week3

MSc Project 2021

Title: Estimating personality in communication
Name: Yuichi Midorikawa

weekX	This indicates when it was done
X-X	This corresponds to the mindmap number

week2

1. Summary of actions agreed during last meeting

1-1. I updated my research steps

1-2. I have selected a dataset to use

Time Plan		Oct				Nov					Dec	My Progress	
		4	11	18	25	1	8	15	22	29	6		
Step	Task	week1	week2	week3	week4	week5	week6	week7	week8	week9	week10	Status	Question(Yes/No)
1	capture a dataset that contains people talking and the text of what they say.	1					buffer				buffer	Finished	
	1-1. Find a dataset to use	1-1										Finished	
2	using the conversation text, do sentiment analysis (A)	2										Finished	
	2-1. Find a model to use	2-1										Finished	
	2-2. Using the model and its data set, perform sentiment analysis		2-2									Finished	
3	from the videos, extract people and body pose (B)	3										Running	
	3-1. Find a model to use	3-1										Running	
	3-2. Using the model and its data set, extract body pose		3-2									Running	
4	from the head, extract facial feature points (C)	4										Running	
	4-1. Find a model to use	4-1										Running	
	4-2. Using the model and its data set, extract facial points		4-2									Running	
5	Then train a model to predict (A) from (B)+(C)			5								-	
	5-1. Predict (A) from (B) + (C)			5-1								-	
6	evaluate and analyse the results.				6							-	
	6-1. Decide a evaluation metrics				6-1							-	
	6-2. evaluate and analyse the results					6-2						-	
7	Write a paper							7				-	

week3

1-1

Research Steps (Updated on week 3)

week3
2-1

Data Preparation

1-1. capture a dataset that contains people talking and the text of what they say.

Youtube

No need to conduct it

1-2. Watch the video and manually annotate each subtitle/frame with a positive/negative. (A') (Use as training data)

Positive/
Negative

Implementation (Using a pre-trained model)

2. using the conversation text, do sentiment analysis (A)

①

Positive/
Negative

3. from the videos, extract people and body pose (B)

4. from the head, extract facial feature points (C)

5. Then train a model to predict (A) or (A') from (B)+(C)

②

Positive/
Negative

Evaluation

6. evaluate and analyse the results.

① Positive/
Negative

② Positive/
Negative

Evaluation metrics:
Accuracy,
Precision,
Recall, and
F1-score...

week2

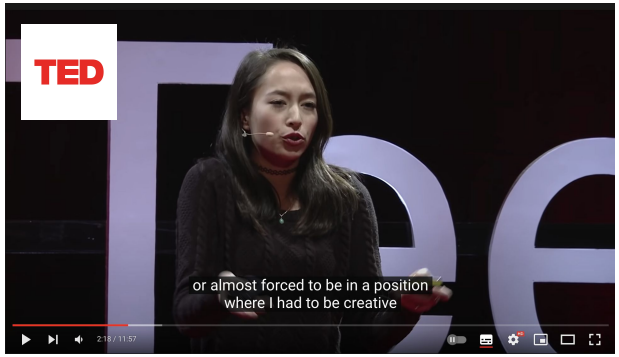
2-2

2-2. I have selected a dataset to use (Step 1-1)

I got the TED video from Youtube to use in this project.

Titile: [Why I Don't Use A Smart Phone | Ann Makosinski | TEDxTeen]

→ I would like to find some other videos with different conditions.



What is the ideal video for Youtube in this project?

- One or two people are talking.
- Facial expressions can be detected.
- Body posture can be detected.
- Emotional expression is as much as possible.
- English subtitles are available.

Research Steps

1. capture a dataset that contains people talking and the text of what they say.
2. using the conversation text, do sentiment analysis (A)
3. from the videos extract people and body pose (B)
4. from the head, extract facial feature points (C)
5. Then train a model to predict (A) from (B)+(C)
6. evaluate and analyse the results.

week3

2-1, 2-2

2. Summary of work done & results this week

2-1. Decided on Emotions Category for this project

2-2. Decided on Text Emotion Classification Model

2-3. Conducted sentiment analysis from a single image

week3

2-1

2-1. Decided on Emotions Category for this project

According to the following paper, the emotion categories **Hierarchical Grouping** and **Ekman** were high F1-score, so for now I will use these emotion categories in this project.

I would like to determine Hierarchical Grouping or Ekman according to the emotional categories that **the facial expressions and body postures can output.**

Emotion category name	Number of emotions	F1-score
Original GoEmotions	27 emotions + neutral	0.46
Hierarchical Grouping	positive, negative, ambiguous + neutral	0.69
Ekman	anger, disgust, fear, joy, sadness, surprise + neutral	0.64

GoEmotions: A Dataset of Fine-Grained Emotions
https://arxiv.org/abs/2005.00547
P. Ekman, Universal facial expressions of emotion
151–158.

week3

2-2

2-2. Decided on Text Emotion Classification Model

The top performing models in the EmotionX Challenge (Hsu and Ku, 2018) all used the pre-trained BERT model.

→ Therefore, I plan to use the **BERT** model in my project as well. I will use an existing, pre-trained model called Go-emotion.

Model	Embeddings	Macro-F1
SVM	-	0.55
Random Forest	-	0.49
BiLSTM+CNN+ Self-Attention	GoobleEmb	0.62
	GloVE	0.62
	FastText	0.63

Table 1: Comparison of BiLSTM+CNN+Self-Attention models for ISEAR

Model	Macro-F1
BERT	0.702
RoBERTa	0.742
DistilBERT	0.693
XLNet	0.731

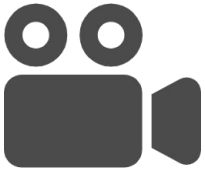
Table 3: Comparison of Transformers models for ISEAR

[1] Chao-Chun Hsu and Lun-Wei Ku. 2018. SocialNLP 2018 EmotionX Challenge Overview: Recognizing Emotions in Dialogues
https://www.researchgate.net/publication/328137931_SocialNLP_2018_EmotionX_Challenge_Overview_Recognizing_Emotions_in_Dialogues
[2] Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA
https://arxiv.org/abs/2104.02041

week3

The process from video collection to emotion recognition analysis

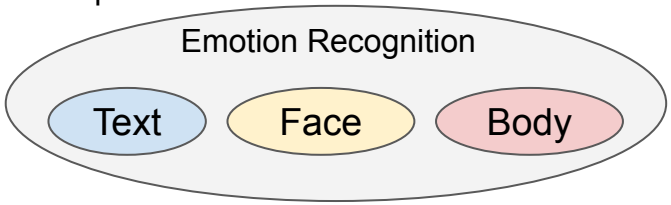
Video (Youtube) → Frames → Selected an image



Step 1 & 2
e.g. The whole Youtube video was 11:57 long, but I cut out the video with the speaker's whole body at 4:39 - 4:48

Step 3
e.g.
When I split the 4:39 - 4:48 video into frames, I got about 250 images.

Step 4



2-3. Conducted sentiment analysis from a single image

Script:
I was like, "Oh well, I want to learn how to skateboard and be cool."

1 frame (selected by hand)

Text

Google BERT

The result of Emotion Recognition from text:
joy

```
In [7]: texts = [
            "I was like,                                    ""
        ]
In [8]: pprint(goemotic                (texts))
[{'labels': ['joy'], 'scores': [0.9993316]}]
```

Face

deepFace

The result of Emotion Recognition from face:
fear

```
In [18]: obj = DeepFac                                    ['age', 'gender', 'race', '
         Action: emotion
In [19]: print(obj)
{'age': 29, 'region': {'x': 541                  66}, 'gender': 'Woman', 'race': {'asian': 51.244670152664185, 'india
n': 14.840932190418                2.4615228176116943, 'white': 4.869714006781578, 'middle eastern': 2.341
9369012117386                hispanic': 24.241232872009277), 'dominant_race': 'asian', 'emotion': {'angry': 1.34370
33630907536, 'disgust': 0.03624255477916449, 'fear': 42.89453625679016, 'happy': 0.0583472428843379,
'sad': 32.361769676208496, 'surprise': 0.6144466809928417, 'neutral': 22.690963745117188}, 'dominant_e
motion': 'fear'}
```

Current	・This week, I performed Emotion Recognition on a single frame using BERT(text) and DeepFace(face). → The results of emotion recognition from text using BERT were different from the results of emotion recognition from facial expressions using DeepFace.
Next	・I would like to perform Emotion Recognition on all frames in the speech, because I don't know at what point in the speech the speaker's emotion is emphasized.

3. Questions to be discussed during the meeting

3-1. I would like to ask for some advice on how to handle frames when analysing sentiment from face and body.
(I'm going to do a sentiment analysis for the frames that the dialogue is in, notjust one frame.)

3-2. (if we have time), I would like to get some tips for finding Facial & Body pose emotion recognition models.
How do I find out any pre-trained models? I could read the papers and find a model that looks good, but it is notpublished. Can I use a paid API or something?

4. Proposed objectives for next week

4-1. I will explore and implement techniques to extract emotions from (1) facial features and (2) body posture. (Step 3-2, 4-2)

4-2. I will perform emotion extraction notfor a single frame, but for the frame in that sentence.

5. Articles read this week

5-1. GoEmotions: A Dataset of Fine-Grained Emotions
https://arxiv.org/abs/2005.00547

5-2. Exploring Transformers in Emotion Recognition: a comparison of BERT, DistillBERT, RoBERTa, XLNet and ELECTRA
https://arxiv.org/abs/2104.02041

5-3. On the Performance Analysis of APIs Recognizing Emotions from Video Images of Facial Expressions
https://www.researchgate.net/publication/329718637_On_the_Performance_Analysis_of_APIs_Recognizing_Emotions_from_Video_Images_of_Facial_Expressions

End

体の姿勢と表情の座標データを使って、感情を予測できるモデルを作成する

テキストから1セリフごとに感情を取得する

1セリフごとから複数のフレームを取得する

そのフレームに対して体の座標と表情の座標を取得する

それらの座標を組み合わせてテキストの

X. 課題

・表情から感情認識するときの精度が良いモデルはどれ？

・姿勢から感情認識するときの精度が良いモデルはどれ？

Emotion Recognition from text

GoEmotions Pytorch
https://github.com/monologg/GoEmotions-pytorch/blob/master/README.md

State-of-the-art Natural Language Processing for Jax, PyTorch and TensorFlow
https://github.com/huggingface/transformers

Emotion Recognition from body posture

VIBE: Video Inference for Human Body Pose and Shape Estimation [CVPR-2020]

https://github.com/mkocabas/VIBE

Emotion Classification in Short Messages

https://github.com/lukasgarbas/nlp-text-emotion

質問(1)

どういったデータセットなのか？（概要を使って説明）

具体的にはどういったデータがあるのか？（テーブルを使って説明）

このデータを使って何をするのか？（＝テーマを改めて説明）

それをどういったスケジュールで進めるのか？

今はどういった作業をしているのか？

代替案

アイデア

（1）fake news （テキスト）

（2）サマリー生成　超長い文章→要約（テキスト）

（3）文章生成（テキスト）

（4）画像生成（画像）

（5）画像＋テキスト：boketeの英語版

（6）芸術系　この絵に対して、どう思う？（アノテーションが必要）

（7）コミュニケーション系＋画像

代替案

Real Life Violence Situations Dataset

https://www.kaggle.com/mohamedmustafa/real-life-violence-situations-dataset

icon
https://icooon-mono.com/