

ADND assignments

General information

Topics and Group creation

- X topics
- max 3 persons / topic
- join to a project using Xoyondo

Expected:

- Deep literature search before jumping into coding
- Planning and pipeline proposal with time estimation
- Shared codebase (GitHub) and documentation (README, research diary) per group
- List of contributions within the research diary (1 slide / person / week)
- Baseline model reproduction, evaluation
- Further contributions, e.g. adding extra modality to the network, trying out different training techniques with proper ablation study, determining the added value

Grading:

Individual grading based on the topics, and supervisors. Example:

1. Literature search and pipeline planning
Weight: 20%
Create slides → max 2 weeks (deadline oct.05), upload it to your topic's canvas discussion
2. Experiments
Weight: 70%
3. Results, producing a finished and coherent project
Weight: 10%

Supervisor: Ádám Fodor

T1: Personality

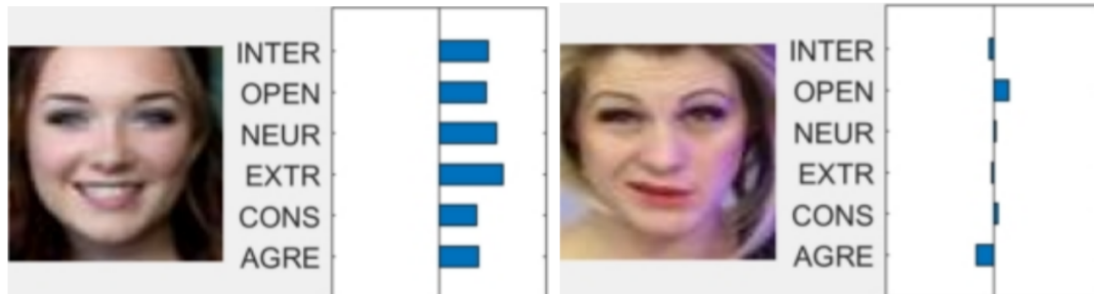


image source: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8999746>

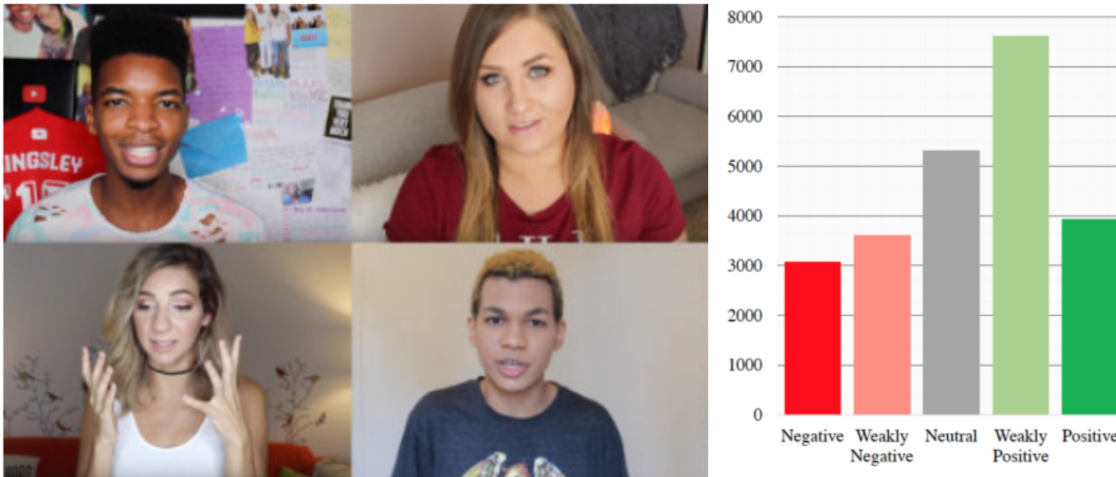
- Task
 - Big Five (OCEAN) personality trait estimation
 - Regression, Binary classification
- Difficulties
 - hard to find open-sourced codebases with reproducible results
 - unbalanced target distribution
 - regression-to-the-mean problem
 - a function, which returns the average GT value (~ 0.52) can reach a high (1-MAE) score: 0.88
 - SOTA models are reaching ~ 0.92 (1-MAE)
 - recent works are exploiting biases in the database: background, attractiveness, gender, age
- Baseline model:
 - MulT: <https://github.com/yaohung/Multimodal-Transformer>
 - SSE-FT: <https://github.com/shamanez/Self-Supervised-Embedding-Fusion-Transformer>
- Dataset
 - First Impression V2 (<https://chalearnlap.cvc.uab.cat/dataset/24/description>)
- Papers
 - 2019: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8999746>
 - 2021: https://openaccess.thecvf.com/content/WACV2021W/HBU/papers/Jacques_Person_Perception_Biases_Exposed_Revisiting
 - Vision + Language: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8424834>
 - Survey: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8769905>
 - ChaLearn: https://link.springer.com/content/pdf/10.1007/978-3-319-49409-8_32.pdf

T2: Emotion



- Task
 - Emotion estimation
 - Classification
- Difficulties
 - mmsdk features hard (or unfeasible) to reproduce (e.g. iMotion's FACET is a paid software)
 - you can use the mmsdk feature set to reproduce published results from a paper or try out a given software from GitHub
 - for the assignment, you must use a different feature set, which makes the results reproducible or adaptable to other datasets as well
- Baseline model:
 - MulT: <https://github.com/yaohungt/Multimodal-Transformer>
 - SSE-FT: <https://github.com/shamanez/Self-Supervised-Embedding-Fusion-Transformer>
- Dataset
 - MOSEI (<http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset>)
 - MELD (<https://affective-meld.github.io>)
 - CREMA-D (<https://github.com/CheyneyComputerScience/CREMA-D>)
- Papers
 - MOSEI introduction: <https://aclanthology.org/P18-1208.pdf>
 - MELD introduction: <https://arxiv.org/pdf/1810.02508.pdf>
 - CREMA-D introduction: https://www.researchgate.net/publication/272081518_CREMA-D_Crowd-sourced_emotional_multimodal_actors_dataset
 - Emotion datasets for general literature search: <https://superkogito.github.io/SER-datasets/>
 - Survey: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9637545>

T3: Sentiment



- Task
 - Sentiment estimation
 - Regression
- Difficulties
 - mmsdk features hard (or unfeasible) to reproduce (e.g. iMotion's FACET is a paid software)
 - you can use the mmsdk feature set to reproduce published results from a paper or try out a given software from GitHub
 - for the assignment, you must use a different feature set, which makes the results reproducible or adaptable to other datasets as well
- Baseline model:
 - MuIT: <https://github.com/yaohungt/Multimodal-Transformer>
 - SSE-FT: <https://github.com/shamanez/Self-Supervised-Embedding-Fusion-Transformer>
- Dataset
 - MOSEI (<http://multicomp.cs.cmu.edu/resources/cmu-mosei-dataset>)
 - MOSI (<http://multicomp.cs.cmu.edu/resources/cmu-mosi-dataset>)
- Papers
 - MOSEI introduction: <https://aclanthology.org/P18-1208.pdf>
 - Survey: <https://reader.elsevier.com/reader/sd/pii/S1566253521001299?token=2E18DC80AD157E87B9BF8054FE9C601EF25CA6433A2F4B2C984E6D5365CA864AB4D17FFEBAB6FB299380AEwest-1&originCreation=20220318131032>

T4: Multilingual



- Task
 - Multilingual Text Analysis
 - NLP, Classification
 - Multimodal experiments on multilingual subsets
 - semantic text similarity
 - multilingual sentiment analysis
- Baseline models:
 - S-BERT <https://arxiv.org/pdf/1908.10084.pdf>
 - XLM-RoBERTa <http://ceur-ws.org/Vol-2826/T4-13.pdf>
- Difficulties
 - experimenting with different modalities
 - multilingual cross-dataset evaluations
 - unique tasks and datasets
- Dataset
 - UDIVA: <https://arxiv.org/pdf/2012.14259.pdf>
 - + 10 multilingual datasets

T5: Conversation

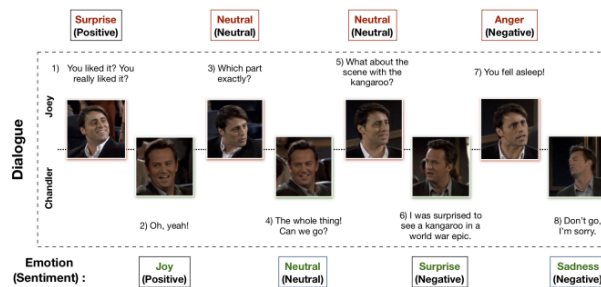
- 1)

P_A You know I am *getting married!* excited

P_B Wow! That's great news. Who is the lucky person? happy
- 2)

P_A Such a disappointing *football match!* disgust

P_B No! I am enjoying it. happy



- Task
 - Emotion recognition in conversations
 - Classification
- Difficulties
 - The research field is heavily using text for conversation analysis (RECOON)
Integrating other modalities (visual and speech) can be difficult due to the lack of data/annotations in variety
- Baseline model:
 - EmoBERTa: <https://github.com/tae898/erc>
- Dataset
 - RECOON: <https://github.com/declare-lab/RECCON>
 - MELD: <https://arxiv.org/pdf/1810.02508.pdf>
 - IEMOCAP: <https://github.com/tae898/erc>
- Papers
 - Multiple methods and codebase: <https://github.com/declare-lab/conv-emotion>
 - Method and Code: <https://github.com/tae898/erc>
 - Method and Code: <https://github.com/declare-lab/conv-emotion/tree/master/emotion-cause-extraction>

Supervisor: Áron Fóthi

T6: Tracking

- Task
 - Tracking birds in a cage with bounding box
- Difficulties
 - Search available labeled datasets
 - Annotate videos
- Database
 - Unlabeled videos
- Papers
 - <https://arxiv.org/pdf/1903.05625.pdf>

T7: Video instance segmentation - Indoor

- Task
 - Furniture segmentation in an indoor environment and on Rats
- Difficulties
 - Synthesize Dataset for training
 - Train Mask2Former algorithm
 - Combine RGB features with Depth
- Database
 - SUN RGB-D
 - Rats
- Papers
 - <https://github.com/facebookresearch/Mask2Former>

T8: Video instance segmentation - Rats

- Task
 - Rat tracking with segmentation
 - Using ClusterRCNN with scribble annotation
- Difficulties
 - Extend VGG VIA with scribble annotation
 - Train ClusterRCNN with scribbles
- Database
 - Rats
- Papers
 - https://link.springer.com/chapter/10.1007/978-3-030-63830-6_37

T9: 3D hand object mesh estimation

- Task
 - Detect objects in hand
- Difficulties
 - Generate scenes with objects and hands
 - Train ACFM https://github.com/fkokkinos/acfm_video_3d_reconstruction
- Database
 - Generated annotated
- Papers
 - <https://arxiv.org/abs/2201.08555>