ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN KHOA HỌC MÁY TÍNH

NHẬN DẠNG
ĐỀ CƯƠNG NGHIÊN CỨU

Giảng viên lý thuyết
PGS. TS. Lê Hoàng Thái

Giảng viên hướng dẫn

Lê Thanh Phong, Nguyễn Ngọc Thảo

Sinh viên thực hiện

Nguyễn Hoàng Đức, Lê Nhựt Nam, Nguyễn Viết Dũng

Tháng 4 năm 2021

Mục lục

1. THÔNG TIN NHÓM

- Thành viên 01: Nguyễn Hoàng Đức - 18120018
- Thành viên 02: Lê Nhựt Nam - 18120061
- Thành viên 03: Nguyễn Viết Dũng - 18120167

2. THÔNG TIN ĐỀ TÀI

Tên đề tài

- Tên đề tài (Tiếng Việt): Sinh trắc học giọng nói
- Tên đề tài (Tiếng Anh): Voice Biometrics

Nguồn tham khảo

- Sách: Chương 8: Voice Biometrics, Handbook of Biometric, Anil K. Jain, Patrick Flynn, Arun A. Ross
- Paper with code: Speaker Recognition

Danh sách các bài báo khoa học tham khảo:

- Deep neural networks for small footprint text-dependent speaker verification, Ehsan Variani (Johns Hopkins Univ., Baltimore, MD, USA), Xin Lei (Google Inc., USA), Erik McDermott (Google Inc., USA), Ignacio Lopez Moreno (Google Inc, Mountain View, CA, US), Javier Gonzalez-Dominguez (Google Inc., USA), 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy
- Multi-Task Learning for Text-Dependent Speaker Verification, Nanxin Chen, Yanmin Qian, Kai Yu (Shanghai Jiao Tong University, China), INTERSPEECH 2015 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, 2015
- X-Vectors: Robust DNN Embeddings for Speaker Recognition, David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey, Sanjeev Khudanpur (Center for Language and Speech Processing & Human Language Technology Center of Excellence, The Johns Hopkins University, Baltimore, MD, USA), 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 2018
- Mirco Ravanelli and Yoshua Bengio. Speaker recognition from raw waveform with sincnet, 2019.

Từ khóa

- Tên từ khóa (Tiếng Việt): Nhận dạng sinh trắc học, Nhận dạng giọng nói, Nhận dạng người nói, Mạng Neural Tích chập, Mẫu thô, Xác minh người nói, Định danh người nói
- Tên từ khóa (Tiếng Anh): Biometric Recognition, Voice Recognition, Speaker Recognition, Convolutional Neural Networks, Raw Samples, Speaker Verification, Speaker Identification

Nội dung trình bày

Nội dung trình bày trước lớp gồm 2 phần: trình bày nội dung tìm hiểu được từ chương 08 - Voice Biometrics sách Handbook of Biometric, trình bày những phương pháp state-of-the-art về lĩnh vực nhận dạng giọng nói trong thời gian gần đây.

Phần trình bày nội dung tìm hiểu được từ sách

- Giới thiệu chung
- Những thông tin nhận dạng trong tín hiệu giọng nói

- Rút trích đặc trưng và phân tách
- Hai công nghệ chính của lĩnh vực nhận dạng giọng nói

Phần trình bày những phương pháp state-of-the-art về lĩnh vực nhận dạng giọng nói

- Giới thiệu
- Động lực nghiên cứu khoa học
- Phát biểu bài toán
- Các công trình tiêu biểu (khoảng 3 công trình)
  - Deep neural networks for small footprint text-dependent speaker verification - Đại diện cho d-vectors (Deep Vectors)
  - Multi-Task Learning for Text-Dependent Speaker Verification - Đại diện cho j-vectors
  - X-Vectors: Robust DNN Embeddings for Speaker Recognition - Đại diện cho x-vectors
  - Speaker recognition from raw waveform with sincnet - Đại diện cho Phân lớp người nói
- So sánh d-vectors, j-vectors và x-vectors
- Demo
- Tài liệu tham khảo

Nội dung báo cáo

- Thông tin nhóm
- Thông tin đề tài
- Nội dung phân công
- Nội dung báo cáo
  - Nội dung tìm hiểu từ sách Handbook of Biometric
    * Giới thiệu chung
    * Thông tin nhận dạng trong tín hiệu giọng nói
    * Rút trích đặc trưng và phân tách thông tin
      · Phân tích theo từng đoạn ngắn
      · Tham số hóa
      · Phân tách ngữ âm và tách từ
      · Phân tách ngữ điệu
    * Công nghệ nhận dạng giọng nói phụ thuộc văn bản
    * Công nghệ nhận dạng giọng nói không phụ thuộc văn bản
  - Các phương pháp SOTA
    * Giới thiệu chung định hướng gần đây
    * Động lực nghiên cứu
    * Phát biểu bài toán: Đầu vào (Input), Đầu ra (Output)
    * Kho ngữ liệu
    * Kiểm định mô hình nhận dạng giọng nói
    * Các độ đo thường dùng trong nhận dạng giọng nói
    * Các công trình tiêu biểu

        - Trong tác vụ rút trích đặc trưng
            - d-vectors
            - j-vectors
            - x-vectors
            - So sánh d-vectors, j-vectors và x-vectors
        - Trong phân lớp người nói
            - Variational autoencoder
            - Multi-domain features
            - SincNet
    - Thực nghiệm: Sử dụng SincNet trong xác minh và định danh người nói
        * Chuẩn bị dữ liệu
        * Xây dựng mô hình
        * Đăng ký
        * Đánh giá

Xây dựng demo cho chủ đề nghiên cứu

- Phương pháp giải quyết vấn đề
    - Dựa trên source code chính thức SincNet từ tác giả
    - Huấn luyện mô hình xác minh người nói bằng tiếng Anh, tiếng Việt
- Dữ liệu thực nghiệm
    - TIMIT Acoustic-Phonetic Continuous Speech Corpus
        * Tác giả: John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, Nancy L. Dahlgren, Victor Zue
        * Năm: 1993
        * DCMI Type(s): âm thanh - sound
        * Sample Type: 1-channel pcm (mono channel)
        * Nguồn thu dữ liệu: microphone speech
        * Ngôn ngữ: tiếng Anh
        * Kích thước: 630 người nói từ 8 đặc trưng giọng Anh Mỹ, 1,3GB
    - Librispeech
        * Tên đầy đủ: LibriSpeech ASR corpus
        * Bài báo: LibriSpeech: an ASR corpus based on public domain audio books
        * Tác giả/ Nhóm tác giả: Vassil Panayotov, Guoguo Chen, Daniel Povey and Sanjeev Khudanpur
        * Được công bố tại hội nghị The international Conference on Acoustics, Speech, & Signal Processing (ICASSP), 2015
        * Mô tả sơ lược: Kho ngữ liệu có kích thước khoảng 1000 giờ nói tiếng Anh với tần số 16kHz
    - Son et al. Dataset
        * Bài báo: Vietnamese Speaker Authentication Using Deep Models

    * Dung lượng của tập dữ liệu: 535 MB
    * Số mẫu trong tập dữ liệu: 400 mẫu
    * Bộ dữ liệu gồm: hai tập Men và Women, mỗi tập con chứa 10 thư mục người nói. Mỗi thư mục người nói chứa 20 đoạn ghi âm, chia ra Long và Short (mỗi loại 10 đoạn)
    * Nội dung câu nói
        · Câu ngắn: “Tôi là sinh viên chuyên ngành công nghệ thông tin"
        · Câu dài: "Tôi là sinh viên Học viện Công nghệ Bưu chính Viễn thông, chương trình đào tạo khá nặng đòi hỏi sinh viên phải học tập và nghiên cứu rất nhiều nhưng tôi tự hào vì đó là ngành đã và đang làm thay đổi cuộc sống xã hội loài người".
    * Điểm hạn chế: Bộ dữ liệu có kích thước khá nhỏ

- Thực nghiệm và đánh giá
  - Trực quan hóa quá trình huấn luyện dựa trên các thông số: average training loss, classification error, average test loss, classification error frame level test data, classification error sentence level test data
  - Dự đoán người nói từ mô hình
  - Xác minh người nói bằng cosine similarity

3. THÔNG TIN TỰ ĐÁNH GIÁ TIẾN ĐỘ

Những công việc đã thực hiện được

Về những kiến thức chương 8 - sách Handbook of Biometrics

- Nắm cái nhìn chung về lĩnh vực Nhận dạng Giọng nói.
- Những yếu tố nhận dạng trong tín hiệu hiệu giọng nói.
- Hai công nghệ chính trong lĩnh vực nhận dạng giọng nói.
- Một số kỹ thuật thủ công trong xử lý một số đặc trưng nhận dạng tín hiệu giọng nói.
- Chuẩn bị slides thuyết trình cho phần này.

Về những kiến thức các phương pháp state-of-the-art trong lĩnh vực Nhận dạng Giọng nói

- Nắm sơ bộ những thành tựu gần đây trong lĩnh vực Nhận dạng Giọng nói.
- Nắm sơ bộ key methods, phương pháp, hiệu năng của một số phương pháp Deep Learning trong rút trích đặc trưng tín hiệu giọng nói.
- Nắm sơ bộ key methods, phương pháp, hiệu năng của một số phương pháp Deep Learning trong phân lớp tín hiệu giọng nói.
- Chuẩn bị slides thuyết trình cho phần này.

Về Tìm kiếm open source / libraries / demonstration để minh họa cho nội dung lý thuyết

- Tìm kiếm và đọc hiểu source code SincNet
- Quá trình hoạt động của các bộ lọc, kiến trúc mạng
- Đã huấn luyện thành công mô hình cho tiếng Anh (train và test trên TIMIT, Librispeech) và tiếng Việt (train và test trên tập Son et al. Dataset)
- Đã kiểm thử similarity trên một vài file âm thanh đầu vào với tiếng Anh, tiếng Việt

Những công việc chưa thực hiện được

Về những kiến thức chương 8 - sách Handbook of Biometrics

- Điều chỉnh slides thuyết trình

Về tìm kiếm open source / libraries / demonstration để minh họa cho nội dung lý thuyết

- Hoàn chỉnh 2 giai đoạn đầu tiên của một Speaker Recognition Pipeline (data preparation và development models, basic testing), chưa thực hiện enrollment, evaluation đầy đủ
- Số lượng epochs huấn luyện cho mô hình tiếng Anh, tiếng Việt còn khá thấp với (100 epochs với TIMIT và Librispeech, 300 epochs Son et al. Dataset)

Hướng giải quyết khó khăn

- Tiếp tục tìm hiểu và phát triển hoàn chỉnh Speaker Recognition Pipeline
- Tìm kiếm một bộ dữ liệu cho tiếng Việt lớn hơn