

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

Bùi Quốc Thịnh – 20520934
Nguyễn Nguyên Khôi – 21521009
Đỗ Minh Khôi - 21521007

Tóm tắt

- Lớp: CS519.N11
- Link Github: <https://github.com/sirrtt/CS519.N11>
- Link YouTube video:



Bùi Quốc Thịnh - 20520934

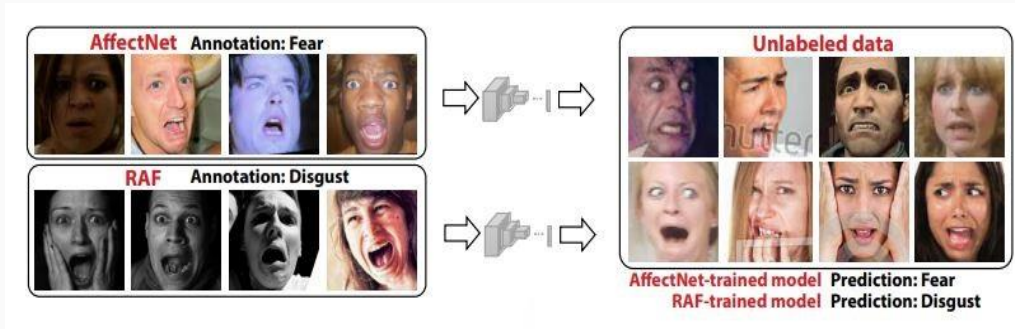


Đỗ Minh Khôi - 21521007



Nguyễn Nguyên Khôi - 21521009

Giới thiệu



Mỗi classifier có cách chú thích cảm xúc khác nhau, dẫn đến nhiều nhãn khác nhau cho cùng một bức ảnh khuôn mặt

Mục tiêu

- Đề xuất một vấn đề tương đối chưa được khám phá: làm thế nào để học một bộ phân loại từ nhiều bộ dữ liệu với các tùy chọn chú thích khác nhau. Giải quyết sự không nhất quán của chú thích 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.
- Các thử nghiệm trên cả dữ liệu tổng hợp và dữ liệu thực xác nhận tính hiệu quả của phương pháp được đề xuất trong việc học từ các nhãn không nhất quán. Đánh giá và so sánh hiệu quả của phương pháp được đề xuất là LTNet và IPA2LT với các phương pháp trước đây từng được nghiên cứu.

Nội dung và Phương pháp

Nội dung:

- Phương pháp đề xuất để đào tạo bộ phân loại từ các bộ dữ liệu được dán nhãn không nhất quán với nhiều nhãn nhiều.
- Có hai loại phương pháp được xem xét: phương pháp xử lý các nhãn không nhất quán và phương pháp xử lý các nhãn nhiều.
- Các phương pháp có nhãn không nhất quán thường sử dụng nhãn mềm trong quá trình đào tạo, nhưng bỏ qua độ tin cậy của chú thích.

Nội dung và Phương pháp

- Các phương pháp ước tính ground truth trong crowdsourcing sử dụng thuật toán EM và các framework xác suất để xác định annotators và ước tính latent truths, nhưng tập trung vào 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.
- Các phương pháp có nhãn nhiều thường tận dụng một tập hợp nhỏ 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 nhiều.
- LTNet cho phép xem xét nhiều chú thích trên mỗi mẫu, cho phép khám phá các mẫu nhiều từ các chú thích này.

Nội dung và Phương pháp

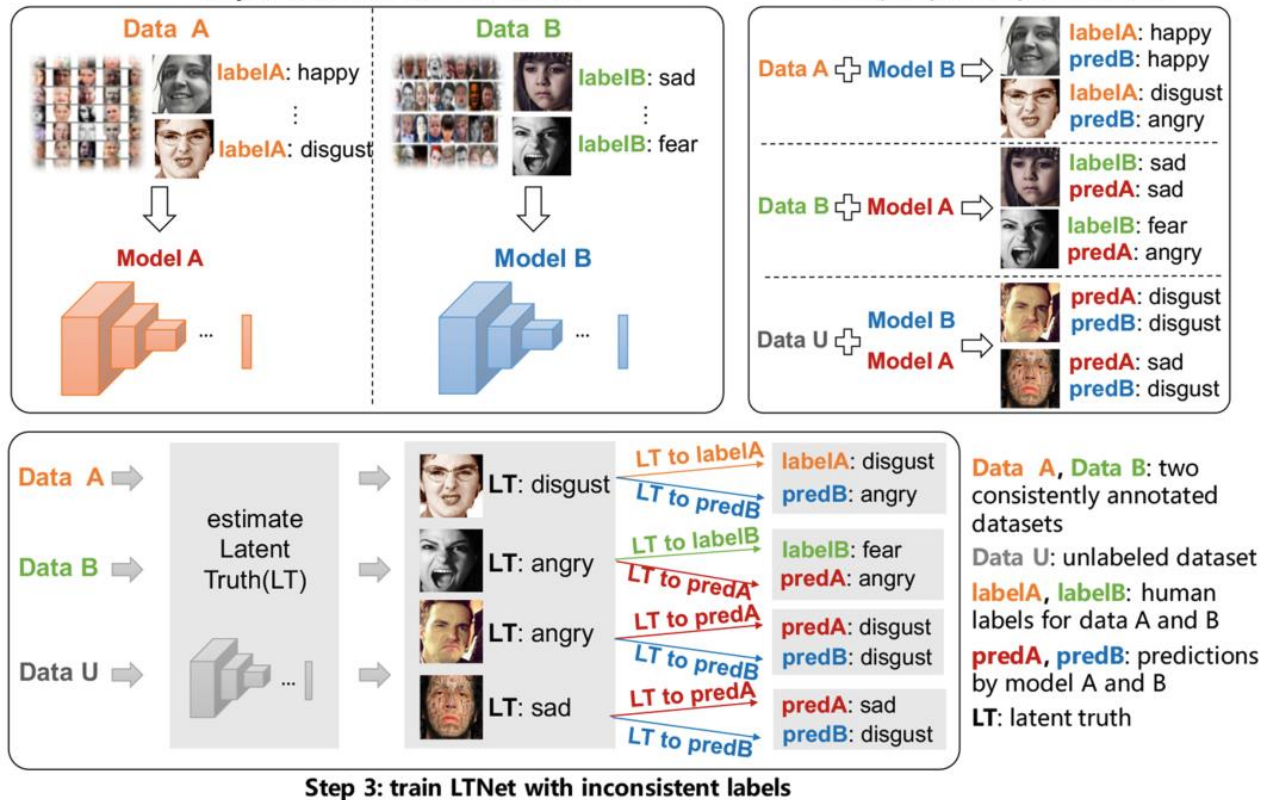
Phương pháp:

- Phương pháp IPA2LT được sử dụng để đào tạo mô hình FER từ nhiều bộ dữ liệu được gán nhãn không nhất quán, sử dụng nhiều dữ liệu chưa được gán nhãn cùng với một số bộ dữ liệu được gán 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.

Nội dung và Phương pháp

Step 1: train machine annotators

Step 2: predict pseudo labels



- Tìm hiểu cách tạo chú thích giả cho cả dữ liệu được gán nhãn và không được gán nhãn của con người.
- Huấn luyện mạng LTNet để khám phá latent truth.
- Ước tính nhãn thực cho hình ảnh khuôn mặt mới

Kết quả dự kiến

- Đánh giá được hiệu suất của các phương pháp khác nhau trong việc xử lý các nhãn nhiễu và không nhất quán trong bộ dữ liệu hình ảnh.
- So sánh LTNet với các phương pháp hiện đại khác để giải quyết các nhãn nhiễu hoặc không nhất quán và độ chính xác kiểm tra của tất cả các phương pháp đã được đánh giá.
- Cho thấy rằng việc sử dụng tất cả các bộ được dán nhãn không nhất quán đã tăng hiệu suất của các mô hình sử dụng một bộ nhiễu duy nhất.
- Kết luận được các phương pháp được đào tạo về dữ liệu hỗn hợp, các phương pháp end-to-end, bao gồm cả LTNet, tốt hơn đáng kể so với các phương pháp từng bước.
- Có một chương trình DEMO để trực quan hóa nghiên cứu

Tài liệu tham khảo

- [1]. Jiabei Zeng, Shiguang Shan, and Xilin Chen: Facial Expression Recognition with Inconsistently Annotated Datasets. ECCV 2018: 222-237.
- [2]. Yu Zhu, Bosong Ding, Chenghua Li, Wanli Qian, Fangya Li, Yiheng Yao, Ruipeng Gang, Chunjie Zhang, Jian Cheng: LTNet: Light Transfer Network for Depth Guided Image Relighting. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops 2021: pp. 243-251.
- [3]. Shikai Chen, Jianfeng Wang, Yuedong Chen, Zhongchao Shi, Xin Geng, Yong Rui: Label Distribution Learning on Auxiliary Label Space Graphs for Facial Expression Recognition. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2020: pp. 13984-13993.
- [4]. Darshan Gera, S. Balasubramanian: Noisy Annotations Robust Consensual Collaborative Affect Expression Recognition. IEEE/CVF International Conference on Computer Vision (ICCV) Workshops 2021: pp. 3585-3592.