2017/01/26

專題內容說明

專題研究 李琳山

- 研究題目：語音訊號處理專題 (Special Projects in Speech Signal Processing)
- 對 象：電機系、資訊系大學部同學大三或大四
- 背景說明：

1. 無線通訊（Wireless Communications）技術及網路環境下日新月異並具多元功能的智慧型手機、可穿戴的 (wearable)如眼鏡 (Glasses)、手錶 (iwatch)及其他未來各種新型的輕便隨身電子設備、車用電子裝置、資訊家電等新型電子設備正在為硬體世界開創全新的空間，也使網路世界的終端設備徹底多元化。這些多元的終端設備功能日益豐富，也正在迅速取代今日的個人電腦，這也就是所謂的後PC（Post-PC）時代的來臨。在輕薄短小的硬體及豐富的軟體應用環境下，由於人的手指不會縮小，原有的鍵盤、滑鼠等個人電腦上網介面將不再方便，語音很顯然將成為最方便自然的網路介面之一。Apple Siri、Google Now、微軟的Cortana等均為現階段成功的全球性產品。

2. 另一方面，在多媒體（Multimedia）技術一日千里的發展下，具備影音、視訊等多媒體功能的手持電子設備已大量出現，而網路上的數位內容（Digital Content）尤其常以多媒體形式呈現如新聞、演講、課程、影片等，它們未必具有文字檔案，卻都帶著語音訊息；它們不易瀏覽，但所帶的語音訊息最適合拿來作為瀏覽多媒體數位內容的基礎。

3. 於是越來越多的使用者必然會透過手持設備用語音指令上網，而網路上的數位內容也多以多媒體及語音形式呈現，適於藉其語音訊息來搜尋。今日上網動作仍主要是用文字指令搜尋文字檔案，其中的文字角色會有越來越多部分由語音取代，語音訊號處理技術也就自然成為新一代軟硬體技術的關鍵部份。

4. 本專題讓大學部同學進行電腦語音訊號處理之初步研究，由電信所、資工所合設的語音實驗室提供部份基礎程式、軟體工具及語音語言資料庫，參與的同學要自行撰寫若干程式，發展基本的語音處理技術，並自行用自己隨時錄製的語音訊號進行實驗。

- 進行方式:

1. 由瞭解基本知識研讀基礎課本或論文開始，在一個學期中，在第一階段先用語音實驗室所提供的資料庫及軟體工具自行建構一套大字彙語音辨識系統並測試其功能，第二階段再在一系列較具挑戰性的研究主題中自行選擇一個主題進行深入研究。

2. 以一個學期為單位，但當然可以延續到第二學期進入更深入的課題。

3. 以每兩人一組最為合適，可以相互討論，共同工作，當然一人一組也是可以的。第一階段完成大字彙語音辨識系統；第二階段選定一個題目，進行研究。

4. 雖然未來主要的應用必將在硬體（例如晶片）上操作，本專題所有工作均為數學模型及軟體程式，使用PC 或工作站。

5. 第一階段的大字彙語音辨識系統及第二階段的每一個題目都會有語音實驗室的研究生學長擔任助教，提供基礎課本或論文、軟體工具、語音資料庫等並引導同學進入主題，但核心部分的研究方向有相當空間由選修同學自行負責，自由發揮，研究生助教僅擔任諮詢角色。學期中每週與老師定期會面討論一次，每組同學隔週需上台報告研究進度一次。

6. 已選修過「數位語音處理概論」課程的同學可以作研究的方向很多，可不受上述限制，在完成前半學期的大字彙語音辨識系統之後，可直接自行選定其他任何題目開始研究，也可以1人自成一組直接進行。

● 配套課程:

『數位語音處理概論』，本學期同時開授，電機系、資訊系大三以上程度均可聽得進去並選修。原則上修本專題研究的同學都應選修這門課程並研讀相關參考文獻，同時動手作專題研究的實驗。這門課可以加強學理基礎，學到豐富的作研究的方法，並讓學生對此一新興領域有完整的瞭解，而本專題研究則提供實際研究的經驗。這門課程與本專題研究有配套設計，兩者相輔相成。

● 研究內容（一）- 第一階段建構大字彙語音辨識系統:

這一部份的目標，是透過建立一個基本型的大字彙語音辨識系統，以電視新聞或課程錄音為辨識對象，讓同學對語音辨識有具體的了解及完整的經驗，並且以此作為進一步研究各項進階技術的基礎。從聲學特徵抽取，聲學模型及語言模型的訓練及測試，整合搜尋及辨識功能，到系統效能的評估等，同學將學習使用Kaldi及SRILM等軟體工具，實際建立一套可以辨識電視新聞或課程錄音的大字彙語音辨識系統。上述每個步驟所牽涉到的相關原理會在配套課程中學到，而程式使用方法與輸入資料格式，以及錯誤訊息等，都可以在助教提供的參考資料與範例中獲得解答。

● 研究內容（二）- 第二階段可能的研究題目:

每一學期的研究題目會有調整。以下所列為上個年度(104年9月-105年6月)第二階段可以選擇研究題目中的若干例子供參考(每2人一組可選一個題目作第二階段的研究主題)。

1. 結構化深層類神經網路之語音處理 (Structured DNN in Speech Recognition)

結構化學習(structured learning)讓機器學習結構化的問題。舉例來說一個句子或者一段聲音就都有它的結構，讓機器學習這個整個結構，而不是切割成一個個小單位分開學習。我們可以把結構化的問題(譬如將整段語音信號辨識成整段文字)用Structured DNN來解決。本專題將會藉由kaldi toolkit來實作，並且使用TIMIT語料庫與全世界競爭。

2. 以語音技術提升線上學習效能

今日網路上有大量公開線上教學課程(Massive open online courses, MOOCs)；但因為這些教學影片的內容可能有深有淺，使用者很有可能不知該從哪些相關影片開始看起。這個題目就是希望讓機器能自動分析、組織課程內容，甚至能將大量教學影片自動規劃成學習地圖。這裡面有非常多研究主題可以思考，例如：假設有兩段來自兩門不同課程的影片，系統如何能自動知道，其中一個是比較基礎的知識，學習者應該先看完後才能理解另一段影片的內容。本研究為台大語音實驗室和麻省理工學院(MIT) 電腦科學暨人工智慧實驗室(CSAIL)的合作計畫。

3. 華語學習對話遊戲 (Dialogue Game for Mandarin Chinese learning)

語音對話系統是語音處理技術的具體呈現，常用馬可夫決策程序(Markov Decision Process, MDP)來建立使用者和機器的互動機制，或者說是利用數學統計及機器學習來得到一套好的對話策略。本專題我們將介紹對話系統，MDP之理論及學習，一直到將這套模型應用至華語學習的情境上，將撰寫MDP程式，助教會從基礎開始教起。

4. 含語音標註的相片搜尋系統 (Retrieval of Photos with Speech Annotations)

我們的手機裡可能存了相當多的照片，在未經特別整理的狀況之下，每次想要找出特定相片總是得花上一些心力。因此，在拍攝相片時，我們可以讓使用者說一句話，假設只有少部分相片(30%)是有聲音的。我們的任務便是利用這些有限的語音以及影像資訊，來找出所有和使用者查詢指令(user query)有關的相片(無論是否有聲音)。我們會嘗試利用不同的技術讓語音及影像的資訊充分的融合與互動，來完成這個任務。

5. 語音文件摘要 (Spoken Document Summarization)

與純文字資料相比，語音文件不易呈現在螢幕上不易瀏覽，故自動選擇少數語句來構成摘要(summary)，盡量涵蓋原本語音文件中的資訊是重要的研究方向。本專題將會利用「數位語音處理概論」、「信號與系統」課程影音以及「公視新聞」等資料庫作為素材，引導同學進行自動語音文件摘要之研究。

● 有興趣參加者：

1. 原則上先將姓名學號及e-mail 地址寄至: lslee@cc.ee.ntu.edu.tw 先行登記。如已組成兩人一組，則將兩人的資料一同登記。可以容納的組數是有限的，有興趣請早登記。
2. 本學期每週定期Meeting暫訂每星期四下午5:30，地點另行公布，第一次是開學第一週2月23日。
3. 相關訊息會公告在網頁上。
4. 第一階段基礎實驗完成的同學可以選定後半學期深入研究的題目，原則上希望每一組作不同的題目。
5. 尚未修過「數位語音處理概論」課程者，本學期請選修「數位語音處理概論」課程。