

PHÂN TÍCH CÁC THÀNH PHẦN TRONG VĂN BẢN PHÁP LUẬT SỬ DỤNG PHƯƠNG HỌC SÂU

Nguyen Truong Son¹, Nguyen Le Minh²

¹Khoa Công nghệ Thông tin,
Trường Đại học Khoa học Tự Nhiên, ĐHQG-HCM

²Viện Khoa học và Công nghệ Tiên tiến Nhật Bản
ntson@fit.hcmus.edu.vn , nguyenml@jaist.ac.jp

Tóm tắt

Phân tích các thành phần của câu trong văn bản pháp luật là một bước quan trọng để xây dựng ác hệ quản lý tri thức trong lĩnh vực luật pháp. Bài báo này đề xuất các phương pháp dựa trên học sâu để nhận ra các thành phần điều kiện-hệ quả trong văn bản pháp luật. Đầu tiên, chúng tôi đề xuất một mô hình dựa trên BiLSTM-CRF mà cho phép tích hợp thêm các đặc trưng từ loại và cú pháp vào các mô hình học sâu. Chúng tôi cũng đề xuất một số mô hình mới như là Multilayer-BiLSTM-CRF và Multilayer-BiLSTM-MLP-CRF để nhận diện các thành phần điều kiện-hệ quả mà bị chồng lấp. Kết quả thực nghiệm trên một số bộ dữ liệu chuẩn đã thể hiện các ưu điểm của mô hình đề xuất.

Từ khóa: Deep learning, Long short-term memory, Sequence labeling, Legal Texts, Recognizing Requisite and Effectuation

ANALYZING LOGICAL PARTS IN LEGAL TEXTS USING DEEP LEARNING

Nguyen Truong Son¹, Nguyen Le Minh²

¹Faculty of Information Technology, University of Science, VNU-HCM

²Japan Advanced Institute of Science and Technology

ntson@fit.hcmus.edu.vn , nguyenml@jaist.ac.jp

Abstract

Analyzing structures of legal sentences is an important step to build knowledge management systems in the legal domain. This paper proposes deep learning-based approaches for recognizing requisite and effectuation (RE) parts in legal texts. First, we propose a deep learning model based on BiLSTM-CRF which allows incorporating external features along with deep learning models such as Part-of-Speech or syntactic features. We also propose several novel models for recognizing overlapping RE parts such as Multilayer-BiLSTM-CRF and Multilayer-BiLSTM-MLP-CRF. Experimental results on several benchmark datasets exhibit the advantages of the proposed approach.

Key words: Deep learning, Long short-term memory, Sequence labeling, Legal Texts, Recognizing Requisite and Effectuation