自然語言處理工程師應用實務班

■ 課程簡介

讓電腦開始聽懂人話的自然語言處理（Natural Language Processing，NLP）是近幾年成長最迅速的 AI 應用，卻也是最具挑戰性的研究領域之一。自然語言處理技術與應用之目標在於幫助機器處理、理解及生成人類的語言。NLP 的機制涉及兩個流程：自然語言理解及自然語言生成。自然語言認知和理解是讓電腦把輸入的語言變成有意思的符號和關係，然後根據目的再處理。自然語言生成系統則是把計算機資料轉化為自然語言。

根據 P&S Intelligence 最新的預測顯示，自然語言處理市場的產值將在 2024 年達到 229 億美元的規模，2019 到 2024 的年複合成長率將達到 19.7%。在過去一年中，人工智慧技術發展進程突破所有人的想像，進入了全新的里程碑：在自然語言處理方面，Google 開源了比 BERT 更優秀的 XLNet 模型，在眾多 NLP 任務的任務中達到 SOTA 的精度，Facebook 則在去年末發表了基於 PyTorch 架構的 PyText，利用其快速迭代的特點來提高自然語言處理的效率與可擴展性，用以加速 NLP 專案的產品化。自然語言處理可以作為雲端服務或者本地服務，並應用於資訊抽取、問題回答、機器翻譯等領域，並且適合為許多垂直行業製作報告，例如汽車、零售和消費品、高科技電子、政府公共管理、銀行業務、金融服務以及保險、醫療保健及生命科學、研究與教育、媒體與娛樂等等。自然語言處理，常見的應用如下圖所示:

自然語言處理，常見的應用。
資料來源｜李宏毅演講 圖說重製｜林婷嫻、張語辰

自然語言處理工程師應用實務班全系列課程，係根據自然語言處理之學習地圖規劃，包含三個模組，學員可根據自己的需求，由淺入深逐步分階段地學習，將帶領學員從深度學習框架及演算法應用到自然語言處理，進而實現具有語言處理功能的人工智慧應用系統。

模組一: 自然語言處理相關之程式設計及深度學習框架
單元 1.1 Python 程式設計及常用函式庫
單元 1.2 深度學習框架: Keras、PyTorch 等

模組二: 自然語言處理工具與資源
單元 2.1 自然語言處理基礎
單元 2.2 詞向量

模組三: 自然語言處理應用剖析
單元 3.1 文本分類
單元 3.2 社群媒體分析

■ 課程目標 (需求)

- 掌握自然語言處理之基礎核心技術，包括程式設計及深度學習框架等。
- 掌握自然語言處理工具與資源，包括：斷詞、詞性標注、特徵選取、語言模型、詞向量等。
- 掌握自然語言處理典範應用之技術開發及其眉角，能夠有效地開發人機間信息交流的產品或服務。
- 學習如何將深度學習的演算法應用到自然語言處理，從而實現具有語言處理功能的人工智慧應用系統。課程設計將從自然語言的資料表示方法開始介紹，並透過此方法與深度學習的CNN、RNN、LSTM等神經網路模型整合成自然語言處理的演算架構，以及學習最終如何使用深度學習框架實現此一可應用於現代人工智慧系統的自然語言處理方法。

■ 課程特色 (賣點)

- 完整的自然語言處理學習地圖：聚焦自然語言處理全系列課程，量身設計課程，學員可依需求，由淺入深逐步分階段地學習，將帶領學員從深度學習框架及演算法應用到自然語言處理，進而實現具有語言處理功能的人工智慧應用系統。
- 堅強師資陣容：授課老師具備完整知識及豐富的業界經驗。課程中將以著名的案例來說明深度學習原理及其應用於自然語言處理的方法，讓學員以最快的速度奠定深度學習與自然語言處理的設計基礎。
- 聚焦自然語言處理，量身設計課程，快速的重點講解：課程中將結合講師講授、經驗分享、實務案例分析、工具軟體應用示範，以增進學習的成效。同時，提供豐富的補充教材，包括設計時參考使用的程式源碼及實驗資料取得方式，讓學員能夠盡快上手，開始設計屬於自己的專案計畫，加入自然語言處理的設計行列。

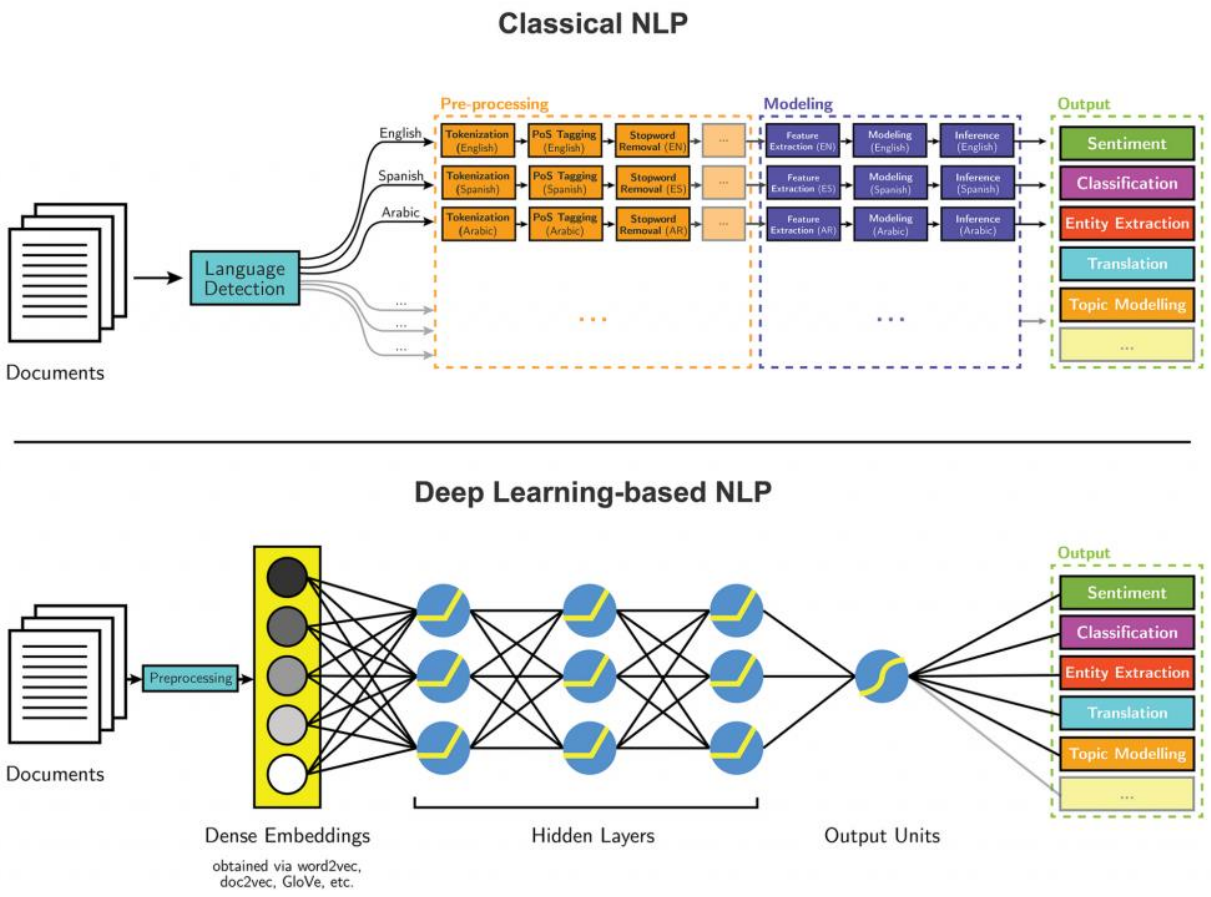
■ 適合對象

- 適合具有基礎程式邏輯的學員
- 從事自然語言處理系統規劃、建置或應用開發者
- 從事自然語言處理邊緣運算裝置、系統應用服務、系統整合等之研發人員及主管。
- 從事自然語言處理創新應用之市場研究及計畫管理開發經理人。
- 對自然語言處理智慧應用實務架構及技術，及應用服務感興趣者

■ 先備知識

- 掌握入門級代數知識，包括變量和係數、線性方程式、函數圖和直方圖。
- 基礎程式邏輯

傳統 NLP 和深度學習 NLP 的區別:



■ 課程大綱 舉辦日期： 七天共七單元，每天 6 小時，全系列共計 42 小時。學員可依需求任選獨立單元。

模組一: 自然語言處理相關之程式設計及深度學習框架

單元 1.1 Python 程式設計及常用函式庫

時間	課程單元	課程大綱	講師
9:00~ 12:00	Python 基礎程式設計	• 整合開發環境：IPython notebook • 資料類型：boolean, numbers, strings • 資料結構：list, tuple, dictionary, set • 程式結構：if else, for, while, function	中央大學電機系 李龍豪助理教授
13:00~ 16:00	Python 資料分析套件	• 資料科學的基石：numpy • 資料處理：pandas • 資料視覺化：matplotlib	中央大學電機系 李龍豪助理教授

單元 1.2 深度學習框架: Keras

時間	課程單元	課程大綱	講師
9:00~ 12:00	Keras 神經網路實作	• 卷積神經網路（CNN） • 遞歸神經網路（RNN） • 長短期記憶模型（LSTM） • 生成對抗網路（GAN）	中央大學電機系 李龍豪助理教授
13:00~ 16:00	Keras 應用範例實作	• 二元分類：IMDB 電影評分類 • 多類別分類：路透社新聞專欄分類 • 迴歸分析：波士頓房價預測 • 影像生成：DeepDream 風格轉換	中央大學電機系 李龍豪助理教授

單元 1.3 深度學習框架: PyTorch

時間	課程單元	課程大綱	講師
9:00~ 12:00	**PyTorch** 深度學習框架	● 深度學習簡介 ● 深度學習框架環境介紹與設定 ● 深度神經網路訓練原則與運作過程 ● 基本深度神經網路實作	東吳大學 巨量資料管理學院 吳政隆助理教授
13:00~ 16:00	**PyTorch** 深度學習模型 實作	● 遞歸神經網絡(RNN、GRU、LSTM) ● 卷積神經網路(CNN) ● 注意力機制(Attention Mechanism) ● sequence-to-sequence 模型 ● 文本分類應用實作 ● 文本生成應用實作	東吳大學 巨量資料管理學院 吳政隆助理教授

模組二: 自然語言處理工具與資源

單元 2.1 自然語言處理基礎

時間	課程單元	課程大綱	講師
9:00~ 12:00	自然語言處理基礎	● 斷詞 ● 詞性標注 ● 特徵選取 ● 語言模型 ● 命名實體識別 ● 句子剖析 ● 相關應用	元智大學資管系 禹良治教授
13:00~ 16:00	語言資源與工具	● 語言資源 (如：語料庫、知識本體) ● 處理工具 (如： Python NLTK)	元智大學資管系 禹良治教授

單元 2.2 詞向量

時間	課程單元	課程大綱	講師
9:00~ 12:00	基礎詞向量模型介紹與工具實作	● 詞向量原理 ● 詞向量工具環境介紹與設定 ● 知名詞向量模型介紹(Word2vec、Doc2vec、FastText)	東吳大學 巨量資料管理學院 吳政隆助理教授
13:00~ 16:00	進階詞向量模型介紹	● ELMO 模型 ● BERT 模型 ● 最新 ALBERT 模型	東吳大學 巨量資料管理學院 吳政隆助理教授

★主辦單位保留調整課程內容、行程與講師之權利

模組三: 自然語言處理應用剖析

單元 3.1 文本分類

時間	課程單元	課程大綱	講師
9:00~ 12:00	機器學習概念	● 監督式與非監督式學習 ● 分類、聚類與迴歸 ● 實驗設計：k-fold cross validation ● 結果解讀：混淆矩陣與偽陰性/偽陽性 ● 案例分享	元智大學資管系禹良治教授
13:00~ 16:00	主題實作	● 文本分類實作 ● 成果發表	元智大學資管系禹良治教授

單元 3.2 社群媒體分析

時間	課程單元	課程大綱	講師
9:00~ 12:00	社群媒體分析	● 情感分析 ● 資料視覺化 ● 案例分享	元智大學資管系禹良治教授
13:00~ 16:00	主題實作	● 情感分析實作 ● 視覺化實作(社群網路)	元智大學資管系禹良治教授

★主辦單位保留調整課程內容、行程與講師之權利

■ 講師簡介

禹良治 (元智大學 資訊管理學系 教授)

【學歷】
- 國立成功大學資訊工程博士

【經歷】
- 元智大學圖書資訊服務處 副資訊長
- DOCOMO Innovations, Palo Alto, CA, USA 訪問學者

【專業領域】:
- 自然語言處理、情感分析、學習科技

李龍豪 (中央大學 電機工程學系 助理教授)

【學歷】
- 國立臺灣大學資訊工程博士

【經歷】
- 科技部人工智慧技術暨全幅健康照護聯合研究中心 博士後研究員
- 國立臺灣師範大學 圖書資訊學研究所 博士後研究員&兼任助理教授
- 國立臺灣師範大學 資訊中心 博士後研究員

【專業領域】:
- 自然語言處理、人工智慧、網路資訊檢索、生醫資訊學

吳政隆 (東吳大學 巨量資料管理學院 助理教授)

【學歷】
- 元智大學資訊管理學系博士

【經歷】
- 東吳大學巨量資料管理學院助理教授
- 中國文化大資訊管理學系助理教授
- 中央研究院資訊科學研究所博士後研究

【專業領域】:
- **Text Mining**
- **Deep Learning**
- **Machine Learning**
- **Sentiment Computing**
- **Stock Market Prediction**
- **Natural Language Processing**

- 主辦單位：財團法人工業技術研究院
- 舉辦地點：工研院光復院區一館 教室 (新竹市光復路二段 321 號一館 教室)

  實際上課地點，請依上課通知為準!
- 舉辦日期： 七天共七單元，每天 6 小時，全系列共計 42 小時。學員可依需求任選獨立單元。
  - 12/07 ： 單元 1.1 Python 程式設計及常用函式庫；講師: 李龍豪助理教授；
  - 12/08 ： 單元 1.2 深度學習框架: Keras；講師: 李龍豪助理教授；
  - 12/09 ： 單元 1.3 深度學習框架: PyTorch；講師: 吳政隆助理教授；
  - 12/15 ： 單元 2.1 自然語言處理基礎；講師: 禹良治教授；
  - 12/16 ： 單元 2.2 詞向量；講師: 吳政隆助理教授；
  - 12/22 ： 單元 3.1 文本分類；講師: 禹良治教授；
  - 12/23 ： 單元 3.2 社群媒體分析；講師: 禹良治教授

- 報名截止日期：該單元上課前一周。
- 課程費用 (含稅、午餐、講義)：折扣後若有個位數則個位數歸零
  - 一般生，任一單元: 每人$4,000，早鳥優惠價減價$300 (3 週前)：每人$3,700 元；
  - 團報(同公司 2 人含以上)優惠價，團報每人減價$300，即任一單元: 每人$3,700 元，早鳥團報價(3 週前)：每人$3,400 元；工研人不適用團報價。
  - 任選二單元含以上，上述費用打九折。任選四單元含以上，上述費用打八五折。
  - 全系列共七單元，原價每人$28,000，享八折優惠價: 每人$22,400，早鳥(3 週前)優惠價: $22,100，早鳥團報(同公司 2 人含以上)每人$21,800。
  - 報名兩單元以上者，請先直接勾選課程並註記於備註欄，助理將於後台修改費用。
- 報名方式：
  - 工研院學習服務網，線上報名：https://college.itri.org.tw/course/all-events/5FFB4FE9-6A23-4504-8DAD-CF9A0701DA3F.html 或

    https://reurl.cc/arzxpY
  - 或請聯絡洽詢黃小姐 03-5732034 或 email 至 itri535579@itri.org.tw

■ 注意事項：

1. 請自帶筆電，實作部分建議筆電配備:

   硬體與作業系統
   - CPU: Intel i5 以上
   - 記憶體: 8 GB 以上
   - 硬碟: 剩餘 50 GB 以上
   - 作業系統: Windows 10、Ubuntu 16 以上 、Mac OSX 10.10 以上

   軟體與其他
   - Google Chrome 瀏覽器
   - Google Colab（需要有 Google 帳號）
   - Python 3.7 或 Anaconda3

   若安裝上有問題，請 email 或來電詢問

2. 為確保您的上課權益，報名後若未收到任何回覆，敬請來電洽詢方完成報名。

3. 因課前教材、講義及餐點之準備及需為您進行退款相關事宜，若您不克前來，請於開課三日前告知，以利行政作業進行並共同愛護資源。

4. 若原報名者因故不克參加，但欲更換他人參加，敬請於開課前三日通知。