

長短期記憶模型之忘記閘提取語意流暢度  
之架構以自閉症小孩說故事為例

**A Lexical Coherence Representation Computational  
Framework using LSTM Forget Gate For  
Autism Recognition**

劉于碩\*、陳金博\*、高淑芬<sup>+</sup>、李祈均\*

**Yu-Shuo Liu, Chin-Po Chen,**

**Susan Shur-Fen Gau and Chi-Chun Lee**

**摘要**

泛自閉症研究指出，由於口語表達能力的遲緩，跟典型孩童相比，自閉症孩童較無法敘說一個流暢的故事，因此在診斷自閉症孩童時，衡量其口語表達流暢度以成為一個診斷的重要指標。然而流暢度的評量，不是需要費時的人工標註，就是得用專家專門設計出的特徵來當指標，因此，本篇研究提出一種自動化直接資料導向的流暢度特徵學習架構，利用長短期記憶模型的遺忘閘導出語意流暢度的特徵，同時我們也利用自閉症觀察診斷量表中的評分細項來測試我們的流暢度特徵，結果上，利用我們提出的語意流暢度特徵來辨識自閉症小孩與典型小孩的任務上能夠達到 92% 的高準確率，對照傳統上使用語法、語詞使用頻率、潛在語義模型分析等地模型有顯著的提升。

---

\* 國立清華大學電機工程學系

Department of Electrical Engineering, National Tsing Hua University

E-mail: drugstore950337@gmail.com; cclee@ee.nthu.edu.tw

<sup>+</sup> 國立台灣大學醫學院心理系

Department of Psychiatry, National Taiwan University Hospital and College of Medicine

E-mail: gaushufe@ntu.edu.tw

這篇論文也進一步隨機打亂字序及句子順序，使典型小孩說故事內容變得不流暢的方式，來驗證我們提出的流暢度的意義，降維並將資料樣本可視化分析後證明我們的提取的特徵含有流暢度的概念。

### Abstract

Autistic children are less able to tell a fluent story than typical children, so measuring verbal fluency becomes an important indicator when diagnosing autistic children. Fluency assessment, however, needs time-consuming manual tagging, or using expert specially designed characteristics as indicators, therefore, this study proposes a coherence representation learned by directly data-driven architecture, using the forget gate of long short-term memory model to export lexical coherence representation, at the same time, we also use the ADOS coding related to the evaluation of narration to test our proposed representation. Our proposed lexical coherence representation performs high accuracy of 92% on the task of identifying children with autism from typically development. Comparing with the traditional measurement of grammar, word frequency, and latent semantic analysis model, there is a significant improvement.

This paper also further randomly shuffles the word order and sentence order, making the typical child's story content become disfluent. By visualizing the data samples after dimension reduction, we further observe the distribution of these fluent, disfluent, and those artificially disfluent data samples. We found the artificially disfluent typical samples would move closer to disfluent autistic samples which prove that our extracted features contain the concept of coherency.

**關鍵詞：**人類行為分析、流暢度、長短期記憶模型、自閉症、說故事

**Keywords:** Behavioral Signal Processing, Lexical Coherence Representation, LSTM, Autism Spectrum Disorder, Story-telling.

## 1. 緒論 (Introduction)

自閉症是一種普遍且又病症複雜的精神症候群，不同於一般較單純的精神疾病，每個自閉症患者之間的症狀異質性非常的高，此一特性也增加了診斷的困難性，因此許多研究致力於設計出各種評量的標準與特徵來做為人工自閉症診斷或者自動化自閉症診斷的標準，此類研究對於自閉症的即早治療與縮短診斷時程有極大幫助，也因此人類行為訊號處理的領域裡越發蓬勃發展。過去許多研究已針對自閉症與典型孩童的各種不同模態行為進行過分析，例如利用臉部不同的表情反應特徵來辨識自閉症的研究(Liu, Li & Yi, 2016)，以及利用孩童之間不同的語音表徵(Marchi *et al.*, 2015; Bone *et al.*, 2014)，諸如聲調或發音咬字等等，也有利用孩童敘事的文字內容去分析語法和用字作為分析自閉症辨

識特徵的研究(Regneri & King, 2016; Chorianopoulou *et al.*, 2017)。我們在此篇論文中提出一種資料導向的語意流暢度特徵提取架構，這個架構能夠利用自閉症以及典型孩童說故事的資料辨識出說故事的資料來自於自閉症孩童或典型孩童。

過去對於自閉症孩童說故事的分析時常使用許多不同的關鍵字特徵或是主題特徵，來觀察自閉症與典型孩童的差異(Rouhizadeh, Prud'Hommeaux, Roark & Van Santen, 2013; Rouhizadeh, Sproat, & Van Santen, 2015)。許多研究更指出相對於典型孩童，自閉症孩童較無法說出一個完整的語法流暢的故事(Losh & Gordon, 2014)，或是自然地描述故事發展內容的因果關係。分析顯示，自閉症小孩描述故事時，傾向於把故事中的一個個元素、事件單獨描述，缺乏流暢的串接或敘述各個情節之間的因果關係(Capps, Losh, & Thurber, 2000; Tager-Flusberg, 1995)。而語法表達的不流暢現象也已有研究觀察證實(Diehl, Bennetto & Young, 2006)。而在一篇 ACL2016 的研究中更利用了許多關鍵字性質標註表達整個文章的劇情走向與分佈，藉由這個導出的劇情發展狀況來當作衡量孩童說故事的流暢性。然而，以上方法皆須仰賴經過訓練的專業人士標註，標註十分費時，且會受限於只能在被設計的情況下使用。

因此，我們的研究使用資料導向的方法，以文字向量當作輸入，輸進長短期記憶模型，使用期訓練好的遺忘閘作為參數，讓機器自己學出表達資料中含有流暢度的時序性語法特徵。實驗中，我們比較以各不同的特徵來訓練針對說故事資料庫的自閉症辨識模型的準確率，而我們所提出的語意流暢度特徵所訓練的模型在實驗中達到 0.92 的準確率，高過於由知名的流暢度特徵，流暢矩陣(Graesser, McNamara, Louwerse & Cai, 2004; McNamara, Graesser, McCarthy & Cai, 2014)所訓練出來的模型。再者，啟發於最新關於理解神經網路黑盒子的研究(Li, Monroe & Jurafsky, 2016; Koh & Liang, 2017)，我們觀察改變原始資料的性質後的語意流暢特徵分佈的改變，來試著理解推敲神經網路這個黑盒子運作，並驗證我們導出的表徵的意義。實驗中，我們將典型孩童所敘述故事的字序和句序分別打亂，來模擬不同程度的不流暢，結果驚訝的發現，在以我們導出的流暢度表徵空間裡，被打亂後的不流暢典型小孩說故事資料分佈，會往無法說出流暢故事的自閉症小孩說出的故事資料分佈靠近，此結果驗證了我們導出的表徵的確含有流暢度的資訊。此篇論文內容安排如下：第二節為資料庫以及架構介紹，第三節為實驗結果分析與呈現，最後第四節為結論。

## 2. 研究方法 (Research Methodology)

### 2.1 資料庫 (Database)

這篇論文中，我們使用兩個資料庫，第一個是主要用來驗證我們所提取出的流暢度特徵辨識力的自閉症與典型孩童說故事資料庫，第二個則是從收錄了許多兒童讀物的童話故事網站，所爬下來的四種童話故事資料庫，第二個資料庫是用來預訓練使用。

### 2.1.1 資料庫一:自閉症與典型孩童說故事資料庫 (Database I)

資料庫一的蒐集方式是由專業培訓過的研究人員引導式的帶領自閉症孩童完成的，使用的故事書名為瘋狂星期二，故事書本身是一本繪本，書中每頁內容，都是一張圖片，與一句短句概要這頁的內容，而小孩子會根據圖片內容描述每一頁的劇情，完成整篇故事的論述。

而說故事的這個情境，便是標準的自閉症診斷觀察量表(Lord, Rutter, Dilavore & Risi, 2008)中的一個單元，此為一國際公認的臨床自閉症診斷量表，我們是與台大兒童醫院共同合作收集。這個量表是設計來觀測孩童的社交互動行為，因此設計了許多行為單元，來觀測並評量受測者的行為表現，單元包刮建構式作業、假扮遊戲、共同互動式遊戲、示範作業、圖片描述、看圖說故事等等。而在我們所使用的看圖說故事單元中，施測者會先引導一段引文，而孩童則會被要求根據後半部的繪本內容，完成後半部的故事，我們用相同的流程收集了典型孩童部分的說故事資料，逐字稿的內容舉例如：

“星期二的晚上，有一群青蛙在池塘上，他們突然乘著荷葉飛了起來，他們飛到了鄰近的小鎮，接下來換你說故事……”

在這個資料庫中我們總共收集了 67 位孩童的資料總共約有 28446 字的資料庫，每篇故事平均約有 424 字，67 位孩童中包括 31 位自閉症孩童(ASD)以及 36 位典型孩童(TD)。表 1 為資料庫相關敘述。

### 2.1.2 資料庫二:童話故事資料庫 (Database II)

資料庫二是從線上收藏不同種類童話故事的網站所收集而來，我們隨機選了四種篇幅長短約莫類似的故事種類，來作為我們預訓練的資料庫，這四種故事分別為俗諺故事、床邊故事、當代童話故事以及謎語故事。每種故事，我們都隨機選了 24 篇故事作為我們的預訓練資料集，總共 96 篇故事、59212 字，表 2 為相關的描述。

## 2.2 語意流暢度表徵 (Lexical Coherence Representation)

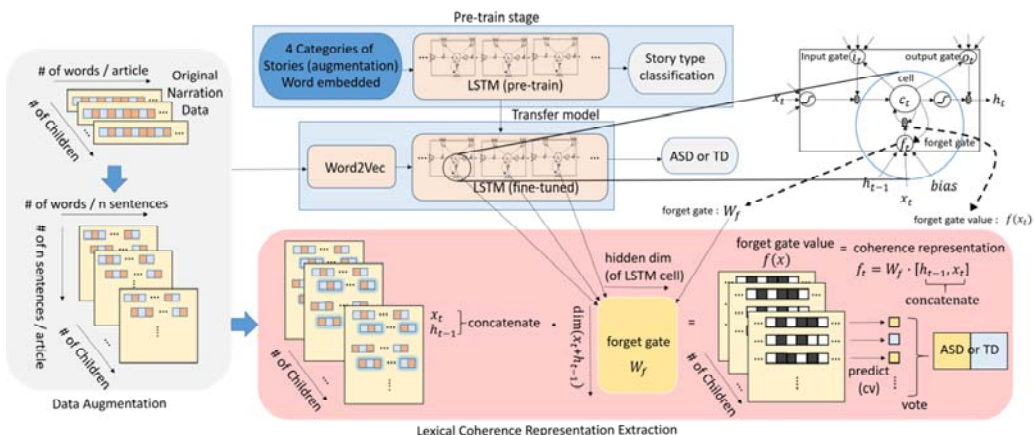


圖 1. 語意流暢度表徵提取實驗架構圖  
[Figure 1. Architecture of lexical coherence representation]

圖 1 呈現了用來提取語意流暢度特徵的架構流程，流程包含了資料擴增、中文文字向量訓練模型、資料庫二之長短期記憶模型預訓練、資料庫一之長短期記憶模型適性訓練以及最終由遺忘闡提取語意流暢度特徵。

### 2.2.1 資料擴增 (Data Augmentation)

研究證實，透過資料擴增的方式能夠使的神經網絡看過更廣泛的資料分佈，進而有助於穩定神經網絡訓練的整體偏差率，在神經網絡翻譯與語音辨識的研究中都已證實這個方法的幫助性(Fadaee, Bisazza & Monz, 2017; Hannun *et al.*, 2014)。在這篇文章中,我們使用一個簡單的滑動視窗方法來執行資料增加兩個資料集。我們用  $n$  個句子作為一筆資料，並使用與整個文檔相同標籤的樣本，我們每次移動一個句子來生成另一個  $n$  句的資料樣本。實驗也比較了由各種不同個句子數組成的資料樣本的結果，分別試驗了  $n$  為 1, 3, 5, 7, 9 的情況。

### 2.2.2 中文文字斷詞 (Chinese Word Segmentation)

近年來的文字相關研究中，都使用文字向量空間作為普遍公認的語意表徵，但不同於英文，中文字在書寫上，並沒有將詞語詞之間隔開，因此在進入文字向量的訓練之前，必須先經過斷詞的程式，這裡我們使用結巴斷詞(Sun, 2012)這個工具來斷詞。結巴使用概率語言模型來找出該句為最可能的詞組合。從統計資料,輸入字串分割模型的  $C = C_1, C_2, \dots, C_n$ , 和輸出  $S = W_1, W_2, \dots, W_m$ , 其中  $m \leq n$ 。對於一個字串  $C$ ，將會有不止一種可能的分割輸出  $S$ 。分割模型的任務是找到概率最高的輸出  $S$  集合。

$$\text{Seg}(c) \underset{SEG}{\operatorname{argmax}} P(S|C) = \underset{SEG}{\operatorname{argmax}} \frac{P(C|S)P(S)}{P(C)}$$

舉例來說:

C: 青蛙紛紛又再度回到水裡

S1: 青蛙 / 紛紛 / 又再 / 度回 / 到 / 水裡

S2: 青蛙 / 紛紛 / 又 / 再度 / 回到 / 水裡

S1 和 S2 便是不同的可能詞組合，根據不同的分割  $S$ ，詞數  $m$  的個數不同。 $m$  越大，實際觀測的可能詞組成概率越小。

### 2.2.3 中文文字轉向量 (Chinese Word2Vec)

與語音信號和視訊訊號不同，文本是離散的，不是連續可微的。為了使它能夠輸入神經網路，成為更好的進一步分析的表示，我們使用文字向量方法將我們語料庫中的每個詞投射到低維空間的特定協調上。因此，我們現在把離散文本轉換成連續可微的詞向量。

文字向量空間(Mikolov, Sutskever, Chen, Corrado & Dean, 2013)是一種資料驅動的學習表示形式，是從一個龐大的通用語料庫中學習的。在學習過程之後，模型將每個單詞

投射到隱藏的維向量空間模型中的座標上。我們可以通過計算不同詞間的協調距離來發現每個詞之間的關係。Mikolov 提出了兩種不同的方法，一是連續的詞袋(CBOW)，另一種是 Skip-gram (Skip-gram)。他還提出了兩種計算效率高的近似層次化和負採樣。利用這兩種有效的逼近方法，該模型可以以一種更有效的方式學習正確的代表方法(Goldberg & Levy, 2014)。

這裡我們使用 CBOW 來構建我們的詞嵌入網路。主要思想是利用 2c 相鄰詞的上下文向量來預測目標詞，模型結構包括 K 個詞彙量字典、輸入層、投影層和輸出層。輸入層的輸入是每個相鄰字的獨熱編碼(one-hot)向量。我們使用  $K \times V$  輸入層維度得到每個相鄰詞上下文向量  $w$ ， $w$  的維度為自訂大小，而後我們得到這些上下文的平均向量  $V(w)$  傳播到輸出層預測目標詞  $w$ 。最終，該模型可以使用反向傳播來更新參數，本篇實驗使用資料庫一以及資料庫二的全部文檔訓練出文字轉向量的網絡模型，而詞頻數過低的項目會被轉成 OOV(out of vocabulary)之向量。

#### 2.2.4 長短期記憶神經網絡 (Long short-term memory neural network)

我們的語意流暢度需要將說故事樣本，投射到長短期記憶神經網絡後，提取遺忘閘的輸出。因此在本節中，我們將首先簡要地描述長短期記憶神經網絡(Hochreiter & Schmidhuber, 1997)。長短期記憶神經網絡是一種定向時間序列神經網路。長短期記憶神經網絡的核心是其中包含的資訊狀態更新參數 $\tilde{C}_t$ :

$$\begin{aligned}\tilde{C}_t &= \tanh(W_C \cdot [h_t, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t\end{aligned}$$

使用長短期記憶神經網絡建模的好處在於它能夠調節長期或短期記憶體上下文中保留的信息量。調節機制是使用閘的結構來完成的，閘的結構被表述為權重矩陣和啟動函數。每個長短期記憶神經網絡有三個閘( $f_t, i_t, o_t$ )：

$$\begin{aligned}f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)\end{aligned}$$

在細節中,忘記閘 $f_t$ ,閘口負責控制過去的資訊被允許通過主要細胞 $\tilde{C}_t$ 的通過比率,我們利用此一機制來當作導出資訊流暢度的一種測量方法。我們首先使用監督式學習預訓練在四種故事類型的標籤上,作為初始化的模型參數,此長短期記憶網絡層的隱藏維度為 128 維。

#### 2.2.5 語意流暢度表徵 (Lexical Coherence Representation)

為了從資訊中導出流暢度的表達特徵，我們首先使用資料集二四類童話故事，訓練出來的模型權重，作為我們預訓練好的初始化參數。然後我們用我們預訓練過的模型通過分享模型參數，去對資料集一去進行適性化的微調，這是一種轉移學習的方式。經過對資

料集一進行微調，得到最後的結果。最後提取所學習的遺忘關矩陣 $W_f$ ，提取後，我們可以計算出，在啟動函數之前，沒有偏倚項的遺忘關的輸出值 $\bar{f}_t$ :

$$\bar{f}_t = W_f \cdot [h_t, x_t]$$

對於每個資料樣本，即  $n$  個句子，每個句子分別 $k_n$ 個字，會有總  $n \times k$  大小的 $\bar{f}_t$ 序列。我們將它們編碼為每個資料樣本都固定維度的表示形式。

$$F = g(\bar{f}_1, \dots, \bar{f}_T)$$

其中  $g$  表示 17 個統計泛函數，最大值(MAX)，最小值(Min)，平均數(Mean)，中位數(Medium)，標準差(Std)，1 百分位數(1%)，99 百分位數(99%)，99 百分位數- 1 百分位(99%-1%)數，偏度(skewness)，峰度(kurtosis)，最小位(min pos)，最大位(max pos)，下四分位數(lower quartile)，上四分位數(upper quartile)，上四分位數範圍(interquartile range)，冪(power)，1 差分(point difference)。這個  $F$  便是每個資料樣本的語意流暢度表徵。

### 2.3 暢度矩陣 (Cohesion Matrix)



圖 2. 流暢度矩陣特徵擷取網站示意圖  
[Figure 2. Extraction procedure of cohesion matrix]

流暢度矩陣被認為是對文本和話語最詳盡的自動評估之一。它計算書寫的一致性，並能用在度量口語和敘寫的內容(McNamara *et al.*, 2014)。測量方法包括相鄰句間的相似度計算、相鄰句或整篇文章中名詞重疊(重複)、文章句子結構的相似度等。例如，相鄰句間的相似度通過每句的 LSA(潛語義分析)向量計算。然後計算兩個 LSA 向量的餘弦相似度。

圖 2 為流暢度矩陣的使用介面，將欲計算流暢度之文章文字內容輸入後，便可勾選各項特徵，得到各特徵值，例如圖中勾選的左邊便是透過 LSA 向量計算語詞間語義的相似度，而右邊則是單純計算不同詞性字詞重複的程度。

我們使用所有的 20 維特徵，通過相關性特徵選擇，我們在選擇 25% 的特性集時獲得最佳性能。所選擇的五個特徵是局部名詞重疊、全功能變數名稱詞重疊、局部句子結構相似、全域句子結構相似以及句子中最小的單詞重複頻率。

局部名詞重疊是指計算相鄰句子中同一名詞的重複頻率，全功能變數名稱詞重疊是指計算同一名詞在一篇文章中千詞的重複頻率。局部和全域句子結構的相似性是指計算相鄰句子結構的相似性，計算整個句子結構的相似性。句子結構的相似性一般採用計算句子之間的編輯距離的方法(Ji & Eisenstein, 2013; Cheng & Liang, 2005)。編輯距離通常定義為將源句編輯為目標句(包括插入、刪除或切換)的成本步驟。最後一個特徵是每個句子中單詞重複頻率的平均值。

## 2.4 頻逆文檔頻率(TF-IDF)

方法主要由詞頻(TF)以及逆向文檔頻率(inverse document frequency, IDF)所組成，其數學式子表達成：

$$TF_{i,j} = \frac{c_{i,j}}{\sum_k c_{i,j}}$$

分子是第  $i$  個單詞在第  $j$  個文章的單詞計數，分母是第  $i$  個單詞在全部文章中的總出現次數。

IDF 值與目標詞在整個語料庫中的通用性有關。

$$IDF_i = \frac{|D|}{1 + |\{d \in D : t \in d\}|}$$

分母是整個語料庫的文章數量，分母是目標詞出現的文章數量。如果目標詞沒有出現在語料庫中，分母將為零。常用的方法是調整分母為  $1 + |\{d \in D : t \in d\}|$ 。

最終得到值的式子如下：

$$TF - IDF_{t,d} = TF_{t,d} \times IDF_t$$



### 3. 實驗設計及結果 (Experimental Setup And Results)

#### 3.1 實驗設計 (Experiment Setup)

在這篇研究中，我們使用我們提出的語意流暢度特徵表示在典型孩童與自閉症孩童之間進行分類。我們比較下列方法：

- (1) 詞頻逆文檔頻率:共 2034 維的特徵值
- (2) 流暢度矩陣:共 20 維的特徵值
- (3) 長短期記憶神經網絡:取出隱藏層的輸出來當作特徵進行分類
- (4) 語意流暢度表徵:為本篇所提出之表徵

實驗中使用了各項特徵作為流暢度特徵的比較，如 Coh-Metrix 主要是利用句子之間的相似度，來衡量一篇文章的流暢，所使用的概念是，當一篇文章裡相鄰的句子語義空間相似度高時，可以代表敘述連貫，反之則是不流暢。而透過 TF-IDF 的方法有機會讓模型透過關鍵詞數去學習，是否一篇文章一直使用過多重複的詞語，造成不流暢的現象，而 LSTM 則是在依序閱讀過段落文字之後，給出一個斷定流暢類或不流暢類的分類模型。我們提出的特徵則是保留了資料的個時間點的語義特徵分布情形，我們將這樣有各資料整個時間點語義分布的向量作為特徵，讓分類器去學習語義時序分布與流暢度的關聯。

圖 1 為以語意流暢度表徵為特徵之實驗架構圖，實驗所使用的分類器皆是支援向量機(SVM)和相關性的特徵選擇，差別僅在於使用來作為辨識基準的特徵抽換成不同的各項特徵。由於每一個兒童敘事包括不同數量的資料樣本(句數)，為了對於單個受試者給出一個單一的標籤，我們會對預測進行投票來決定最終個人的標籤。而準確率的驗證則採用交叉驗證，使用的評量方法是非加權平均召回率(UAR)。

再者，我們也利用我們提取出的流暢度特徵去對自閉症觀察診斷量表中的評分細項去做高低分分類，其中包括刻板的使用單字或片語程度、會話流暢性，以及報告事件的能力。

#### 3.2 實驗結果 (Experiment Result)

首先，觀察特徵選擇的結果，在 TFIDF 方法方面顯示連接詞、副詞和常見的短語在漢語中，如“然後”、“還好”、“這樣”、“什麼”都是自閉症孩童的敘述重點頻率明顯高於典型孩童。關於故事內容的幾個關鍵字，例如“屋頂”、“老奶奶”、“地板”、“住”，恰恰相反，於 TD (Typically developing)受試者的敘述頻率中顯著來的高。此外，流暢度矩陣的重要特徵是“名詞重複”相鄰的句子“和”名詞是整體的重複文章“相鄰句的句子相似性”。“整篇文章的句子相似”。而這些發現，確實符合自閉症患者使用較多虛詞，較無法表達流暢的口語的情況，也印證過去的研究得到的結果。

**表1. 不同特徵訓練模型準確率**  
**[Table 1. Performance of different features]**

Features	UAR/p value	Accuracy
Coh-Matrix	0.73/0.12	0.74
TFIDF	0.77/0.10	0.78
TFIDF + Coh-Matrix **	0.80/0.04	0.79
LSTM **	0.85/0.03	0.86
Lexical Coherence Representation **	0.92/0.007	0.91

在分類準確度方面，表 1 總結我們的實驗結果。獲得的最佳準確度是通過使用我們提出的語意流暢度特徵，達到了 UAR 92%。它比其他比較的特徵流暢度矩陣和 TFIDF 方法的 73% 和 77% 都要出色，且是顯著來的高。而且，有趣的一點是當與使用長短期記憶模型直接執行分類的特徵擷取，去進行比較時，使用我們所提出的語意流暢度特徵所訓練的模型，在辨識典型孩童與自閉症孩童文本內容的任務上還能有 8% 的召回率優勢，我們所提出的語意流暢度特徵是由長短期記憶模型的內部參數衍生而來的，而其似乎比使用整個長短期特徵模型提取的特徵更具有指標性。最後，對於不同句子數量來做資料擴充對於準確率的影響，表 2 列出了實驗的結果。資料擴充步驟中選定的使用句數，最佳數量似乎在 n=5 左右時會得到最好的效果。

**表2. 不同資料擴增參數下的準確率**  
**[Table 2. Performance of different augment set]**

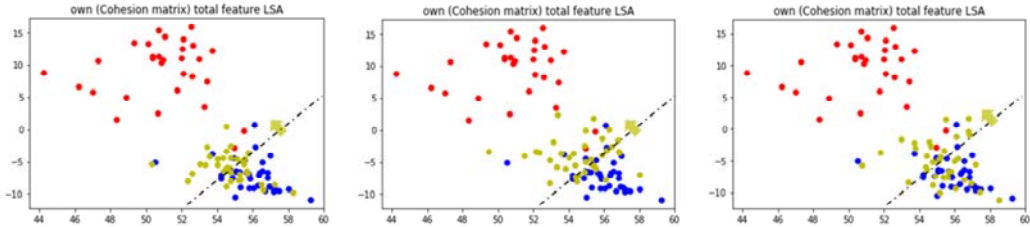
# of sentence	LSTM	Lexical Coherence
n=1	0.76	0.78
n=3	0.85	0.87
n=5	0.78	0.92
n=7	0.79	0.81
n=9	0.71	0.81

而在對自閉症診斷觀察量表(ADOS)有關口語能力的評分細項中(表 3)，我們發現此篇所提出的語意流暢度特徵，的確與評分細項中跟流暢度最相關的”會話流暢性”最相關，在高低分兩類分類的任務中能達到 80.48% 的準確率。

**表3. 流暢度相關行為屬性**  
**[Table 3. Attributes of behavior related to coherence]**

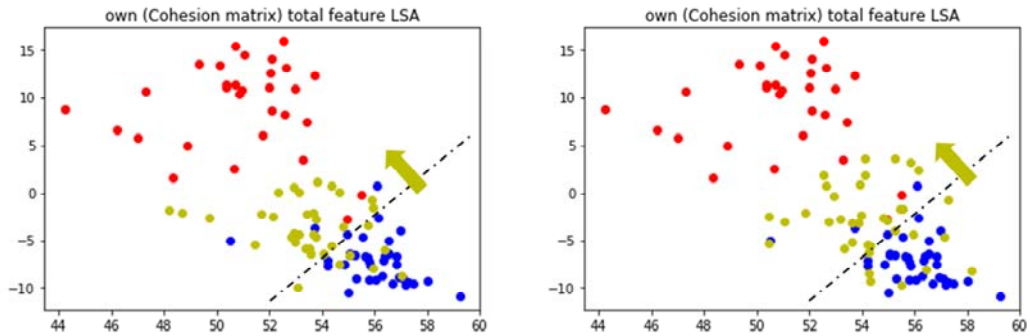
編號	細項內容	高低分類準確率
A4	刻板的使用單字或片語	0.5378
A8	會話流暢	0.8048
A7	報告事件	0.4208

### 3.3 實驗分析 (Analysis)



**圖 3.** 由左至右分別為打亂字序、句子順序、刪除連貫句子中特定字詞  
 [Figure 3. From left to right, the green dots represent the samples of random word order, random sentence order, delete random word in sentences]

我們的語意流暢度特徵模型實現了高識別準確性，但由於模型的複雜性，它缺乏了特徵的直接解釋性。因此，我們採用了類最近發表於 ICLR 的類似方法(Li *et al.*, 2016; Koh & Liang, 2017)來進一步瞭解分析我們的特徵。在這個分析中，我們首先通過隨機打亂典型小孩資料樣本中的單詞順序、句子順序，或是刪除連貫句中的隨機詞語，模擬出語無倫次的樣本，利用潛在語義分析來降維將特徵表示成二維圖，來觀察視覺化資料分布。然後我們進一步同時使用兩種我們上述所模擬的不流暢機制(同時打亂字序或句序、同時刪除特定字與打亂句序)，來觀察不同程度的不流暢情形在我們的空間中的資料分布狀況。



**圖 4.** 由左至右分別為同時打亂字序或句序、同時刪除特定字與打亂句序  
 [Figure 4. From left to right, the green dots represent the samples of random word and sentence order, delete random word and random sentence order]

藍點表示典型孩童的資料樣本，紅點代表自閉症孩童的資料樣本，黃點代表典型孩童故事的樣本通過模擬敘述不流暢的機制改變後的資料分布。觀察圖 4 能有趣的發現，黃色的點在典型孩童與自閉症孩童之間，顯示我們的表徵能反映出我們上述三種不流暢機制的情況。此外，隨著我們引入更多的不流暢機制(同時打亂字序或句序、同時刪除特定字與打亂句序)進入典型孩童樣本，黃點會更靠近不流暢的自閉症樣本。這個實驗分析表明，我們提出的特徵能夠反映不同的流暢機制，並且衡量文本不同的流暢程度。

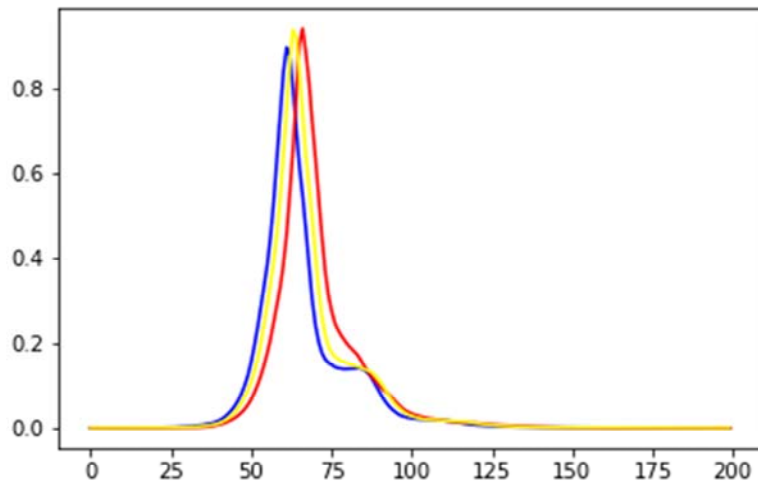


圖5. 樣本值分布圖  
[Figure 5. histogram of samples]

我們也畫出自閉症樣本、典型孩童樣本，以及導入不流暢機制的典型孩童樣本的值分布圖，從圖 5 能發現，導入不流暢機制的典型孩童樣本，的確在資料分布上更趨近於不流暢的自閉症孩童樣本。進一步我們也仔細比較各樣本中的統計值，並畫成樣本值分布圖來觀察資料分布的情況，在最大值(Max)、偏度的樣本值分布圖中，我們能明顯看出黃色的資料分布從原藍色分布區域移往紅色的資料分布區域。

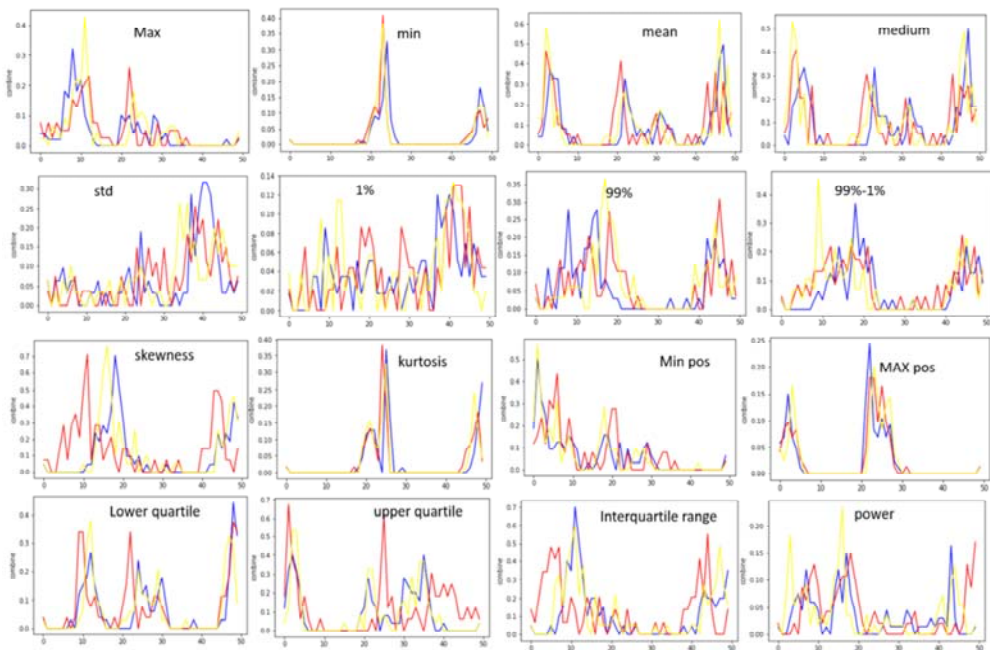


圖6. 樣本各統計值分布圖  
[Figure 6. histogram of samples with statistical analysis]

#### 4. 結論 (Conclusions)

本文提出了一種新的資料導向的語意流暢度特徵學習架構，其精隨是利用長短期記憶神經網絡架構內的遺忘閘中，資訊隨時間傳導的機制，模擬人類閱讀時會對不同時間點資訊有不同權重的特性，來推算出流暢度特徵。用我們所提出的語意流暢度訓練的模型，在自閉症辨識能達到 92% 的識別準確率，證明優於傳統衡量詞彙內容和連貫性的方法。最後，透過在典型孩童的資料樣本中導入類似不流暢的特徵(類比語無倫次的情況)，我們可以從視覺化後的資料樣本分佈點圖觀察到，將資料樣本映射到語意流暢度特徵的空間後，被導入不流暢特徵的典型小孩樣本點會移動得更加接近於自閉症小孩樣本。這樣的結果推論，即使我們的長短期記憶模型架構並非直接對流暢度標籤做學習，但在其內部架構的元素中似乎含有代表語意連貫性的流暢度的元素，甚至還能實現較高的準確率。

而在未來發展方向，首先是能去分析我們所導出的語意流暢度與計算語言學中設計的特徵之間的關係，若能找到與計算語言學中設計的特徵如語意、關鍵字、語法重複這些特徵之間的關係，便能更加瞭解神經網絡理解文章的方式。此外，流暢度也可以往多模態的方向發展，如語調流暢和手勢協調或臉部、肢體動作方面。最後未來可以套用更新的架構例如 Attention-based LSTM，並應用在現實世界的行為資訊分析，提高臨床價值將繼續是研究的中心目標。

#### 參考文獻 (References)

- Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P. & Narayanan, S. (2014). The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody. *Journal of Speech, Language, and Hearing Research*, 57(4), 1162-1177. doi: 10.1044/2014\_JSLHR-S-13-0062
- Bone, D., Lee, C.-C., Chaspari, T., Gibson, J., & Narayanan, S. (2017). Signal processing and machine learning for mental health research and clinical applications [perspectives]. *IEEE Signal Processing Magazine*, 34(5), 195-196.
- Capps, L., Losh, M., & Thurber, C. (2000). The frog ate the bug and made his mouth sad: Narrative competence in children with autism. *Journal of abnormal child psychology*, 28(2), 193-204. doi: 10.1023/A:1005126915631
- Cheng, S.-Y. & Liang, T. (2005). Chinese sentence similarity computing and applying. In *Proceedings of the 17th Conference on Computational Linguistics and Speech Processing*, 113-124.
- Chorianopoulou, A., Tzinis, E., Iosif, E., Papoulidi, A., Papailiou, C., & Potamianos, A. (2017). Engagement detection for children with autism spectrum disorder. In *Proceedings of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)*, 5055-5059. doi: 10.1109/ICASSP.2017.7953119
- Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative coherence of highfunctioning children with autism spectrum disorders. *Journal of abnormal child psychology*, 34(1), 83-98. doi: 10.1007/s10802-005-9003-x

- Fadaee, M., Bisazza, A., & Monz, C. (2017). Data augmentation for low-resource neural machine translation. Retrieved from arXiv preprint arXiv:1705.00440.
- Goldberg, Y. & Levy, O. (2014). word2vec explained: deriving mikolov *et al.*'s negative-sampling wordembedding method. Retrieved from arXiv preprint arXiv:1402.3722.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2), 193-202. doi: 10.3758/BF03195564
- Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., ...Ng, A. Y. (2014). Deep speech: Scaling up end-to-end speech recognition. Retrieved from arXiv preprint arXiv:1412.5567.
- Hochreiter, S. & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735
- Ji, Y. & Eisenstein, J. (2013). Discriminative improvements to distributional sentence similarity. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 891-896.
- Koh, P. W. & Liang, P. (2017). Understanding blackbox predictions via influence functions. In *Proceedings of the 34th International Conference on Machine Learning (PMLR 70)*, 1885-1894.
- Li, J., Monroe, W., & Jurafsky, D. (2016). Understanding neural networks through representation erasure. Retrieved from arXiv:1612.08220.
- Liu, W., Li, M. & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8), 888-898. doi: 10.1002/aur.1615
- Lord, C., Rutter, M., Dilavore, P. C., & Risi, S. (2008). *ADOS: Autism diagnostic observation schedule*. Boston, MA: Hogrefe.
- Losh, M. & Gordon, P. C. (2014). Quantifying narrative ability in autism spectrum disorder: A computational linguistic analysis of narrative coherence. *Journal of autism and developmental disorders*, 44(12), 3016-3025. doi: 10.1007/s10803-014-2158-y.
- Marchi, E., Schuller, B., Baron-Cohen, S., Golan, O., B"olte, S., Arora, P., & H'ab- Umbach, R. (2015). Typicality and emotion in the voice of children with autism spectrum condition: Evidence across three languages. In *Proceedings of Sixteenth Annual Conference of the International Speech Communication Association (InterSpeech 2015)*, 115-119.
- McNamara, D., Graesser, A., McCarthy, P., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511894664
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of the*

*26th International Conference on neural information processing systems(NIPS'13)*, 3111-3119.

- Narayanan, S. & Georgiou, P. G. (2013). Behavioral signal processing: Deriving human behavioral informatics from speech and language. In *Proceedings of the IEEE, 101(5)*, 1203-1233.
- Regneri, M., & King, D. (2016). Automated Discourse Analysis of Narrations by Adolescents with Autistic Spectrum Disorder. In *Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning*, 1-9. doi: 10.18653/v1/W16-1901
- Rouhizadeh, M., Prud'Hommeaux, E., Roark, B., & Van Santen, J. (2013). Distributional semantic models for the evaluation of disordered language. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)conference, 2013*, 709-714.
- Rouhizadeh, M., Sproat, R., & Van Santen, J. (2015). Similarity measures for quantifying restrictive and repetitive behavior in conversations of autistic children. In *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015*, 117-123.
- Sun, J. (2012). JiebaChinese word segmentation tool. Retrieved from <https://github.com/fxsjy/jieba>
- Tager-Flusberg, H. (1995). Once upon a Ribbit: Stories Narrated by Autistic Children. *British Journal of Developmental Psychology*, 13(1), 45-59. doi: 10.1111/j.2044-835X.1995.tb00663.x

