圖書資訊學刊　第18卷第2期　（2020.12）　頁73-101
doi: 10.6182/jlis.202012_18(2).073

基於檢索方法的中文幽默對話系統之建置應用與評估

Implementation and Evaluation of a Retrieval-based Chinese Humor Chatbot

曾元顯[1,6]　許瑋倫[2]　吳玟萱[3]　古怡巧[4]　陳學志[5,6]

Yuen-Hsien Tseng[1,6], Wei-Lun Hsu[2], Wun-Syuan Wu[3], Yi-Ciao Gu[4], Hsueh-Chih Chen[5,6]

摘　要

本研究發展幽默語料庫，開發相關的技術，實作一個基於檢索方法的「破冰機器人」系統，讓使用者透過對話找出相關的笑話，在與人互動時打破冰冷的氣氛，活絡陌生、尷尬的情境，最後評估其運用成效。透過資訊系統開發研究法的循環步驟，經過回饋後加入Word2Vec的查詢擴展、關鍵詞查詢提示，以及好笑話的隨機推薦等功能，讓使用者找不到笑話的比例從25.4%降低到16.7%，而系統達到的破冰效果從27.9%提升到39.9%。綜合而言，本研究不僅蒐集編製了近5,000則正體中文幽默語料庫，也建置中文幽默對話應用系統，語料與程式公開於https://github.com/SamTseng/icebreaker。本研究結論提供了實證經驗與意涵：自動化豐富笑話語料並確保其幽默程度，以及提供推薦功能，是提升此類服務成效的重要研發工作。

關鍵字：計算幽默、中文幽默對話、幽默語料、對話系統、破冰機器人

Abstract

This research is to construct a humorous corpus, develop related technologies, implement a retrieval-based “icebreaker chatbot” system which allows users to find relevant jokes for use in relaxing an unduly formal atmosphere when interacting with people, and finally evaluate its effectiveness. Through the iterative steps of the information system development research method, query expansion based on Word2Vec technology, frequent keyword prompts, and random

[1,2,3,4] 國立臺灣師範大學圖書資訊學研究所
Graduate Institute of Library & Information Studies, National Taiwan Normal University, Taipei, Taiwan
[5] 國立臺灣師範大學教育心理與輔導學系
Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei, Taiwan
[6] 科技部人工智慧生技醫療創新研究中心
MOST AI Biomedical Research Center, Tainan, Taiwan
* 通訊作者Corresponding Author: 曾元顯Yuen-Hsien Tseng, E-mail: samtseng@ntnu.edu.tw

recommendation of good jokes are added after user feedback. The results are that the proportion of user queries that fail to find jokes is reduced from 25.4% to 16.7% and that the icebreaking effect achieved has been increased from 27.9% to 39.9%. The importance of this research not only compiled a corpus of nearly 5,000 Chinese jokes, but also built a Chinese humor dialogue system, which have both been publicized at https://github.com/SamTseng/icebreaker for future use and verification. Empirical experience and implications of this study include: automating the richness and quality of joke corpus and providing recommendation service are important R & D efforts to improve the effectiveness of such services.

Keywords: Computational Humor; Chinese Humorous Dialogue; Humor Corpus; Dialogue System; Icebreaker Chatbot

Extended Abstract

1. Introduction

Humorous dialogue constitutes an important element in interpersonal communication, and it is also one of the important processes involved in human–computer interaction. The human–machine dialogue interface, which is commonly referred to as a chatbot in practice, has flourished since 2016. Applying humor techniques in human–machine dialogue services can often reduce user complaints and improve service quality.

The purpose of this study is to create a humor corpus, develop related technologies, implement a retrieval-based icebreaker chatbot system that allows users to find relevant jokes for lightening an unduly formal atmosphere when interacting with people, and finally evaluate its ice-breaking effectiveness of retrieved jokes.

Through the iterative steps of information system development, query expansion based on Word2Vec technology, frequent keyword prompts, and random recommendation of good jokes were added after user feedback was received. The results indicates that the proportion of user queries that failed to find jokes was reduced from 25.4% to 16.7% and the success rate of icebreaking increased from 27.9% to 39.9%.

The study findings may offer insights into applying humor in human–machine dialogue services in libraries, such as online reference services.

2. Development of a Chinese Humor Corpus

To commence our study, a humor corpus, commonly in the form of a joke collection, was

Note. To cite this article in APA format: Tseng, Y.-H., Hsu, W.-L., Wu, W.-S., Gu, Y.-C., & Chen, H.-C. (2020). Implementation and evaluation of a retrieval-based Chinese humor chatbot. *Journal of Library and Information Studies, 18*(2), 73-101. doi: 10.6182/jlis.202012_18(2).073 [Text in Chinese].

To cite this article in Chicago format: Yuen-Hsien Tseng, Wei-Lun Hsu, Wun-Syuan Wu, Yi-Ciao Gu, and Hsueh-Chih Chen. "Implementation and evaluation of a retrieval-based Chinese humor chatbot." *Journal of Library and Information Studies* 18, no. 2 (2020): 73-101. doi: 10.6182/jlis.202012_18(2).073 [Text in Chinese].

required. To develop a traditional Chinese humor corpus with sustainable and extendable value, we adopted the following steps: (1) select sources for collecting humorous jokes, (2) analyze the joke contents and define the revised Dublin Core fields (metadata) necessary for the corpus, (3) collect the jokes through various methods, and 4) remove near-duplicate jokes.

To diversify the joke contents, we searched and evaluated numerous joke sources and collected 5,615 jokes from 43 sources, including 27 public websites (2,777 jokes), 11 joke collection books (895 jokes), 3 free apps (156 jokes), personal collections (1,427), and jokes obtained from Facebook (360).

After analyzing a sample of jokes, we decided to catalogue the joke collection on the basis of the Dublin Core Metadata Element Set. The subsequent metadata was finally categorized into 16 fields. Some of the important fields include source identifier, source publication date, joke collection date, sharer, author, joke content, and funniness level (1 to 5). Notably, identifying the real author/creator or the owner of a joke's copyright was challenging because jokes are often narrated or revised in various manners before being disseminated in social media or websites or collected in books or apps. These important fields provided as much information as possible on the origin of the joke. Additionally, finding content for some fields was difficult, and such fields were left empty for future development.

As the collection of jokes expanded, the possibility of collecting duplicates from different sources also increased. Therefore, we subsequently developed a full-text matching tool based on a "bag of words" model and TFxIDF term weighting to detect near-duplicate jokes for removal. This step reduced the number of jokes from 5,615 to 4,696, with each removal verified by the authors.

3. Retrieval-based Chatbot: Icebreaker

To apply the above corpus in an effective manner, the following five prominent characteristics of humor were considered: subjectivity, regionality, cultural value, trending topics, and language differences. Each person may respond differently to the same joke based on their mood, understanding, or familiarity with the joke they have seen. Therefore, we constrained our application system to a specific scenario and for a certain group of people.

As a result, we designed a retrieval-based chatbot on LINE (a common social platform in Taiwan similar to Facebook Messenger) called Icebreaker for use by college students who had to make an oral presentation for their final project. This free chatbot allowed users to find relevant jokes to start their public presentation with in order to relax an unduly formal atmosphere, which is basically how we incentivized college students across various social media channels to use the chatbot. The chatbot had quick feedback buttons for users to rate the funniness level (from 1 to 3, because our preliminary test revealed the ineffectiveness of using a 5-point scale to evaluate funniness in this scenario) and to report whether the jokes helped achieve the icebreaking effect. A high funniness level reported by students did not certainly lead to the icebreaking effect, and vice versa, in our application experiment, although these two variables are positively and highly correlated.

4. Evaluation

We conducted two experiments at the end of each of two consecutive semesters. From June 8 to June 21, 2019, 67 users had made 493 valid joke queries. Additionally, during the other period (from December 22, 2019 to January 10, 2020), 132 users had made 1,344 queries. As expected, our simple poll showed that most users were aged 18-25 years.

The first experiment involved applying a vector space model for joke retrieval such that our Icebreaker could yield context-relevant jokes instead of just finding a random joke like Siri does. Of the 493 joke queries, the funniness level of the retrieved jokes was rated 298 times by the users. Additionally, the feedback demonstrated that 83 queries had achieved the expected icebreaking effect, with a ratio of 27.9% (83/298). Our log file revealed that 125 queries (25.4% = 125/493) did not yield any jokes (i.e., any retrieved jokes dissimilar to the query based on a similarity threshold were removed from being presented to the users).

According to the open-ended user feedback in the first experiment, we improved our chatbot through query expansion with Word2Vec, the random recommendation of good jokes, and query term suggestion. In the second experiment, 132 users made 1344 joke queries, and 639 feedback items reported that they were funny. Among the 1344 queries, 1038 (77.2% = 1038/1344) were prompted through the random joke function and 306 (22.8% = 306/1344) were made through keyword searches. The icebreaking ratio was 39.9% (255/639), indicating an increase of 12% from the first experiment (27.9% = 83/298). The query expansion based on Word2Vec reduced the search failure ratio from 25.4% to 16.7% (51/306).

5. Conclusion

Many reasons influence the icebreaker's effectiveness, including (1) the quality of the joke (the funnier the joke, the more effective it can be); (2) the availability of/accessibility to good jokes; (3) the atmosphere when telling a joke (if someone in the audience is an easy laugh, it will affect other audience members; this was difficult to verify from our experiments and could only be evidenced from the literature).

This research involved not only compiling a corpus of nearly 5,000 traditional Chinese jokes but also building a Chinese humor dialogue system, both of which have been published at https://github.com/SamTseng/icebreaker for future use, verification, and study. Implications of this study include the richness and quality of the joke corpus (collecting more jokes and identifying their level of funniness automatically) and the automatic recommendation feature, both of which require more R&D efforts to improve the effectiveness of such services.

Acknowledgement

This work is supported by the Ministry of Science and Technology (MOST 107-2221-E-003-014-MY2 and MOST 109-2634-F-002-023).

壹、前言

近年來Facebook、LINE等各大社交與即時通信軟體公司，推出人機對話平台，以「聊天機器人」或稱為「對話機器人」（chatbot）、「對話系統」形式，讓各商家使

用其應用程式介面（application programming interface, API）對廣大的使用者透過對話介面（conversational user interface, CUI）提供各種主動、被動的客戶服務，如：報名、通知、查詢、客服等。由於透過社交通訊平台的互動，可識別個別使用者，且新增服務不需使用者安裝軟體，再加上行動載具上語音識別與合成技術的進步與易用特性，使得CUI被視為是繼網頁（Web）與應用程式（App）介面之後，第三種重要的人機互動使用介面。光是2017年Facebook Messenger（Johnson, 2017）、LINE上面都各有超過10萬個CUI系統（或稱chatbot、bot）在其平台上提供使用者各項服務。

雖然人機對話系統的發展，已達商業應用階段，但多數以關鍵語句比對知識庫方式回應使用者，使用起來仍有生硬、呆板、低於預期之憾。要讓對話系統進步到更智慧、更人性的階段，需要更多語言處理、人機互動甚至幽默機制的相關研究。

幽默對話是人際溝通中一項重要的元素。具備幽默感的溝通過程，常可消解使用者抱怨的程度（Binsted, 1995），也是機器無法提供有效服務時候的備案（Bellegarda, 2014），甚至是主動留住使用者或贏得信任的利器。過去的研究也發現（Bryant & Zillmann, 1989; Mcghee & Frank, 2014），適當地運用幽默於課堂中，可提高學生的注意力，幫助他們學習得更有樂趣；幽默也能刺激開創性的想法，提高學生課堂的參與和互動；而在考試中，幽默可協助學生降低焦慮水平並改善他們的表現。此外，廣告、娛樂等產業，也是幽默應用的場域（Mihalcea & Strapparava, 2006b）。

然而技術上要實現流暢、合宜、適時的幽默對話系統，至今仍非常困難。對話機器人不僅需要辨識對話中的幽默氛圍，也要在適當時機生成幽默文意，亦即需做到幽默辨識（humor recognition）與幽默生成（humor generation）。即便是商業大廠如Apple、Google、Microsoft的對話系統，也僅能做到簡易的一次性幽默回應。例如，對Siri說：「講笑話」，Siri可以回應：「問：加油站不適合哪一種人工作？答：油腔滑調的人」，或是回應：「有一天火柴的頭很癢，他抓一抓就燒起來了。」若對Siri說：「講個中秋節笑話」，則其無法回應應景的笑話。

基於幽默在未來對話系統的重要性，以及目前可達到的技術，本研究之目的，在嘗試建置一套應用於幽默情境的中文對話機器人，簡稱為「破冰機器人」。具體而言，透過在臺灣普及率很高的LINE平台，建置基於檢索方法的對話介面，設定其應用情境為：使用者要與人互動，特別是要上台報告或演講、甚至初次接觸陌生人時，先以當前情境相關的主題詞彙查詢破冰機器人，獲得笑話文本後，自己講出來，以做為打破緊張、陌生、尷尬等氣氛的開場白。

為了實作這樣的破冰機器人，我們運用圖書資訊領域的採、編、典、藏、用概念，先評估各種笑話來源約40幾種，再廣

泛的取樣各來源的笑話約5,000則，使此語料具備多樣性，繼而編製詮釋資料，如笑話來源、笑話主題、好笑程度等，以便於後續的典藏、持續的更新；最後實作系統並上網公告此破冰機器人供網友使用，進而評估此項服務的效益、缺失，從中歸納出此類服務系統的特性，希望提供後續應用（如圖書館線上參考服務）輔助或是改進的參考。本研究產出的語料、程式、數據，呼應研究資料管理（research data management）的精神（Corrall, Kennan, & Afzal, 2013），開放於https://github.com/SamTseng/icebreaker，以激發我們謹慎檢查數據之正確性，並提供大眾後續便利的驗證、比較、研究與應用。

貳、文獻探討

本節先介紹文字對話系統的研發概況，繼而綜述國內外幽默語料庫建置與應用情形，再說明幽默生成的多種方法，最後概述幽默對話系統的研究，以做為啟發本研究的背景知識。

一、文字對話系統

本研究嘗試建置一套應用於幽默情境的中文對話機器人，需先瞭解文字對話系統的研發現況，而這方面國內外近年來發展迅速。分析已發表之文獻，歸納文字對話系統的技術，主要分為三種方法：規則法（rule）、檢索法（information retrieval, IR）、使用深度神經網路（DNN）的文字序列對應生成法（seq2seq generation）。由這三種，可再加以變化、組合，如IR + Re-ranking（重新排序）、IR + seq2seq、IR + seq2seq + Re-ranking等（Qiu et al., 2017）。分述如下：

規則法：可運用Artificial Intelligence Markup Language（AIML）人工智慧標註語言（Wallace, 2003），針對某特定領域，人工撰寫對話語料庫（Gang, Bo, Chen, Yi, & Zi, 2014），或是從客服實際對話紀錄中，將問題與回應轉成AIML知識庫，讓AIML比對引擎，自動比對使用者問題，並做適當回應。其優點是初期系統建置快速、系統的回應可預期並容易解釋；缺點是當要回應的領域範圍越來越大時，建構並維持知識庫的成本（人力、時間、經驗）會越來越高。目前網路上可公開取得的AIML知識庫，英文約有4萬多條問答規則，而中文僅有簡體中文約1,715條規則（多為招呼對話），正體（繁體）中文則完全沒有，也沒有正體中文的AIML知識庫比對引擎。除了運用AIML建構知識庫，也可以自訂語法與規則建構對話知識庫，以客製化的方式建構對話系統。Newyear與McNeal（2014）以AIML發展圖書館服務的對話機器人，透過對話紀錄的分析，調校其成效，讓此系統的回答正確率從12%逐步提升到83%。Kane（2015）也報導已上線運作至今的圖書館常見問題對話機器人，其發展出來的AIML知識庫均開源可下載（https://escholarship.org/uc/

item/4bs6s3hs），可惜這些參考服務機器人並不具備幽默對話的能力。

檢索法：在給定大量對話語料的情況下（如電話客戶服務對話），透過先進的資訊檢索方法，將使用者詢問的問題，去檢索對話語料庫中的問題，將最相似問題的回應，傳回給使用者，當作系統的對話回應（Ji, Lu, & Li, 2014）。其優點是可運用現今相當成熟的資訊檢索技術與工具，進行人機對話；其缺點是需要大量且清理乾淨的對話語料庫，才有可接受之成效。

序列對應生成法：運用近年來進展快速的深度神經網路，如回歸神經網路（recurrent neural networks, RNN; Wen et al., 2015）、長短期記憶體（long short-term memory, LSTM; Wen et al., 2015）等技術，並將（中文）文句斷詞（segment）後的詞彙（word）透過Word2Vec（Mikolov, Sutskever, Chen, Corrado, & Dean, 2013）等技術轉成隱含語意詞向量（word embedding vector），然後針對大量的人類對話語料，訓練出從文字序列產生另一種文字序列（seq2seq）的人工神經網路（Li et al., 2016），用來回應使用者詢問的問題。其優點是可產生出對話語料中沒有出現過的回應，擴增人工撰寫的回應，降低成本；缺點是訓練資料不足時容易產生不一致的回應或是無意義的文句。例如，開源碼CakeChat建議訓練資料最好有5千萬文字大小的對話，否則效果不佳（Ivanov, Khalman, Smetanin, Rodichev, & Fedorenko, 2019）。

二、計算幽默、笑話語料庫

在語料處理、人機互動、人工智慧的應用等領域，自1995年以來，已累積不少幽默辨識與幽默生成的技術研究，統稱為計算幽默（computational humor）。這些研究的目標，在探索幽默的計算模型（Bergen & Coulson, 2006）、增進人機溝通與使用者體驗（Morkes, Kernal, & Nass, 1999; Nijholt, 2006）、提升創意、動機、專注與記憶（Stock & Strapparava, 2006）、甚至協助溝通障礙人士以輔助其人際互動（Ritchie, Manurung, Pain, Waller, & O'Mara, 2006）等。

由於幽默機制的深度理解非常困難，計算幽默的研究，大略分為幽默的辨識（humor recognition）與幽默的生成（humor generation）。不論是幽默辨識或幽默生成，都需要事先建構幽默或笑話語料，以探索幽默特徵，便於後續的辨識與生成。底下簡述笑話語料的蒐集與其相關應用的研究。

Mihalcea與Strapparava（2006a）以10個英文小笑話（one-liner）做為種子查詢句，從網路上查詢包含該笑話之網頁（其網址必須含有oneliner、one-liner、humor、humour、joke、funny等詞），從中剖析出更多小笑話（例如：跟種子笑話一樣列在<li>的HTML標籤〔tag〕中），最後得出16,000句笑話（他們隨機檢視200個，約有9%的雜訊）。其範例有：「Change is inevitable, except from a vending machine」。除此之外，為了幽默辨識，他們也蒐集了與笑話文

字特性類似的負範例（非笑話），做為機器分類（辨識）笑話的學習語料。

前述的幽默語料是相對靜態不變的。在社交網路上內容快速變化的情況中，能夠辨識幽默文句，是值得研究的議題。Zhang與Liu（2014）便針對Tweet（推文），以自動下載、人工逐一判讀方式蒐集Twitter上幽默、非幽默文句，以及從http://textfiles.com/下載，再人工選取長度類似並排除重複的非Twitter上之幽默文句，各1,000句。其範例有：「when nothing goes right... go left」。

Chen與Lee（2017）根據TED演講的文字謄本，半人工選取了4,726幽默句（聽眾有笑聲之句子），並從該句子的前後7句，隨機選擇一句做為負範例，以進行笑話分類的測試。此語料雖有其創意，但有笑聲的句子，常需伴隨著前後文或是投影片上的文字才會好笑，否則該句本身並不一定好笑。

Blinov、Bolotova-Baranova與Braslavski（2019）則從各個公開的線上資源蒐集了俄羅斯語言的大量笑話語料，約15萬則，另蒐集了15萬則非笑話，也是為了機器笑話分類而建的語料。他們從笑話與非笑話語料中各取樣1,000則，以群眾外包方式判定，結果約有238則（12%）為雜訊。

上述研究都將幽默辨識轉換成二元的是、非分類問題，實際上幽默文句有其幽默程度的差異。語意評估國際工作坊（International Workshop on Semantic Evaluation, SemEval）在2017年舉辦的評測任務，便依此觀點進一步探索計算幽默的議題。主辦方Potash、Romanov與Rumshisky（2017）針對喜劇競賽電視節目任務主題的推文，半人工地蒐集整理了約8個月共112個主題、12,734條推文，然後要求參與者進行幽默程度的比較任務（因此，幽默是有情境的，有些甚至需要外部知識）。例如：「The host of Singled Out #BadJobIn5Words」比「Donut receipt maker and sorter #BadJobIn5Words」幽默。

Moudgil（2017）從多個公開網站以網頁擷取程式下載了231,657則英文笑話，並公布於開源碼網站：https://github.com/amoudgl/short-jokes-dataset。其範例有：「My wife and I were happy for twenty years. Then we met.」。由於數量眾多，跟前述的語料一樣，是否每則都是笑話，可能因人而異。例如：「Take my advice. I'm not using it.」可能對性格嚴肅的人，就不好笑。

在中文部分，中國大陸的任璐等人（2018）研擬一個適用漢語體系的「中文笑話語料庫」，他們收錄了33,025則笑話，並採用兩種分類方法，一種為按主題分類，另一種則按笑話產生原因分類。此中文笑話語料庫強調不對幽默跟笑話做區分，其標註體系，包含了笑話篇名、場景、人物、關鍵詞、幽默程度、幽默方式及笑話類別。另外，李璠（2017）用環境語篇之語料庫，來反映現實環境議題；李廣偉、戈玲玲與劉朝暉（2016）則建構了「言語幽默漢英平行歷時語料庫」；而劉鋒與張京魚（2015）自建的小型多媒體語料庫，則是分析學生的幽默對

話得來。臺灣則有鄭昭明、陳學志、詹雨臻、蘇雅靜與曾千芝（2013）蒐集160則各類型具代表性之中文笑話，以線上問卷評定方式邀請396位參與者，對每則笑話進行「理解程度」、「好笑程度」及「厭惡程度」的九點量尺評定，以供後續研究者挑選適當的笑話作為眼動反應、生理回饋、腦波反應等研究所需材料。這些都呈現了幽默的多樣性及實用性。

以上的描述可歸納出笑話有：主觀性、地域性、文化性、時事性以及語言差異等至少5種特性，適切的運用並不簡單，若沒有人工標記幽默程度，並多人進行確認，恐怕並非每則笑話對所有的人都有幽默的效果。

三、幽默生成技術

幽默笑話的生成技術，過去多使用語言資源與模版（template）來產生特定種類的幽默，然而近年來也有利用資訊檢索、深度神經網路的方法出現，簡述如下。

(一) 規則式幽默生成法

Binsted與Ritchie（1997）基於雙關語（pun、punning riddle）型態的語意與語法規則模版，設計出幽默雙關語產生器JAPE（Joke Analysis and Production Engine）。

Özbal與Strapparava（2012）根據同音異義的雙關語（homophonic puns）與隱喻（metaphors），結合英文WordNet與ConceptNet等資源，來產生創意性的名稱，特別是幽默的新詞（neologism）。

Stock與Strapparava（2003）在HAHAcronym計畫中利用失諧（incongruity）理論對既有的英文字頭語（acronym）產生有趣的詮釋版本。例如將技術方面的字頭語以宗教方面的詞彙重新詮釋，或是由使用者提供概念由系統產生新的有趣字頭語。其中一例是將原為International Joint Conference on Artificial Intelligence（人工智慧之國際聯合會議）的IJCAI，轉換成：Irrational Joint Conference on Antenuptial Intemperance（婚前不節制之荒謬聯合會議）。Stock與Strapparava（2006）提到他們運用WordNet Domains的領域標示，將宗教與技術、性別與宗教等領域對立成為失諧（incongruity）的來源，依此建立30萬則幽默。其使用到的語料資源，有發音詞典、WordNet、WordNet Domains、WordNet-Affect（基於WordNet擴展的情感詞彙）、首字母縮寫詞語法、字頭語文法、常識句庫、諺語和陳腔濫調（clichés）的語料庫以及成語典。使用這些資源的演算法包括傳統的語言處理元件，如：文句解析器（parser）、詞彙形態分析器（morphological analyzer）和特定的推理組件（reasoner）。

(二) 以資訊檢索為主的幽默生成法

在蒐集了16,000句小笑話後，Mihalcea與Strapparava（2006b）將相關笑話附加在電子郵件回應訊息或講義（lecture notes）上。其方法乃利用事先訓練好的分類器先將email訊息標示為快樂或悲傷其中一類，再運用隱含語意分析法（latent semantic analysis,

LSA），計算跟電子郵件訊息最後面30%的內容最相似的笑話，附加到被標示為快樂的電子郵件上。小規模的使用者評估調查顯示，其成效不錯，使用者會喜歡這樣的功能。

上述LSA類似資訊檢索方法找出適合的幽默句，只用到16,000句語料。Blinov、Mishchenko、Bolotova與Braslavski（2017）則以3種檢索方法實驗了更大量的語料。他們針對Yahoo!Answer幽默類別中使用者的問題，試驗了BM25、QTR（query term reweighting）、Doc2Vec（Le & Mikolov, 2014）3種資訊檢索方法，分別查詢約30萬條幽默的英文推文，以瞭解哪種資訊檢索方法效果較佳。實驗結果有些令人意外：傳統的BM25比其他兩種需要更多資源與處理步驟的方法稍好，其中QTR會用到Stanford CoreNLP的文句剖析以及名稱實體辨識（name entity extraction, NER），而Doc2Vec則用到英文的Wikipedia語料來訓練。作者解釋可能的原因是此幽默語料已經夠大，以致這三種方法之間的差異不明顯。其蒐集大量推特幽默語料的過程，是先從幾個「top funny Twitter accounts」的網站列表（http://www.hongkiat.com/blog/funny-twitter-accounts/），篩選出超過20,000個追隨者的帳號共103個，再根據這些帳號下載其被按讚或分享數超過30個的推文，然後再刪除重複者而獲得。其中一例，如：「Life is a weekend when you are unemployed」。此30萬條幽默的推文下載程式已公開提供使用（https://github.com/micyril/humor.）。

(三) 以機器翻譯、序列生成技術為主的幽默生成法

Du、Wan與Ye（2017）以相聲語料為素材，比較了資訊檢索、統計式翻譯以及seq2seq深度神經網路方法，針對逗哏角色（leading comedian）的發言，模擬捧哏角色（supporting role）的應對。他們從相聲書籍、相關網站、相聲錄音等1,551個相聲劇本中，擷取整理出150,000對兩個角色的對話，再以人工標記其中隨機選出的6,000對，結果有348對為幽默對話，以這些做為訓練與測試樣本。他們將逗哏的發言視為自動翻譯的輸入，捧哏的回應視為要翻譯出來的文字，除了採用統計式翻譯技術中的翻譯模型與語言模型外，也加入了Yang、Lavie、Dyer與Hovy（2015）提出的4種特徵來建立幽默偵測模型。實驗結果顯示此種加入幽默偵測模型的翻譯技術，效果比檢索、seq2seq都好一些。他們認為，可能語料不夠大，致使檢索與seq2seq的方法，都不夠好。而相聲的創作與表演越來越稀少，後續要蒐集更大語料將有困難（Du, Wan, & Ye, 2017）。

四、 幽默對話系統

上述的研究，較少實際應用在對話系統中。Augello、Saccone、Gaglio與Pilato（2008）報告了他們建構於Yahoo! Messenger的英文幽默對話系統，在當時只要結交Funnybot07@yahoo.it即可與之對話。他們運用AIML既有開源的知識庫資源，做為一般性對話的基礎，再蒐集笑話語料建

構AIML笑話知識庫，讓機器被要求時能做適當的回應。為了能「瞭解」使用者輸入的幽默對話，他們運用The Carnegie Mellon University Pronouncing Dictionary（CMU pronunciation dictionary）與WordNet實作了Mihalcea與Strapparava（2006a）以押頭韻、反義詞、成人俚語為特徵的幽默偵測方式，來回應使用者的笑話。其範例展示頗具娛樂效果，而小規模（自行上網蒐集的幽默、非幽默語料各100個短句）的偵測實驗，可達到73%的偵測準確度。

Sjöbergh與Araki（2009）以模組化方式設計一套日文聊天機器人，可依照不同使用者類型做個別化的回應，並可彈性增加其功能。其幽默辨識與生成模組雖然簡單，但仍勝過當時日本兩款聊天機器人的幽默能力。其運作方式是將使用者輸入轉發給各模組，各模組回覆後估計其回覆的自信度（數值在0到1之間），系統依使用者偏好調整每個模組的權重，模組權重及自信度相乘後，有最高數值者即被選為該次輸入的回應模組。例如歡迎模組，若偵測不到問候輸入，則輸出自信度為0，否則為1。其笑話模組，擷取輸入的內容詞，去搜尋內含3,000個笑話的語料庫，輸出其中一個笑話並給出數值1的自信度，否則為0。其幽默偵測模組，僅簡單比對使用者的輸入是否在笑話資料庫中，若是，則回應諸如：「哈哈，好笑！」之類的文句，並給出自信度1，否則為0。若同時有多個模組自信度不為0，如歡迎模組與笑話模組，則各自的自信度乘上該模組的權重後，若是歡迎模組數值較高，顯示該使用者較嚴肅不愛看笑話。依此方式，可針對使用者客製化該系統的回應，而添加新的功能模組也非常容易。

從以上的文獻評述可知，語料的蒐集、語言處理所需的外部資源，是極為重要的項目。根據認知理論發展出來的幽默特徵，在語料不夠大的時候，對自動化幽默偵測與生成會有幫助。而深度神經網路的方法，則是具有潛力的技術與方向。

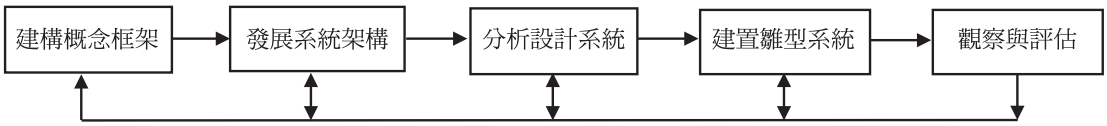
參、研究方法

對於以系統發展為主的研究，Nunamaker、Chen與Purdin（1990）提出了一套系統發展在資訊系統研究的方法（systems development in information systems research）。其研究流程包含了下列五個步驟的循環：(1)建構概念框架、(2)發展系統架構、(3)分析設計系統、(4)建置雛型系統，及(5)觀察與評估系統，如圖一所示。

本研究就上述步驟依序進行，並在本節先介紹前3項，再按步驟內容細節的重要程度分節探討。

一、建構概念框架

依據前一節的文獻探討，我們歸納出幽默具有：主觀性、地域性、文化性、時事性以及語言差異等至少5種特性。這些特性，讓每個人對同一個笑話，會因為目前的心情、理解的能力或是對笑話的熟習程度，而有不同的反應（覺得有趣或無趣）。為此，



圖一　Nunamaker等人（1990）提出的系統發展在資訊系統研究的方法示意圖

我們限定幽默對話系統的應用範圍，以便有效探討幽默對話系統的效益，從中獲取相關技術與應用的經驗。亦即，本研究針對特定對象、特定情境，提出破冰機器人對話系統的構想。其最終的目標，是期待在未來沒有限定範圍的應用中，本研究之經驗能適切地應用於輔助各項對話系統使其增加幽默對話的能力。

二、發展系統架構

依照上述的概念框架，並根據過去文獻的研究經驗，簡化前述Sjöbergh與Araki（2009）的設計，本研究提出圖二的系統架構，做為研究中文幽默對話的基礎。圖三則為使用者使用此系統的流程。

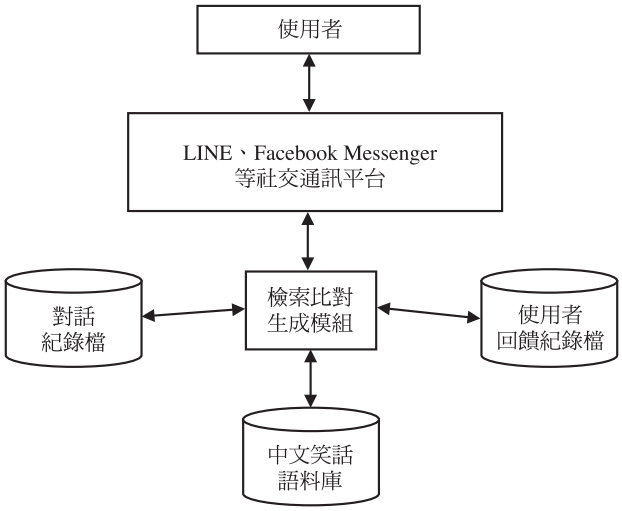
三、分析設計系統

圖二的系統架構中，較具關鍵性與挑戰性的工作，是笑話語料庫之建置以及檢索比對生成模組，將在後續兩節中，做較為詳盡的說明。

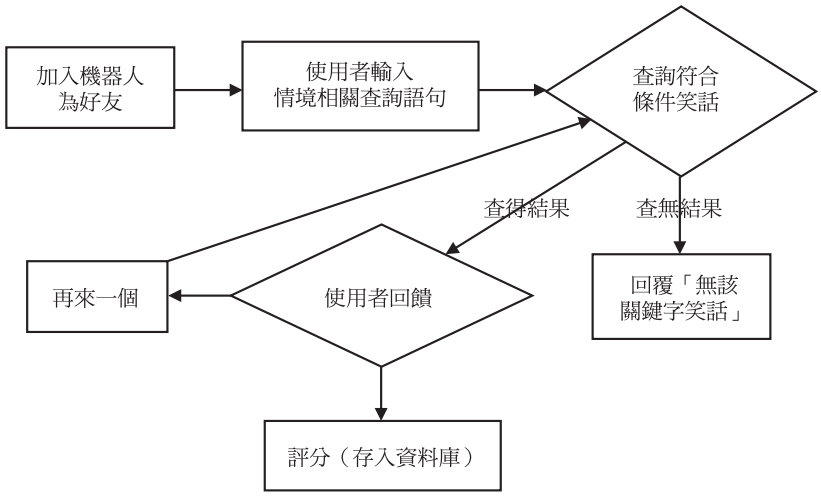
肆、笑話語料庫之建置

為了使笑話語料的內容多樣化，我們搜索和評估很多笑話來源，目前為止從43個來源收集了5,615則笑話，其中包括27個公開網站（共2,777則笑話）、11本笑話書籍（共895則）、3個免費App（共156則）、一位個人的蒐藏（1,427則）、以及一位專門蒐集改寫笑話的臉書朋友貼過的笑話（360則）。在分析了一些笑話樣本之後，我們決定參考都柏林核心集（Dublin Core）對此笑話語料進行編目，共有16個欄位，如表一所示。其中重要欄位包括：出處來源、發布日期、笑話收集日期、分享者、作者、笑話內容、笑話主題類別（共分9個類別），以及笑話的好笑程度（1至5共5個等級）。有些欄位無法得知其內容者先留白，後續有機會再進行加工編目時補充。另外，欄位資源識別代號（Identifier）標示笑話來源的網址或ISBN，如篇名「要求加薪」的笑話，其資源識別代號即為原笑話網址http://www.ak9k.com/6185.html。ID為笑話語料庫的唯一辨識碼：J開頭為個人蒐藏的笑話；L開頭則是收集的笑話；S開頭為後來新增學校類的笑話。流水號（Number）則為蒐集時笑話在各來源的流水編號。

由於笑話來源眾多，難免有重複或類似的內容，我們參考Tseng與Teahan（2004）的作法，以Python程式語言的gensim套件，透過向量空間模型全文檢索方法，偵測相似度



圖二　本研究規劃的中文幽默對話系統架構圖



圖三　中文幽默對話系統對話流程

高的笑話，將相似度高者再經人工確認後剔除，最後留下4,696則笑話。

本研究標記每則笑話的幽默等級或好笑程度從1到5，1為最不好笑，5為非常好笑。多數的笑話由2人評定（來自網站、書籍、App者），再取得共識決；而個人蒐集者則已有好笑程度的標記；臉書下載者則都一律給予程度4（好笑）之標記。表二顯示好笑程度及其則數，其中好笑程度3至5的比例佔72.2%。

表一　Dublin Core與幽默語料庫元素比較表

項目	都柏林核心集	幽默語料庫
(1)	題名（Title）	(1) 來源篇名（Source title） (2) 替代篇名（Alternative title）
(2)	著者（Creator）	(3) 作者（Creator）
(3)	主題和關鍵詞（Subject）	(4) 來源主題（Source subject） (5) 笑話主題（Subject）
(4)	簡述（Description）	(6) 內容（Text content）
(5)	出版者（Publisher）	刪除
(6)	其他參與（Contributor）	(7) 分享者（Sharer）
(7)	出版日期（Date）	(8) 公開日期（Public date） (9) 蒐集日期（Collection date）
(8)	資源類型（Type）	(10) 資源類型（Type）
(9)	資料格式（Format）	刪除
(10)	資源識別代號（Identifier）	(11) 資源識別代號（Identifier）
(11)	來源（Source）	刪除
(12)	語言（Language）	(12) 語言（Language）
(13)	關連（Relation）	刪除
(14)	涵蓋時空（Coverage）	刪除
(15)	版權規範（Rights）	刪除
		新增： (13) ID、(14)流水號（Number）、(15)笑話長度（Length）、(16)好笑程度（Level）

表二　笑話語料庫的好笑程度與其數量統計表

好笑程度	1	2	3	4	5	總計
則數	384	912	1,851	1,310	239	**4,696**
百分比	8.2	19.4	39.4	27.9	5.1	100.0

伍、檢索比對生成模組

此模組負責將使用者輸入的查詢語句，比對笑話語料，依照相關或相似程度，回應一則笑話給使用者。若使用者輸入相同查詢語句（或選擇再來一則笑話），則搭配對話紀錄檔的運用，回應下一則相關的笑話，直到沒有相關的笑話為止。為記錄使用破冰機器人的實際成效，此模組亦負責使用者回饋的應對與記錄，以便後續的分析、評估。然而其最重要的任務，是資訊檢索的部分。

使用者的輸入，可以是一句跟其當時情境相關的語句或詞彙，如「找出中秋節相關的笑話」或是僅有「中秋節」一詞。而要比對的笑話，有將近5,000則。為求快速的回應，我們採用向量空間模型（vector space model, VSM）資訊檢索技術，並於後續採用Word2Vec詞向量進行查詢擴展（query expansion），以提升檢索的成效。

VSM是經典的傳統資訊檢索方法（Salton, 1989）。其將語料中每份文件的重要詞彙（有主題意義的詞彙），都視為向量中的一個維度，而詞彙在文件中的出現次數（term frequency, TF）以及在整個語料中出現篇數的倒數（Inverse document frequency, IDF）的乘積（TF × IDF），常做為該維度的權重。如此n篇文件的語料庫若共有m個詞彙，就形成一個$m \times n$的矩陣，其中每一行向量對應到每一篇文件，而每一列向量則對應到每一個重要詞彙。依向量餘弦公式（cosine），可計算任意兩文件或是兩詞彙的相似度。

另一種VSM的表示法，則跟語料無關，單純以「獨熱編碼」（one-hot encoding）表示。亦即m個詞彙，每個詞彙都佔一個維度，該詞彙在該維度上的值為1，其餘為0。例如，假若全部詞彙只有3個：政治、經濟、運動，則其獨熱表示法，分別為[1,0,0]、[0,1,0]、[0,0,1]。其優點是(1)詞彙跟其向量的對應，只需簡單的查表即可；(2)很多機器學習演算法，只能做二分法，亦即偵測一個詞彙有出現或沒出現，因此需用到獨熱編碼。其缺點是m經常數以萬計。

上述兩種VSM的問題，在於用個別詞彙做為向量的維度：若有不同詞彙卻語意相近時，因屬不同維度，也無法增加其相似度，造成詞彙不匹配問題（vocabulary mismatch）。例如：「宇宙」跟「太空」，以VSM表示的話，其相似度為0。因此，在1990年左右，隱含語意索引法（latent semantic indexing, LSI）或是稱做隱含語意分析法（latent semantic analysis, LSA）被提出來（Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990）。其運用線性代數的奇異值分解（singular value decomposition, SVD）方法，將$m \times n$的矩陣降維（dimension reduction）轉換出$d \times d$的主題矩陣，其中$d < m$且$d < n$。亦即，語料C被降維，並以新的矩陣來近似整個語料，如下式：

$$C_{m\times n} = U_{m\times r}\Sigma_{r\times r}(V_{n\times r})^T \approx U_{m\times d}\Sigma_{d\times d}(V_{n\times d})^T \quad (1)$$

個別文件（或詞彙）的向量仍可從這個降維的矩陣算出近似值，然後依此亦可算出任意兩篇文件（或是任意兩個詞彙）的相似度。此種降維的作法，讓語意相近的文件（詞彙），被放在同一維度，解決了前述詞彙不匹配的缺點。如圖四範例所示（見註一）。

圖四中語料C矩陣有「星際大戰」等5篇文件，若只蒐錄7個詞彙，各詞彙出現的次數（或權重）表示在等號左邊的矩陣。等號的右邊有三個矩陣，分別為詞彙到主題的U矩陣、主題矩陣、以及主題到文件的V

$$
\begin{array}{c}
T1\\T2\\T3\\T4\\T5\\T6\\T7
\end{array}
\begin{bmatrix}
\mathbf{1} & \mathbf{1} & \mathbf{1} & 0 & 0\\
\mathbf{3} & \mathbf{3} & \mathbf{3} & 0 & 0\\
\mathbf{4} & \mathbf{4} & \mathbf{4} & 0 & 0\\
\mathbf{5} & \mathbf{5} & \mathbf{5} & 0 & 0\\
0 & \mathbf{2} & 0 & \mathbf{4} & \mathbf{4}\\
0 & 0 & 0 & \mathbf{5} & \mathbf{5}\\
0 & \mathbf{1} & 0 & \mathbf{2} & \mathbf{2}
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{0.13} & 0.02 & -0.01\\
\mathbf{0.41} & 0.07 & -0.03\\
\mathbf{0.55} & 0.09 & -0.04\\
\mathbf{0.68} & 0.11 & -0.05\\
0.15 & \mathbf{-0.59} & 0.65\\
0.07 & \mathbf{-0.73} & -0.67\\
0.07 & \mathbf{-0.29} & 0.32
\end{bmatrix}
\times
\begin{bmatrix}
12.4 & 0 & 0\\
0 & 9.5 & 0\\
0 & 0 & 1.3
\end{bmatrix}
\times
$$

星	地	星	電	海
際	心	際	子	角
大	引	效	情	七
戰	力	應	書	號

$$
\begin{bmatrix}
\mathbf{0.56} & \mathbf{0.59} & \mathbf{0.56} & 0.09 & 0.09\\
0.12 & -0.02 & 0.12 & \mathbf{-0.69} & \mathbf{-0.69}\\
0.40 & -0.80 & 0.40 & 0.09 & 0.09
\end{bmatrix}
$$

圖四　降維矩陣範例

矩陣。從對角線矩陣可知，此語料C其實只有兩個主題比較重要（姑且稱為「動作科幻」、「文藝愛情」兩個主題），且重要程度分別為12.4與9.5，而第三個主題權重只有1.3，相較之下可以忽略。因此，等號右邊的三個矩陣都可以再縮減成7 × 2、2 × 2、2 × 5的矩陣。這時，若再多收納兩個新詞彙T8（如：宇宙）、T9（如：太空），且其在這5篇文件出現的情況分別為：T8=[5,0,0,0,0]與T9=[0,4,5,0,0]，則此兩詞彙的cosine相似度為0。但若其向量轉換到主題空間，亦即各自乘以縮減後的V矩陣，如下：

$$
T8^{*}=\begin{bmatrix}5 & 0 & 0 & 0 & 0\end{bmatrix}\times
\begin{bmatrix}
0.56 & 0.59 & 0.56 & 0.09 & 0.09\\
0.12 & -0.02 & 0.12 & -0.69 & -0.69
\end{bmatrix}^{T} \quad (2)
$$

$$
=\begin{bmatrix}2.81 & 0.63\end{bmatrix}
$$

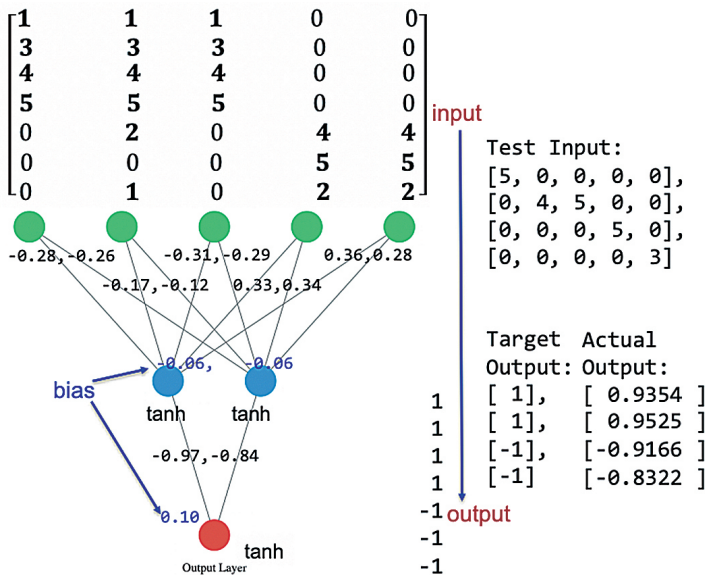
$$
T9^{*}=\begin{bmatrix}0 & 4 & 5 & 0 & 0\end{bmatrix}\times
\begin{bmatrix}
0.56 & 0.59 & 0.56 & 0.09 & 0.09\\
0.12 & -0.02 & 0.12 & -0.69 & -0.69
\end{bmatrix}^{T} \quad (3)
$$

$$
=\begin{bmatrix}5.18 & 0.52\end{bmatrix}
$$

則轉換後的詞彙向量T8*=[2.81, 0.63]與T9*=[5.18, 0.52]，其cosine相似度高達0.99，表示詞彙T8與T9屬於同一主題的訊息相當明確。

而現今神經網路的學習能力，可將C矩陣送入學習，得出如上述將五維詞向量轉換成二維詞向量的功能，如圖五（筆者自製）神經網路與其連結上的權重所示。其中間兩個神經元的輸出值，相當於公式(2)或是(3)的二維向量；亦即相同主題的五維整數向量輸入，會有相似的二維實數值輸出。

而獨熱編碼的詞向量，也可從高維度的整數向量，以各種嵌入語意的方式降維成低維度的實數向量，稱為詞嵌入（word embedding）過程。Mikolov、Sutskever、Chen、Corrado與Dean（2013）發展出稱為Word2Vec的詞嵌入技術，其運用You shall know a word by the company it keeps（Firth, 1957）的原則，透過大量的語料，將每個詞彙都轉換成300維（見註二）的實數向量



圖五　透過神經網路的學習，可將五維的整數詞向量，轉成二維的實數詞向量

（如T8*與T9*，只是此兩詞彙向量只有二維），並具有下列加、減法的類比特性：

Word2Vec（國王）－ Word2Vec（男人）＋ Word2Vec（女人）≈ Word2Vec（皇后）　　(4)

亦即國王的詞嵌入向量，減去男人的詞向量，再加上女人的詞向量，會近似皇后的詞向量。又如，圖六的詞向量運算程式，透過詞嵌入向量，在四個詞彙中，可計算出跟早、午、晚餐比較起來，最不相同的是豆漿這個詞彙；而跟中秋節最相似的詞彙，有：端午節、元宵節等，而可以運用於查詢擴展。換句話說，這些詞向量，幾乎抓住了人們對於詞彙語意的認知，只是這些詞向量都是數字而已。

```
>>> import gensim
>>> model = gensim.models.word2vec.Word2Vec.load("wiki.zh.model")
>>> print(model.wv.doesnt_match(['早餐', '豆漿', '午餐', '晚餐']))
豆漿
>>> print(model.wv.most_similar('中秋節', topn=4))
[('端午節', 0.8020), ('元宵節', 0.7719), ('清明節', 0.7596), ('重陽節', 0.7514) ]
```

圖六　詞向量運算範例

陸、建置雛形系統

我們在LINE平台上，建立一個機器人帳號，後台介接一套基於資訊檢索的笑話查詢互動系統，並記錄使用者的查詢語句以及回饋資訊。其互動介面，如圖七所示。其中使用者可對名稱為「小明同學」的破冰機器人（可用手機上的語音輸入）提問任何詞彙或語句，「小明同學」會找出最相似的笑話提供使用者參考，並請使用者回饋該則笑話的好笑程度，繼而詢問此則笑話對使用者而言是否有達到破冰的效果。其中破冰（icebreaker）指彼此不認識的人見面時活躍氣氛的遊戲或笑話（Cambridge Advanced Learner's Dictionary, 2019）。故只要可以緩和當下的氣氛，而順利開始互動，則稱其有達到破冰效果。這些訊息都記錄在資料庫中，以供後續的分析使用。

雖然對話系統主要以文字方式跟使用者互動，然而實際的商用系統會提供選項按鈕，來限制並簡化使用者的文字輸入。畢竟，有時點一下按鈕，比語音或文字輸入都要快速、方便，因此這樣的選項按鈕，也是對話過程的一部份。在圖七的例子中，使用者輸入「中秋節」後，根據系統的回應，接著點選「有點好笑」，LINE的App就幫使用者送出「有點好笑」的文字給系統，不僅對使用者而言省時省力，也讓系統蒐集到用詞一致的對話回應。

此外，圖七也顯示此系統不像Siri那樣每次只能透過一輪對話講出一個笑話，還(1)可讓使用者以查詢語句觸發不同的笑話，而能與使用者的使用情境結合；(2)會記錄使用者查過什麼詞彙、看過哪些笑話，而對同一使用者、相同查詢提供不重複的笑話。

柒、觀察與評估系統

本研究之實驗情境設定為：「使用者上台報告前使用破冰笑話機器人查詢笑話，現場說出查獲並自認為可用之笑話，自己講出來後，可否緩解現場陌生、緊張之氣氛，而達到破冰的效果？」故將實驗時間設在大學期末報告週：第一次實驗時間為2019年6月8日至6月21日之間；第二次實驗時間為2019年12月22日至2020年1月10日之間，因第二次實驗期間適逢本國總統大選，故實驗時間延長一週。實驗前利用社群網站Facebook、Plurk及Dcard進行上述應用情境的宣傳。第一次實驗吸引了75人使用，其中僅有8位是我們認識的人，扣除後為67人共查詢493次；第二次實驗共有141人使用，其中9位是認識者，扣除後為132人共查詢1,344次。為即時蒐集到使用者的建議，另外製作了回饋表單於聊天室中，使用者可隨時回饋意見。初步分析顯示使用者年齡多分布在18至25歲，與預期的使用族群一致。

然而，在此必須強調，此項實驗沒有強制性。我們僅宣傳鼓勵使用者於上台報告前可運用此系統來輔助講出笑話，以達破冰的效果。但使用者是否用在預期或類似的情境中，我們不得而知。也有可能使用者事先試用，而將整個對話與回饋流程跑完。因此，此實驗可視為是一種模擬實驗。但因為使用

圖七　建構在LINE平台上的破冰機器人「小明同學」的使用範例

註： 圖左為使用者輸入「中秋節」的笑話範例；圖中及圖右為連續操作範例：使用者輸入「講笑話」，點選「好笑」、再點選「有」的連續操作畫面。

者是自願、非我們安排的人員在操作系統，此實驗數據仍有其參考價值。

在建置笑話語料庫時，我們為求細緻，將笑話的幽默程度分成5等級。但在圖七中我們讓使用者回饋的好笑程度為3等級。這是因為我們前導試驗發現一般使用者不在乎需要將笑話程度區分為5個詳細等級，3個等級在使用上最直覺，而容易選擇回饋。

一、第一次實驗評估

在首次實驗中，我們只用到向量空間的資訊檢索技術，沒有用到Word2Vec的查詢擴展。在67位使用者做的493次笑話查詢中，有298次有回饋好笑程度。而回饋紀錄顯示有83次查詢達到破冰效果，比例為27.9%（= 83 / 298），略高於四分之一。表三呈現出語料庫中標記的好笑程度跟破冰效果關係，計算p值為0.16，其統計檢定的顯著性為不明顯。表四呈現使用者回饋好笑程度與破冰效果的關係，其p值為4.52E-33，顯示使用者認為越好笑的笑話，破冰成功的比例越高。比較表三與表四，顯示破冰效果以使用者認知的好笑話為準，這相當合理，但也顯示語料庫中好笑程度的標記與使用者的認知有差距。

在第一次實驗期間，從另外製作的回饋表單中，共收到了19位使用者的意見回饋，如圖八所示。其中有8位認為笑話量不足，7位希望「小明同學」有隨機功能，有4人

表三　第一次實驗破冰效果與語料庫中好笑程度的交叉分析

	有達到破冰效果		沒有達到破冰效果	
好笑程度	次數	百分比	次數	百分比
1	4	4.8	16	7.4
2	11	13.3	29	13.5
3	31	37.3	101	47.0
4	27	32.5	58	27.0
5	10	12.1	11	5.1
總計	83	100.0	215	100.0

表四　第一次實驗破冰效果與使用者回饋好笑程度的交叉分析

	有達到破冰效果		沒有達到破冰效果	
好笑程度	次數	百分比	次數	百分比
不好笑	6	7.2	166	77.2
有點好笑	30	36.2	41	19.1
好笑	47	56.6	8	3.7
總計	83	100.0	215	100.0

圖八　部分使用者的回饋意見

認為有些查到的笑話講出來會有些尷尬，像是黃色笑話、禁忌笑話等（語料庫的黃色笑話、禁忌笑話共有532則，占總笑話約10%的比例），而期待改進。

二、第二次實驗評估

根據第一次評估的回饋意見，發現使用者利用關鍵字查詢笑話，找不到笑話的次數為125次（第一次查詢就找不到笑話48次，加上找到過後再選「再來一個」而沒有笑話的有77次），占了25.4%（= 125 / 493）的比例。為此，我們在第一次實驗後，下載了2019年8月20日的中文維基百科文章，處理後共7.91GB純文字檔，運用Word2Vec技術（Mikolov, Chen, Corrado, & Dean, 2013）與gensim工具訓練出2,238,637個詞彙的詞嵌入向量（word embedding vectors），依此來擴展使用者查詢詞找不到笑話的情況。例如，有使用者以「雞排」找不到笑話，透過Word2Vec找出「雞排」的前10個近似詞為：「臭豆腐、小籠包、黑輪、滷肉飯、關東煮、滷味、火鍋、酒論、排骨、雞腿」，依此再查詢笑話語料庫，即可找出相關的笑話。

仔細觀察第一次實驗的對話紀錄，發現一些使用者並沒有特別想找什麼笑話（換用的查詢詞主題差別大），回饋意見中也顯示使用者有不知如何找到好笑話的困擾。因此第二次實驗除了增加上述查詢擴展的功能外，我們也新增隨機笑話提示功能，以及查詢關鍵詞的提示，並只回傳語料庫中好笑程

圖九　根據首次實驗後改進的功能

度3至5分的笑話，以改善使用者不知從何檢索，以及檢索結果品質的問題。

新功能的介面，如圖九所示。在LINE介面的最下方將選單開啟後，可點選「隨機查詢」如圖左；圖中為點選「推薦字」呈現語料庫中最常出現的詞彙，以提供查詢詞選用的參考；若還不清楚系統的使用，可點選「功能查詢」，以顯示簡短的操作說明，如圖右。

第二次實驗有132位使用者做了1,344次笑話查詢，有639次有回饋好笑程度，其中有255次達到破冰效果，比例為39.9%（= 255 / 639），約四成，亦即從首次的27.9%（= 83 / 298），提升了12%。在1,344次查詢中，隨機笑話功能的查詢共有1,038次，比例為77.2%（= 1,038 / 1,344），而以關鍵詞查詢有306次（22.8% = 306 / 1,344）。可見多數的使用者確實沒有特別想找什麼主題的笑話；但有超過二成（22.8%）仍以查詢詞想找情境相關之笑話，此比例仍不可忽視。

表五與表六呈現好笑程度與破冰效果的交叉分析結果。表五經計算其*p*值為0.01，其統計檢定顯著性為明顯。表六的*p*值則為3.16E-70，結果與第一次實驗類同。進一步分析，利用查詢功能達到破冰效果的比例為17.3%（= 44 / 255）；利用隨機功能達到破冰效果的比例為82.7%（= 211 / 255）。由此可知若只傳回好笑程度為3至5分的笑話，確實可提升破冰效果的比例。

表五　第二次實驗破冰效果與語料庫中好笑程度的交叉分析

	有達到破冰效果		沒有達到破冰效果	
好笑程度	次數	百分比	次數	百分比
1	1	0.4	6	1.5
2	6	2.3	16	4.2
3	118	46.3	202	52.6
4	118	46.3	129	33.6
5	12	4.7	31	8.1
總計	255	100.0	384	100.0

表六　第二次實驗破冰效果與使用者回饋好笑程度的交叉分析

	有達到破冰效果		沒有達到破冰效果	
好笑程度	次數	百分比	次數	百分比
不好笑	8	3.1	263	68.5
有點好笑	95	37.3	97	25.3
好笑	152	59.6	24	6.2
總計	255	100.0	384	100.0

而運用Word2Vec技術擴展使用者查詢詞後，找不到笑話的次數為51次（第一次查詢就找不到笑話22次，加上找到過後再選「再來一個」而沒有笑話的有29次），讓查不到笑話比率降為16.7%（= 51 / 306），相較於第一次實驗減少了8.7%（= 25.4% – 16.7%）。此顯示運用Word2Vec技術進行查詢擴展有其成效，可降低使用者找不到笑話的挫折感。

表三到表六顯示，不好笑的笑話，少數仍能達到破冰的效果。深入分析對話紀錄，使用者認為不好笑的笑話，在實際應用時仍有可能達到破冰效果的例子，如下：

老師：「小明，你今天是值日生，要負責打菜哦」

小明：「好的」

接著，小明走到餐桶前，舉起拳頭直直往餐桶打。

老師驚訝地說：「小明你在幹嘛？」

小明：「不是要打菜嗎？」

這則笑話在語料庫中評分為2，共有8位使用者查詢到：2位使用者覺得有點好笑及好笑，有達到破冰效果；1位使用者覺得不好笑，但有達到破冰效果；1位使用者覺得有點好笑，沒有達到破冰效果；4位使用者覺得不好笑，沒有達到破冰效果。顯示笑話好不好笑，因人而異；而能否達到破冰效果，除了真正好的笑話之外，還有其他因素。例如找到的是情色、不雅等笑話，雖然好笑但不敢講出來，而被使用者標記為沒有達到破冰效果。

進一步分析影響笑話幽默的原因，可分為笑話內容本身及個人的主觀認知。前者如笑話內容的易理解性，亦即能夠讓人理解的笑話，才能讓人發笑。後者如鄭昭明等人（2013）認為性別是影響因素之一，像是男生對情色笑話的好笑程度高於女生。另外，社會互動、群眾心理現象也有影響：當跟別人在一起的時候，笑的機會也比獨自一人的時候多30倍（Provine, 2001）。也就是說，面對一群人講出笑話，雖然也許是一個不夠好笑的笑話，但只要群眾裡其中一個人發笑，就會影響其他人也會跟著笑。這些都說明，影響笑話達到幽默的因素頗為複雜，有時即便不是夠好的笑話，也可達到破冰的效果。

捌、結論

有多種原因影響破冰機器人的成效：(1)笑話語料的品質：越多好笑的笑話，越能達到效果；(2)好笑話的可及性：越容易找出好的笑話（除查詢外，加上各種提示），越能達到效果；(3)講笑話的場景：若聽眾中有人笑點低、容易發笑，也會影響其他聽眾；但這一點難以從我們的實驗證實，只能從過去文獻的發現來推論。

本研究以資訊系統開發研究法的循環步驟，實作出基於檢索技術的對話系統，並評估其運用的成效。經過紀錄檔分析、使用者回饋的改進後，加入Word2Vec的查詢擴展、關鍵詞查詢提示，以及好笑話（好笑程度3至5分）的隨機推薦，讓使用者找不到笑話的比例從25.4%降低到16.7%（降低了

8.7%），而系統達到的破冰效果從27.9%提升到39.9%（提升了12%）。

隨機笑話的功能，讓使用者連檢索的動作都不需要，其使用率（82.7%）比查詢（17.3%）高出許多。這與目前各種資訊平台，根據使用者的使用紀錄，自動推薦而降低使用者自己檢索內容的功能類似（如YouTube、Facebook），印證了在某些應用上推薦功能已凌駕檢索功能的趨勢。後續的各類對話系統，建議應運用類似的推薦功能，以符合現今使用者的習慣與期待。

綜合而言，本研究的重要性，不僅編製了正體中文幽默語料庫，也建置中文幽默對話應用系統，並把這些語料與系統開源在公開網站：https://github.com/SamTseng/icebreaker供後續便利的研究、驗證與利用。此外，在研究過程與結果中，提供了實證經驗與意涵：亦即笑話語料的豐富程度與品質（蒐集更多笑話且自動標記好笑程度），以及相對於被動檢索的自動推薦功能，是提升此類服務成效的重要研發工作。

後續的研究，期望能以生成式方法產生幽默機制，平順自然的整合到圖書館虛擬參考服務或常見問答中，達到消解使用者抱怨程度、贏得信任等功用，以進一步提升圖書館服務的品質。

註釋

註一：此例改自https://www.youtube.com/watch?v=K38wVcdNuFc&t=10

註二：此維度亦可設為50維、100維、500維或是其他數量的維度，視應用而定。

誌謝

本研究感謝科技部研究計畫補助，計畫編號：MOST 107-2221-E-003-014-MY2、MOST 109-2634-F-002-023。

參考文獻 References

任璐、楊亮、徐琳宏、樊小超、刁宇峰、林鴻飛（2018）。中文笑話語料庫的構建與應用。*中文信息學報，32*(7)，20-29。【Ren, Lu, Yang, Liang, Xu, Linhong, Fan, Xiaochao, Diao, Yufeng, & Lin, Honfei (2018). Construction and application of Chinese joke corpus. *Journal of Chinese Information Processing, 32*(7), 20-29. (in Chinese)】

李廣偉、戈玲玲、劉朝暉（2016）。言語幽默漢英平行歷時語料庫及其檢索系統的構建與應用。*外語電化教學，172*，60-65。【Li, Guang-Wei, Ge, Ling-Ling, & Liu, Zhao-Hui (2016). The construction and application of diachronic parallel corpus of Chinese-English verbal humor and its retrieval system. *Technology Enhanced Foreign Language Education, 172,* 60-65. (in Chinese)】

李璠（2017）。基于自建語料庫對環境幽默語篇的多維度分析。*環球市場信息導報，21*，102-106。【Li, Fan (2017). [Ji yu zi jian yu liao ku dui huan jing you

mo yu pian de duo wei du fen xi]. *Global Market Information Guide, 21*, 102-106. (in Chinese)】

劉鋒、張京魚（2015）。基於多媒體語料庫的小學生幽默話語會話分析。*山東師範大學外國語學院學報：基礎英語教育，17*(2)，15-21。【Liu, Feng, & Zhang, Jin-Yu (2015). [Ji yu duo mei ti yu liao ku de xiao xue sheng you mo hua yu hui hua fen xi]. *Journal of Basic English Education, 17*(2), 15-21. (in Chinese)】

鄭昭明、陳學志、詹雨臻、蘇雅靜、曾千芝（2013）。臺灣地區華人情緒與相關心理生理資料庫—中文笑話評定常模。*中華心理學刊，55*(4)，555-569。doi: 10.6129/CJP.20121026【Cheng, Chao-Ming, Chen, Hseuh-Chih, Chan, Yu-Chen, Su, Ya-Ching, & Tseng, Chien-Chih (2013). Taiwan corpora of Chinese emotions and relevant psychophysiological data-normative data for Chinese jokes. *Chinese Journal of Psychology, 55*(4), 555-569. doi: 10.6129/CJP.20121026 (in Chinese)】

Augello, A., Saccone, G., Gaglio, S., & Pilato, G. (2008). Humorist bot: Bringing computational humour in a chat-bot system. In F. Xhafa & L. Barolli (Eds.), *The second International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 703-708). Los Alamitos, CA: IEEE Computer Society. doi: 10.1109/CISIS.2008.117

Bellegarda, J. R. (2014). Spoken language understanding for natural interaction: The Siri experience. In J. Mariani, S. Rosset, M. Garnier-Rizet, & L. Devillers (Eds.), *Natural interaction with robots, knowbots and smartphones* (pp. 3-14). New York, NY: Springer. doi: 10.1007/978-1-4614-8280-2_1

Bergen, B., & Coulson, S. (2006). Frame-shifting humor in simulation-based language understanding. *IEEE Intelligent Systems, 21*(2), 59-62.

Binsted, K. (1995, August). *Using humour to make natural language interfaces more friendly*. Paper presented at the AI, ALife and Entertainment Workshop, Montreal, Canada.

Binsted, K., & Ritchie, G. (1997). Computational rules for generating punning riddles. *Humor: International Journal of Humor Research, 10*(1), 25-76. doi: 10.1515/humr.1997.10.1.25

Blinov, V., Bolotova-Baranova, V., & Braslavski, P. (2019). Large dataset and language model fun-tuning for humor recognition. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 4027-4032). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.18653/v1/P19-1394

Blinov, V., Mishchenko, K., Bolotova, V., & Braslavski, P. (2017). A pinch of humor for short-text conversation: An information retrieval approach. In G. Jones et al. (Eds.), *Lecture Notes in Computer*

Science: Vol. 10456. Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2017. (pp. 3-15). Cham, Switzerland: Springer. doi: 10.1007/978-3-319-65813-1_1

Bryant, J., & Zillmann, D. (1989). Chapter 2: Using humor to promote learning in the classroom. *Journal of Children in Contemporary Society, 20*(1/2), 49-78. doi:10.1300/J274v20n01_05

Cambridge Advanced Learner's Dictionary. (2019). *Icebreaker.* Retrieved from https://dictionary.cambridge.org/zht/dictionary/english-chinese-traditional/icebreaker

Chen, L., & Lee, C. M. (2017). *Predicting audience's laughter using convolutional neural network*. Retrieved from https://arxiv.org/abs/1702.02584

Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends, 61*(3), 636-674. doi: 10.1353/lib.2013.0005

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407. doi: 10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Du, S., Wan, X., & Ye, Y. (2017). *Towards automatic generation of entertaining dialogues in Chinese crosstalks*. Retrieved from https://arxiv.org/abs/1711.00294

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In J. R. Firth (Ed.), *Studies in linguistic analysis* (pp. 1-32). Oxford, England: Blackwell.

Gang, W. Y., Bo, S., Chen, S. M., Yi, Z. C., & Zi, M. P. (2014). Chinese intelligent chat robot based on the AIML language. In J. Lawry (Chair), *2014 Sixth International Conference on Intelligent Human-machine Systems and Cybernetics* (Vol. 1, pp. 367-370). Los Alamitos, CA: IEEE Computer Society. doi: 10.1109/IHMSC.2014.96

Ivanov N., Khalman M., Smetanin N., Rodichev A., & Fedorenko D. (2019). CakeChat: Emotional generative dialog system [Computer program]. Retrieved from https://github.com/lukalabs/cakechat

Ji, Z., Lu, Z., & Li, H. (2014). *An information retrieval approach to short text conversation*. Retrieved from https://arxiv.org/abs/1408.6988

Johnson, K. (2017). *Facebook Messenger hits 100,000 bots*. Retrieved from https://venturebeat.com/2017/04/18/facebook-messenger-hits-100000-bots/

Kane, D. A. (2015). *ANTswers: An interactive library FAQ*. Retrieved from https://escholarship.org/uc/item/4bs6s3hs

Le, Q. V., & Mikolov, T. (2014). *Distributed representations of sentences and documents*. Retrieved from https://arxiv.org/abs/1405.4053

Li, J., Monroe, W., Ritter, A., Galley, M., Gao, J., & Jurafsky, D. (2016). *Deep reinforcement learning for dialogue generation*. Retrieved from https://arxiv.org/abs/1606.01541

Mcghee, P. E., & Frank, M. (2014). *Humor and children's development: A guide to practical applications*. Oxford, England: Routledge.

Mihalcea, R., & Strapparava, C. (2006a). Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence, 22*(2), 126-142. doi: 10.1111/j.1467-8640.2006.00278.x

Mihalcea, R., & Strapparava, C. (2006b). Technologies that make you smile: Adding humor to text-based applications. *IEEE Intelligent Systems, 21*(5), 33-39. doi:10.1109/MIS.2006.104

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. Retrieved from https://arxiv.org/abs/1301.3781

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Proceedings of the 26th International Conference on Neural Information Processing Systems* (Vol. 2, pp. 3111-3119). Red Hook, NY: Curran.

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human-computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human-Computer Interaction, 14*(4), 395-435. doi: 10.1207/S15327051HCI1404_2

Moudgil, A. (2017). *Short-jokes-dataset* [Python scripts]. Retrieved from https://github.com/amoudgl/short-jokes-dataset

Newyear, D., & McNeal, M. (2014). *Extending library services with AI conversational agents*. Retrieved from https://www.slideserve.com/garth/ai-conversational-agents

Nijholt, A. (2006). Embodied conversational agents: "A little humor too." *IEEE Intelligent Systems, 21*(2), 62-64.

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems, 7*(3), 89-106. doi: 10.1080/07421222.1990.11517898

Özbal, G., & Strapparava, C. (2012). Computational humour for creative naming. In A. Nijholt (Ed.), *Proceedings 3rd International Workshop on Computational Humor* (pp. 15-19). Amsterdam, Netherlands: Centre for Telematics and Information Technology.

Potash, P., Romanov, A., & Rumshisky, A. (2017). SemEval-2017 task 6: #HashtagWars: Learning a sense of humor. In S. Bethard, M. Carpuat, M. Apidianaki, S. M. Mohammad, D. Cer,

& D. Jurgens (Eds.), *Proceedings of the 11th International Workshop on Semantic Evaluation* (pp. 49-57). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.18653/v1/S17-2004

Provine, R. R. (2001). *Laughter: A scientific investigation*. London, England: Penguin Books.

Qiu, M., Li, F.-L., Wang, S., Gao, X., Chen, Y., Zhao, W., ... Chu, W. (2017). AliMe Chat: A sequence to sequence and rerank based chatbot engine. In R. Barzilay & M.-Y. Kan (Eds.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers, pp. 498-503). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.18653/v1/P17-2079

Ritchie, G., Manurung, R., Pain, H., Waller, A., & O'Mara, D. (2006). The STANDUP interactive riddle-builder. *IEEE Intelligent Systems, 21*(2), 67-69.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Sjöbergh, J., & Araki, K. (2009). A very modular humor enabled chat-bot for Japanese. In H. Kameda (Chair), *Proceedings of Conference of the Pacific Association for Computational Linguistics (PACLING 2009)* (pp. 135-140). Sapporo, Japan: Hokkaido University.

Stock, O., & Strapparava, C. (2003). Getting serious about the development of computational humor. In A. Cohn (Ed.), *Proceedings of the 18th International Joint Conference on Artificial Intelligence* (pp. 59-64). San Franciso, CA: Morgan Kaufmann.

Stock, O., & Strapparava, C. (2006). Automatic production of humorous expressions for catching the attention and remembering. *IEEE Intelligent Systems, 21*(2), 64-67.

Tseng, Y.-H., & Teahan, W. J. (2004). Verifying a Chinese collection for text categorization. In M. Sanderson (Chair), *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 556-557). New York, NY: Association for Computing Machinery. doi: 10.1145/1008992.1009118

Wallace, R. (2003). *The elements of AIML style*. Retrieved from https://files.ifi.uzh.ch/cl/hess/classes/seminare/chatbots/style.pdf

Wen, T.-H., Gasic, M., Kim, D., Mrksic, N., Su, P.-H., Vandyke, D., & Young, S. (2015). *Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking*. Retrieved from https://arxiv.org/abs/1508.01755

Yang, D., Lavie, A., Dyer, C., & Hovy, E. (2015). Humor recognition and humor anchor extraction. In L. Màrquez, C. Callison-Burch, & J. Su (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 2367-2376). Lisbon,

Portugal: Association for Computational Linguistics. doi: 10.18653/v1/D15-1284

Zhang, R., & Liu, N. (2014). Recognizing humor on Twitter. In J. Li & X. S. Wang (Chairs), *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp. 889-898). New York, NY: Association for Computing Machinery. doi: 10.1145/2661829.2661997

（投稿日期Received: 2020/5/19　接受日期Accepted: 2020/7/23）