Lời cam đoan

Lời cảm ơn

Mục lục

Danh sách hình vẽ

Danh sách bảng

Danh sách từ viết tắt

BAĐT Bệnh án điện tử

Chương 1

Tổng quan

1.1 Giới thiệu đề tài

1.2 Mục tiêu và phạm vi đề tài

1.3 Cấu trúc luận văn

Chương 2

Các công trình liên quan

2.1 Bệnh án điện tử

2.2 Các hướng khai thác dữ liệu bệnh án điện tử

Chương 3

Kiến thức nền tảng

Chương 4

Tổng quan hệ thống

4.1 Nội dung bài toán

4.2 Quy trình huấn luyện hệ thống phân loại

Dựa vào hệ thống tốt nhất của thử thách i2b2 năm 2011, mô hình phân giải đồng tham chiếu mà nhóm sử dụng để hiện thực hệ thống là mô hình cặp thực thể. Tư tưởng cơ bản của mô hình này là xác định xem hai khái niệm bất kì có đồng tham chiếu với nhau hay không, sau đó gom nhóm các cặp đồng tham chiếu giao nhau (có một khái niệm chung) lại để tạo thành các chuỗi đồng tham chiếu. Để xác định tính đồng tham chiếu giữa hai khái niệm bất kì, ta cần huấn luyện một model phân loại dựa trên dữ liệu mẫu. Ngoài ra, việc xác định các khái niệm lớp Person có phải đang chỉ về bệnh nhân hay không đóng góp một phần lớn để phân loại đúng tính đồng tham chiếu của các cặp khái niệm ở lớp này. Cũng theo Yan Xu (2011), các đại từ đa phần tham chiếu tới khái niệm thuộc lớp Person, Problem, Test hay Treatment nào đó đứng trước, do đó việc xác định một đại từ thuộc lớp nào trong 4 lớp này là một việc quan trọng.

Như vậy mục đích của quy trình huấn luyện là xây dựng tổng cộng 6 SVM model, trong đó 4 SVM model nhằm mục đích phân loại và đánh giá độ tin cậy đồng tham chiếu của các cặp khái niệm Person-Person, Problem-Problem, Test-Test và Treatment-Treatment; 1 SVM model để xác định các khái niệm Person có là bệnh nhân hay không và 1 SVM model để phân loại các đại từ (các khái niệm lớp Pronoun) vào một trong bốn lớp Person, Problem, Test và Treatment. Đầu vào của quy trình này là toàn bộ các văn bản BADT với các khái niệm đã được trích xuất và gán nhãn. Sau khi tiền xử lý, hệ thống xây dựng các mẫu huấn luyện bao gồm: Person, Person-Person, Problem-Problem, Test-Test, Treatment-Treatment và Pronoun từ danh sách các khái niệm. Sáu tập mẫu này được trích xuất thuộc tính và đưa vào để huấn luyện 6 SVM model. Thư viện SVM được nhóm sử dụng là LibSVM.

4.3 Quy trình phân giải đồng tham chiếu

Quy trình phân giải đồng tham chiếu sử dụng 6 SVM model đã được huấn luyện để

Chương 5

Chi tiết hệ thống

5.1 Tiền xử lý

5.2 Xây dựng cặp khái niệm

5.3 Rút trích đặc trưng

Nhóm Person

Nhóm Patient

Nhóm Pronoun

Nhóm Problem/Test/Treatment

5.4 Gom cụm và xây dựng chuỗi đồng tham chiếu

5.5 Đánh giá hiệu năng

Hệ đo MUC

Hệ đo B-CUBED

Hệ đo CEAF

Chương 6

Thí nghiệm đánh giá

6.1 Tập dữ liệu

6.2 Kết quả

Chương 7

Tổng kết