

NHẬN DIỆN CẢM XÚC KHUÔN MẶT TRÊN BỘ DỮ LIỆU ĐƯỢC CHÚ THÍCH KHÔNG NHẤT QUÁN

Bùi Quốc Thịnh^{1, 4}

Đỗ Minh Khôi^{2, 4}

Nguyễn Nguyên Khôi^{3, 4}

{¹20520934, ²21521007, ³21521009}@gm.uit.edu.vn

⁴Trường ĐH Công nghệ Thông tin, ĐHQG HCM

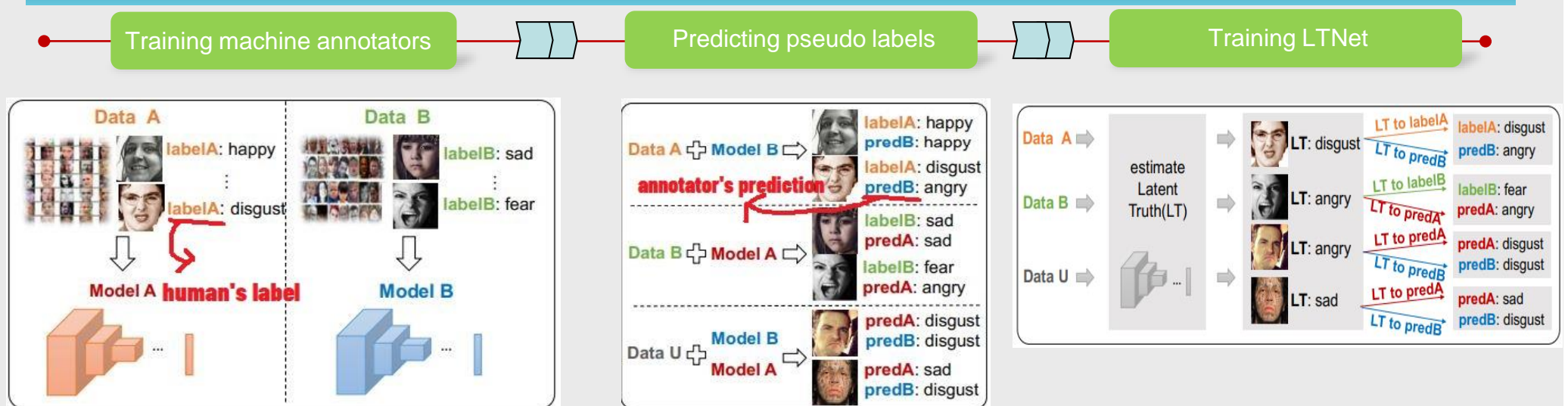
Mục tiêu

- Tìm hiểu cách phân loại từ nhiều bộ dữ liệu với các nhãn khác nhau để giải quyết sự gắn nhãn không nhất quán trong các bộ dữ liệu FER khác nhau.
- Giới thiệu framework IPA2LT để huấn luyện mô hình FER từ nhiều bộ dữ liệu được gắn nhãn không thống nhất và dữ liệu không được gắn nhãn ở quy mô lớn.
- Đánh giá và so sánh hiệu quả của LTNet và IPA2LT với các phương pháp trước đây từng được nghiên cứu.

Lý do chọn đề tài

- Tự động hóa việc nhận diện cảm xúc khuôn mặt (facial expression recognition - FER) đóng vai trò quan trọng trong việc thấu hiểu hành vi của con người.
- Các mô hình hiện nay khó có thể tạo ra một classifier để hoạt động với dữ liệu mới, không được chú thích.
- Chúng tôi đề xuất mô hình IPA2LT và kiến trúc LTNet để khắc phục nhược điểm trên.

Giới thiệu



Mô tả

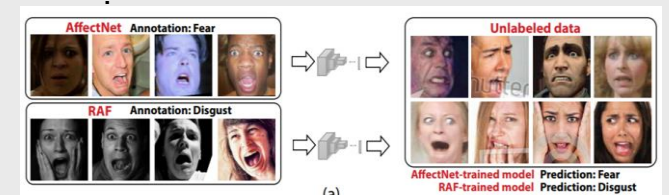
1. Nội dung

- Đề xuất phương pháp đào tạo bộ phân loại từ các bộ dữ liệu được đánh nhãn không nhất quán và nhiễu, xem xét hai loại phương pháp xử lý nhãn không nhất quán và nhãn nhiễu.
- Các phương pháp có nhãn không nhất quán sử dụng nhãn mềm hoặc ước tính ground truth bằng thuật toán EM và các framework xác suất.
- Các phương pháp có nhãn nhiễu sử dụng dữ liệu sạch để đánh giá chất lượng nhãn hoặc đào tạo trình trích xuất đặc trưng hoặc giả định các ràng buộc hoặc phân phối bổ sung trên các nhãn nhiễu.
- LTNet khác với các phương pháp khác bằng cách xem xét nhiều chú thích trên mỗi mẫu, cho phép khám phá các mẫu nhiễu từ các chú thích này.
- Mục tiêu của LTNet là đào tạo công cụ dự đoán cho các mẫu chưa nhìn thấy và nắm bắt mối quan hệ giữa dữ liệu đầu vào và nhãn thực.

2. Phương pháp

- Phương pháp IPA2LT được sử dụng để đào tạo mô hình FER từ nhiều bộ dữ liệu được gắn nhãn không nhất quán, sử dụng nhiều dữ liệu chưa được gắn nhãn cùng với một số bộ dữ liệu được gắn nhãn bởi con người.
- Mỗi mẫu có nhiều hơn một chú thích, bao gồm cả những chú thích quan sát được hoặc dự đoán. LTNet là một mô hình end-to-end được nhúng với sơ đồ khám phá các latent truths từ nhiều nhãn được quan sát (hoặc dự đoán) và hình ảnh đầu vào.
- Mục tiêu của LTNet là tìm giá trị tiên đoán của các latent truth để đưa ra dự đoán tốt nhất cho bộ dữ liệu đầu vào. Tổng quan mô hình cũng được đề cập tại phần giới thiệu.

- Từ phương pháp IPA2LT đề xuất, tìm hiểu cách tạo chú thích giả cho cả dữ liệu được gắn nhãn và không được gắn nhãn của con người bằng cách sử dụng các dự đoán của mô hình phân loại.



- Tiếp theo, tìm hiểu cách huấn luyện mạng LTNet để khám phá latent truth khi xem xét các nhãn không nhất quán và hình ảnh đầu vào.
- Từ đó, trong quá trình suy luận, LTNet sau khi được huấn luyện có thể được sử dụng để ước tính nhãn thực cho hình ảnh khuôn mặt mới.