，目的是為了使用新語言特徵取代現有 parser 的斷詞、POS 資訊，然後產生新的文脈訊息，代入語音合成系統合成語音，並與傳統方法 parser 做比較；新系統與舊系統的差別只有文脈訊息的不同，舊系統使用 parser 做文本的分析，而新系統使用字元層級的 word2vec 與 RNNLM 來分析文本，為求其公平性，新舊系統皆使用中英夾雜語料，訓練語句長度為段落訓練。而我們要比較的是新舊系統合成音檔之偏好，選用的合成語料是訓練語料中沒有被訓練到的部分，最後新舊系統比較的音檔皆為同一句話以示公平，而測試者不知道哪個音檔為新系統所合成避免分數灌水。

4.1 實驗設定

4.1.1 語料

我們使用的訓練語料、合成語料皆為我們與台灣數位有聲書協會合作錄製的” NTUT Audiobook Corpus Vol.2”，在此語料庫中我們請專業錄音員為我們錄製男聲語料，所有音檔皆在專業錄音室錄音，錄製時是以段落為單位唸完，保留語句之間的連結性；合成語料則從其中各別抽取中文 280 句及中英夾雜 160 句來做合成，抽出的句子皆不在訓練語料當中。表 3 為訓練語料資料表。

表3. 訓練語料資料表

[Table 3. Statistics of the speech corpus for speech synthesis experiments.]

	中文語料	中英夾雜語料-CE	純英文
文本內容出處	生命科學大師：遺傳學之父 孟德爾的故事(張文亮著)	線上文本 (工研院提供)	CMU
訓練語料總句數	約 4800 句	約 3500 句	約 990 句
訓練語料每句詞數	20-35 詞	10-30 詞	5-15 單字
訓練語料時間長度	約 172 分鐘	約 201 分鐘	約 79 分鐘

4.1.2 文脈訊息求取方法與設定

新方法與舊方法中只有前級文本分析不同，後級的語音合成系統(HMM-based Speech Synthesis System, HTS)的設定，兩者完全相同；舊系統文脈訊息依然採用 parser 來分析文本，新系統的文脈訊息則是去除 parser 產生的資訊，改使用字元層級的 word2vec、RNNLM 來分析，我們用 word2vec 將字元歸類成 64 類，RNNLM 隱藏層設 256 維，並為隱藏層中每個 neurons 的狀態設定門檻(threshold)把輸出量化成 0 或 1。在問題集中我們去掉有使用斷詞和 POS 的項目，然後添加使用 word2vec 和 RNNLM 產生的新項目，然後建立新的決策樹。表 4 為舊系統與新架構所使用的語言特徵，其中舊系統中紅色斜體字部分在我們新架構中將被廢除，新加入的藍色斜體字部分為 word2vec 與 RNNLM 的輸出參數。

表4. 傳統TTS系統和新架構的文脈訊息差異

[Table 4. Comparison of linguistic features used by the traditional and the proposed speech synthesis approaches.]

	傳統語言特徵	新架構語言特徵
音素(PHONE)	音素在音節中的位置	音素在音節中的位置
音節(SYLLABLE)	<i>音素數量，在詞中的位置</i>	X
詞(WORD)	<i>音節數量，在短語中的位置</i>	X
短語(CLAUSE)	<i>詞數量，在句子中的位置</i>	在句子中的位置
句子(UTTERANCE)	短語數量，在段落中的位置	短語數量，在段落中的位置
段落(PARAGRAPH)	句子的數量	句子的數量
WORD2VEC 類別	X	<i>字元是屬於哪一類</i>
RNNLM	X	<i>字元間前後時間順序</i>

4.1.3 語音合成設定

本研究的中英夾雜語音合成系統使用” NTUT Audiobook Corpus Vol.2” 語料和 HTS 合成；首先我們中文和英文聲音編碼統一使用 X-SAMPA 編碼為標準。在訓練語音合成模型時，所有錄音皆為 48KHz，聲學特性我們取 34 維梅爾倒頻譜係數(MFCCs)，音調(pitch)輪廓每 5 毫秒至 25 毫秒為一音框(frame)長度，最後每個音素(phone)我們使用 5 個狀態(state)的 HMMs 來訓練。

4.2 評估方法

系統偏好的評估方式是傳統使用 parser 的語音合成系統與我們提出使用字元階層提取特徵的新語音合成系統來做比較。我們以聲音的相似度、自然度和可理解度的評估方式，將測試音檔給 10 位以國語為母語的人士進行評分，新舊系統偏好度測試為 2 選 1 方式，為標準 A/B/X 測試，不存在兩者皆好；而新舊系統評分採平均主觀值分數(mean opinion score, MOS)來評估，其評分方式為 1~5 分。表 5 為測試音檔設定。

表 5. 測試音檔設定
 [Table 5. Statistics of the synthesized speech database for all evaluation experiments]

	NTUT Audiobook Corpus Vol.2	
測試類型	中文	中英夾雜
總句數	60 句	40 句
音檔數	10 個	20 個
每句字數	10-20 字	10-20 字

4.3 實驗結果

圖 5 與圖 6 分別為新舊系統在純中文與中英夾雜的相似度、自然度與可理解度偏好的比較，表 6 與表 7 則分別為新舊系統在純中文與中英夾雜的主觀 MOS 分數比較。評比偏好部分，從測試結果中我們可以發現，在圖 5 與圖 6 中不管是在相似度、自然度與可理解度來看，大部分測試者偏好新系統。而由表 6 來看新系統的聲音比傳統架構所合成出來的聲音稍微自然與相似原語者的聲音，但在表 7 中英夾雜測試分數看起來新舊系統分數差別不大，所以只能說兩者大概相當。不過概觀來看可以發現使用字元層級(character-level)的 word2vec 與 RNNLM 來取代傳統 parser 進行文脈訊息擷取，能合成出相當貼近人聲的聲音。

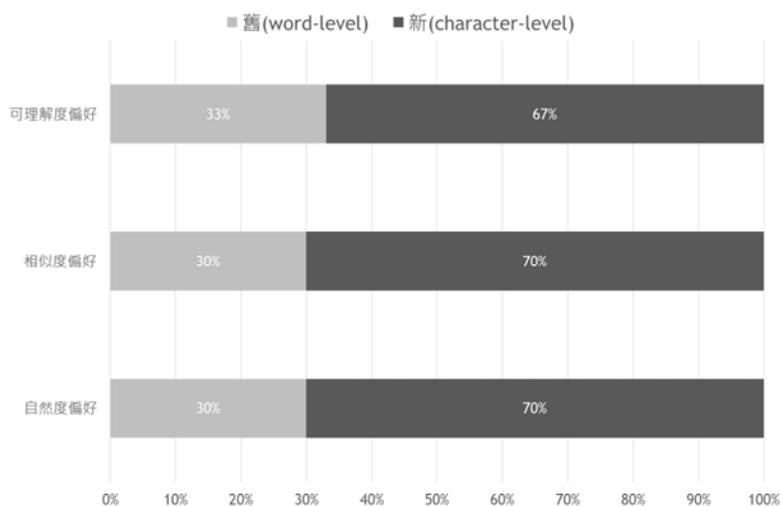


圖5. 新舊系統純中文偏好比較

[Figure 5. Experimental results of the A/B/X preference test on pure Chinese sentence synthesis]

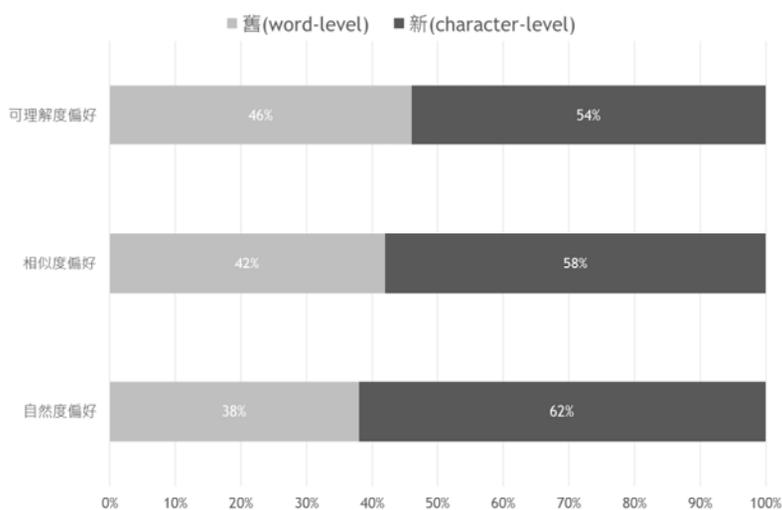


圖6. 新舊系統中英夾雜偏好比較

[Figure 6. Experimental results of the A/B/X preference test on mixed Chinese-English sentence synthesis]

表 6. 新舊系統純中文聲音 MOS 主觀分數比較
[Table 6. Comparison of the MOS scores of the conventional and the proposed approaches on pure Chinese sentence synthesis]

純中文	傳統語音合成系統	新語音合成系統
自然度評分	3.09	3.44
相似度評分	3.25	3.38
可理解度評分	4.27	4.27

表 7. 新舊系統中英夾雜聲音 MOS 主觀分數比較
[Table 7. Comparison of the MOS scores of the conventional and the proposed approaches on mixed Chinese-English sentence synthesis]

中英夾雜	傳統語音合成系統	新語音合成系統
自然度評分	3.26	3.22
相似度評分	3.26	3.36
可理解度評分	3.24	3.24

4.4 實驗討論

在本次實驗中新系統大概獲得 3 點多分，在市面上的 TTS 合成系統多為 4 分以上，雖比不上大公司的合成系統，不過這分數還算是合理，證明使用字元層級的文脈訊息在語音合成上是可行的。對此我們猜測新架構會更好的原因，是因為傳統 parser、POS 是人所設計的，但是真正口語上我們不一定會這樣唸，而我們新系統則是不靠人工設計，讓機器自己從語料當中學習句子關係，因此可能合成出來的聲音會比較流暢。不過在中英夾雜測試下兩者分數差距不大，可能是 RNNLM 在中英夾雜的英文句子部分少，所以無法有效得知字元在句子狀態，不過整體看來新架構確實有比較好的成績。但是本研究只是初步實驗新架構方法以實驗數據來證明真的會比傳統好，真正比較好的詳細原因需要未來繼續深入探討。

5. 結論

在本研究中，我們將一般語音合成中的文本分析做替換，將以前以詞為單位求取文脈訊息的方式，替換成以字元為處理單位。用 word2vec 求取字元的語意屬性與文法角色分類和利用 RNNLM 猜測字元在句子中的狀態，以這種方式能夠避開斷詞、POS 需要大量人工標註資料庫的缺點；而新舊系統在各項評比中都是新系統合成出來的聲音較為優良，所以我們提出的字元層級的文脈訊息擷取方法確實能達到相當甚至超越傳統方式的成績。

致謝

本研究感謝教育部『大學以社教機構為基地之數位人文計畫』（A36 號）與科技部專題計畫（MOST 104-2221-E-027-079, 105-2221-E-027-119 and 103-2218-E-027-006-MY3）支持。

參考文獻 References

- Brill, B. (1992). A SIMPLE RULE-BASED PART OF SPEECH TAGGER . In *ANLC '92 Proceedings of the third conference on Applied natural language processing*, 152-155.
- Ding, C., Xie, L., Yan, J., Zhang, W. & Liu, Y. (2015). Automatic prosody prediction for Chinese speech synthesis using BLSTM-RNN and embedding features. In *proceedings of 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 98-102.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R. & Schmidhuber, J. (in press). LSTM: A Search Space Odyssey. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, Retrieved from <https://arxiv.org/abs/1503.04069>
- Licstar, (2013 年 7 月 29 日)。Deep Learning in NLP (一)詞向量和語言模型。【部落格文字資料】。取自 <http://licstar.net/archives/328>。[Licstar. (2013, July 29). Deep Learning in NLP (1) Word embedding and Language model [Web blog message]. Retrieved from <http://licstar.net/archives/328>]
- Mikolov, T. & Zweig, G. (2012). Context dependent recurrent neural network language model. In *proceedings of 2012 IEEE Spoken Language Technology Workshop (SLT)*. doi: 10.1109/SLT.2012.6424228
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *proceedings of Advances in neural information processing systems 26 (NIPS 2013)*, 3111-3119.
- The Stanford Natural Language Processing Group. (2015). Stanford-Parser Version 3.6.0 Release in 2015/12/09: <http://nlp.stanford.edu/software/lex-parser.shtml>
- Zheng, X., Chen, H. & Xu, T. (2013). Deep Learning for Chinese Word Segmentation and POS Tagging. In *Proceedings of the 2013 Conference on EMNLP*, 647-657.
- 唐大任 (2002)。中文斷詞之研究 (碩士論文)。取自 <http://140.113.39.130/cgi-bin/g32/tugsweb.cgi?o=dntucdr&s=id=%22NT900435069%22.&searchmode=basic> [Tang, D.-R. (2002). *A Study of Chinese Parser*(Master's thesis). Retrieved from <http://140.113.39.130/cgi-bin/g32/tugsweb.cgi?o=dntucdr&s=id=%22NT900435069%22.&searchmode=basic>]
- 黃昭銘 (2010)。改善條件隨機域模型於中文斷詞 (碩士論文)。取自 http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search-c/view_etd?URN=etd-0203110-093833 [Huang, J.-m. (2010). *An Enhanced Conditional Random Field Model for Chinese Word*

Segmentation (Master's thesis). Retrieved from http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search-c/view_etd?URN=etd-0203110-093833]