

國立高雄大學統計學研究所

113 學年度書報討論題目暨摘要登記表

主動學習中初始點選擇方法之於網絡資料的應用

王淵棠

摘要

在標記有限之網絡數據的情境下，Bilgic (2010)的研究提出了一種結合半監督學習與降維技術的主動式學習建模方法，以有效提升網絡數據的分類性能。然而主動式學習中，初始樣本的選擇會在迭代過程中影響模型的預測性能。因此在 Bilgic (2010)的主動式學習架構上，我們於初始樣本選取考量上納入節點中心度 (Node Centrality) 概念，包含度中心性、接近中心性、page rank 等指標，基於不同權重配置整合節點中心度之指標，以建構不同初始樣本選擇策略，並將其應用於網絡數據以探究其對於模型收斂及分類準確率提升之影響性。我們亦將修改之演算法應用於其他網絡數據資料，並與原始主動式學習演算法及隨機抽樣演算法進行比較，以驗證其在不同數據結構之分類準確性上的表現。

關鍵詞：主動式學習、初始樣本、節點中心度

指導教授簽名：

齊湘伶