

BẢN NHẬN XÉT KHÓA LUẬN TỐT NGHIỆP

(HƯỚNG NGHIÊN CỨU)

Tên đề tài: Kết hợp các đối tượng và đặc trưng lân cận cho bài toán phân lớp đa nhãn

Sinh viên thực hiện: - *Trần Xuân Quý* - 18120231

- *Trần Hữu Chí Bảo* - 18120288

Giảng viên hướng dẫn: *Lê Hoài Bắc*

1. Chủ đề và ý tưởng nghiên cứu:

Phân lớp dữ liệu là một trong những bài toán cơ bản của ngành Khoa học Dữ liệu. Trong thực tế, dữ liệu có từ rất nhiều nguồn khác nhau, với sự hỗ trợ của máy tính, có thể trích rút ra các tri thức từ dữ liệu, từ đó hỗ trợ con người đưa ra các quyết định. Đối với việc học có giám sát, bài toán phân lớp dữ liệu về bản chất là xây dựng một hàm ánh xạ từ tập dữ liệu (dataset) của miền ứng dụng vào một tập nhãn cho trước. Trong đó, một đối tượng phân lớp hoặc một mẫu huấn luyện có thể được biểu diễn bằng các tập đặc trưng khác nhau. Phân lớp dữ liệu đã nhận được sự quan tâm rất lớn của các nhà nghiên cứu trong nhiều lĩnh vực khác nhau như: học máy, thống kê, hệ chuyên gia... Công nghệ này cũng được sử dụng trong nhiều lĩnh vực khác nhau như: thương mại, marketing, nghiên cứu thị trường, bảo hiểm, y tế, giáo dục...

Bài toán phân lớp đa nhãn được chia thành hai hướng giải quyết chính là tiếp cận chuyển đổi bài toán và tiếp cận thích nghi thuật toán. Tiếp cận chuyển đổi bài toán là phương pháp chia nhỏ bài toán đa nhãn thành một hoặc nhiều bài toán phân lớp đơn nhãn (phân lớp đa lớp hoặc phân lớp nhị phân) hoặc các bài toán hồi quy. Theo hướng tiếp cận này, các bộ phân lớp đơn nhãn được huấn luyện sau đó các dự đoán phân lớp đơn nhãn được kết hợp lại thành kết quả dự đoán đa nhãn. Phương pháp này khá phù hợp đối với các bài toán dữ liệu ít và nhãn không có mối liên quan và

có thể sử dụng linh hoạt nhiều phương pháp để phân lớp đơn nhãn như SVM, Naive Bayes, kNN và Neural network.

Trong khóa luận này, các sinh viên đã thực hiện đề tài "*kết hợp các đối tượng và đặc trưng lân cận cho bài toán phân lớp đa nhãn*". Ở đây, áp dụng kỹ thuật phân lớp k-NN cho bài toán phân loại đa nhãn dựa trên kỹ thuật chọn phần tử láng giềng trên một số tập dữ liệu thực tế đã được tiền xử lý nhằm giảm chi phí tính toán. Dựa trên đối tượng và đặc trưng, lần lượt áp dụng các kỹ thuật của phương pháp lọc cộng tác dựa trên người dùng và dựa trên từng mục và sau cùng là tổ hợp tuyến tính của các thuật toán trên.

2. Phương pháp nghiên cứu:

Sử dụng phương pháp kết hợp các đối tượng và đặc trưng lân cận cho bài toán phân lớp đa nhãn

3. Kết quả khóa luận:

- Dựa vào kết quả nghiên cứu của Len Feremans và cộng sự (tài liệu tham khảo số [15] của Khóa luận), kết hợp với quy trình cập nhật, khóa luận đã đề xuất thuật toán kết hợp dựa trên đối tượng và dựa trên đặc trưng đã cho thấy kết quả được tốt hơn.
- Thuật toán LCIF là thuật toán đã xử lý một số vấn đề của quá trình phân lớp dữ liệu đa nhãn: chuẩn hoá dữ liệu, lựa chọn thuộc tính tốt nhất, mối quan hệ giữa các nhãn, .. Quá trình triển khai, cài đặt thử nghiệm cùng với các đánh giá hiệu năng mô hình phân lớp LCIF đã được tiến hành. Và đã thu được nhiều kết quả có ý nghĩa thực tiễn, cũng như các kết quả mở ra những hướng nghiên cứu tiếp theo.
- Khóa luận cũng thực hiện cài đặt một phương pháp tính độ tương đồng khác cho dự đoán đa nhãn dựa trên đối tượng với hy vọng phần nào giải quyết một số vấn đề về số lượng và giá trị của những đặc trưng chung giữa các đối tượng có thể ảnh hưởng tới kết quả tương đồng truyền thống, từ đó cải thiện hiệu năng của mô hình cho bài toán phân lớp đa nhãn.

4. Quá trình thực hiện:

Có nhiều kiến thức về Y – Sinh là các tri thức mới mà các sinh viên phải tìm hiểu. Các sinh viên đã có nhiều nỗ lực trong nghiên cứu và tìm tòi. Các sinh viên đã chứng tỏ được khả năng tự nghiên cứu và có phương pháp làm việc tốt. Đã vận dụng được các tri thức đã học vào bài toán thực tế.

Chương trình Demo thể hiện tốt nội dung của khóa luận. Các sinh viên có kỹ thuật lập trình khá tốt

5. Trình bày trước hội đồng:

- Trình bày và trả lời câu hỏi khá tốt
- Slide báo cáo tốt.


6. Công bố khoa học/ ứng dụng thực tế:

Chưa có công bố,

Đánh giá xếp loại: Khóa luận đạt yêu cầu tốt của khóa luận tốt nghiệp Đại học.

TP.HCM, ngày 15 tháng 03 năm 2022

Giảng viên hướng dẫn



Lê Hoài Bắc