



講師：陳沛甫

主題：

Machine learning for risk assessment in minimally invasive cardiac surgery

摘要：

由於傷口較小以及恢復較快速，minimally invasive cardiac surgery (MICS) 手術愈來愈常使用於治療心臟冠狀動脈和瓣膜疾病。The European System for Cardiac Operative Risk Evaluation (EuroSCORE) 可以依據病患因子、心臟因子和手術因子預測心臟手術的風險。現行使用的 EuroSCORE II 是在 2012 年發表的運用邏輯式回歸的風險預測模型。他能夠在心臟手術有很好的風險預測鑑別力，但是在 MICS 手術卻會高估風險。因此更適合 MICS 的風險模型還有待研究。

機器學習常用於建立風險預測模型，這些方法包括 random forest、gradient boost machine、support vector machine 和 deep neural network。運用非線性模型、決策樹、以及類神經網路等方式，這些模型在臨床風險預測超越傳統方法。另外，對於多類別的輸入變項，可運用 label embeddings 的方式將類別轉化成多維度的向量，以得到更適合預測模型的特徵值。

自然語言處理是讓電腦了解文字資料的技術。因為類神經網路的發展，新的自然語言處理技術有愈來愈高的準確度和更快的處理效能，例如 Word2Vec, Global Vectors (GloVe), Embeddings from Language Models (ELMo), and Bidirectional Encoder Representations from Transformers (BERT) 等模型。



其中，相對於 Word2Vec 和 GloVe 是不考慮上下文語意的，BERT 和 ELMo 是可考慮上下文語意關聯的語意式語言模型 (contextualized language model)。BERT 所運用的語意式文字 embeddings 方式能夠用多維度的向量，以維度間的距離保存文字間的語意關係。加上 multi-head attention，使得 BERT 能夠關注整句話的文字間的關係。在使用相關文章預訓練後，BERT 可以有效從臨床文字記錄中取得有意義的 embeddings。

自然語言處理能夠運用臨床文字預測病患風險。術前診斷的文字能夠判斷病患病情的嚴重度，手術方式的文字也能夠判斷手術的複雜度。運用自然語言處理加上深度學習，我們訓練模型學習並用結構化和非結構化資料，預測 MICS 後的死亡風險。預期加入自然語言處理後，模型運用術前診斷和手術方式能夠增強預測的準確度。