基於標準答案的問句產生方法

陳嘉銘

國立雲林科技大學資訊管理所 碩士生

kui15230@gmail.com

陳重臣

國立雲林科技大學資訊管理所 教授

jcchen@yuntech.edu.tw

黃昭義

國立雲林科技大學前瞻學士學位學程 助理教授

joyh@yuntech.edu.tw

摘要

聊天機器人大部分應用於社群媒體、網頁導覽、電子商務等應用，主要提供客戶即時獲取資訊，而目前大多數的聊天機器人系統都是基於搜尋關鍵字，再從資料庫中尋找最匹配的應答句回饋給客戶。大部分資料庫中的問句都是採用人工產生的方式，且基於各個領域的不同知識，需要專業的人員去衡量產生的問句品質，如此將耗費大量時間及人力成本。本研究將提出一個從標準答案產生出問句的方法，並透過 Word2vec 做相似度比對，將問句與問句之間的關係連結起來，協助人工產生問句語料庫。

關鍵詞：聊天機器人、問答系統、Word2vec、網路爬蟲

Abstract

Most of the chat bots are used in social media, web navigation, e-commerce and other applications, mainly to provide customers with instant access to information, and most of the current chatbot systems are based on search keywords, and then find the best match from the database. The response sentence is fed back to the customer. Most of the questions in the database are artificially generated, and based on different knowledge in various fields, professional personnel are required to measure the quality of the questions generated, which will consume a lot of time and labor costs. This study will present a method for generating questions from standard answers, and through Word2vec to do similarity comparisons, link the relationship between questions and questions, and assist in artificially generating query corpus.

Keyword: Chatbot, Question and answer system, Word2vec, Web crawler