

以深層類神經網路標記中文階層式多標籤語意概念

Hierarchical Multi-Label Chinese Word Semantic Labeling using Deep Neural Network

周瑋傑 Wei-Chieh Chou
國立交通大學電機工程學系
Department of Electrical Engineering
National Chiao Tung University
m0450743.eed04g@g2.nctu.edu.tw

王逸如 Yih-Ru Wang
國立交通大學電機工程學系
Department of Electrical Engineering
National Chiao Tung University
yrwang@mail.nctu.edu.tw

摘要

傳統上對超過 100 個階層式標籤分類可以使用扁平 (flatten) 標籤做分類，但如此會喪失架構樹 (taxonomy) 的階層資訊。本研究旨在對廣義知網中文詞彙做概念分類與標記，提出考慮廣義知網架構樹階層關係之深層類神經網路訓練方法，其輸入為詞彙樣本點的詞向量，詞向量方面本研究亦提出考慮上下文前後關係之 2-Bag Word2Vec，而各階層的訓練結果有不同的重要性，所以在模型的最後使用最小分類誤差法，賦予各階層在測試階段時不同的權重。實驗結果顯示階層式 (hierarchical) 分類預測正確率會比扁平分類還高。

Abstract

Traditionally, classifying over 100 hierarchical multi-labels could use flatten classification, but it will lose the taxonomy structure information. This paper aimed to classify the concept of word in E-HowNet and proposed a deep neural network training method with hierarchical relationship in E-HowNet taxonomy. The input of neural network is word embedding. About word embedding, this paper proposed order-aware 2-Bag Word2Vec. Experiment results shown hierarchical classification will achieved higher accuracy than flatten classification.

關鍵詞：詞向量、類神經網路、最小分類誤差、廣義知網、階層式分類、多標籤分類

Keywords: Word2Vec、neural network、minimum classification error、E-HowNet、hierarchical classification、multi-label classification