Mobile Affective Inference:
Recent Methods, Applications and Challenges

Ou Changkun   Seminar Media Computer Science, LMU I Munich, Germany I 2018-02-26
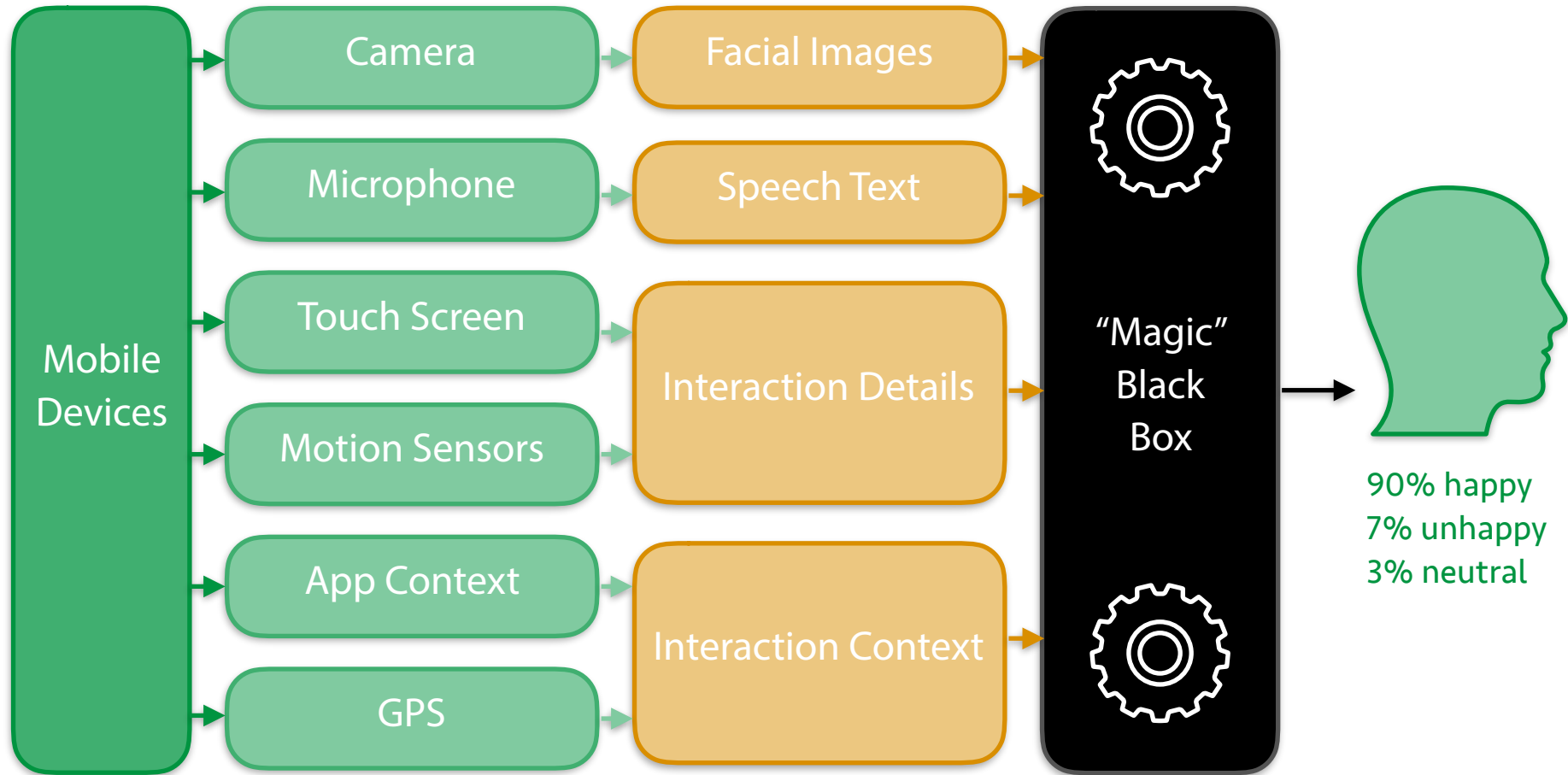
hi@changkun.us

Agenda

- Motivation

- Methods & Models
  - Vision Aspects
  - Speech Aspects
  - Interaction Aspects
  - Multimodal

- Applications

- Challenges

Motivation

- What is *Emotion Inference*?
    - Emotion Inference (aka *Emotion Recognition*) is the process of identifying human emotion, mostly from facial expression
    - Human's affect *expressed* by **various** **channels** in a specific ***context***
    - Emotion recognition relies on massive labeled channel data

- Why Emotion Inference?
    - Intelligence, Emotion-aware User Interfaces, etc.

- Let's survey recent 5-year papers

A General Framework

Mobile Devices

- Camera → Facial Images
- Microphone → Speech Text
- Touch Screen → Interaction Details
- Motion Sensors → Interaction Details
- App Context → Interaction Context
- GPS → Interaction Context

"Magic" Black Box →

90% happy
7% unhappy
3% neutral

Vision Aspects

Facial-based Method & CNN Models

- **Facial expression** is the most important channel for emotion expression;

- **C**onvolutional **N**eural **N**etwork is the recent **breakthrough** model that rules the entire computer vision area;
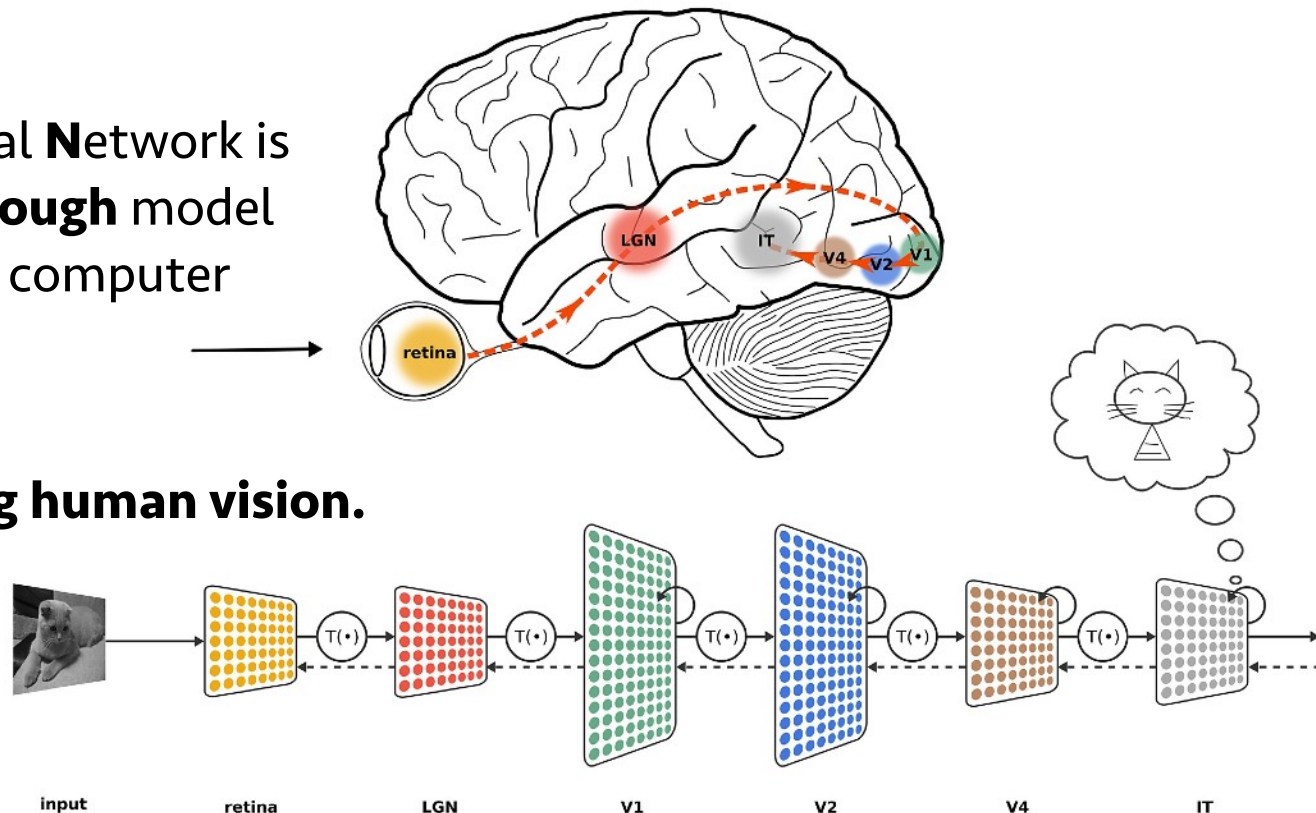
- Basic Idea: **Modeling human vision.**

Image source: https://figshare.com/articles/Ventral_visual_stream/106794

input   retina   LGN   V1   V2   V4   IT

Facial Emotion Inference Steps

- Step0: Prepare dataset [Moolahosseini et al. 2017]

- Step1: Finding Face [He et al. 2017]

- Step2: Emotion Recognition [Howard et al. 2017] [Howard et al. Feb. 2018]
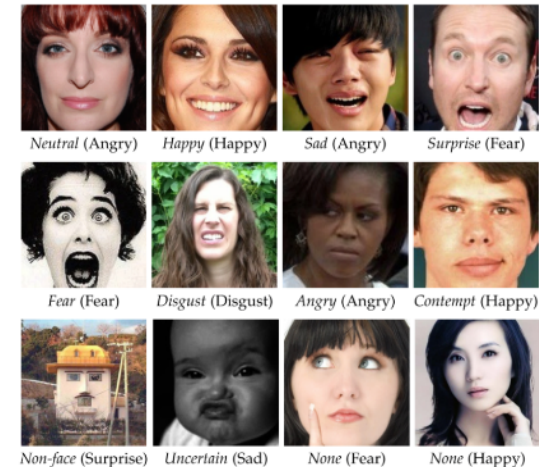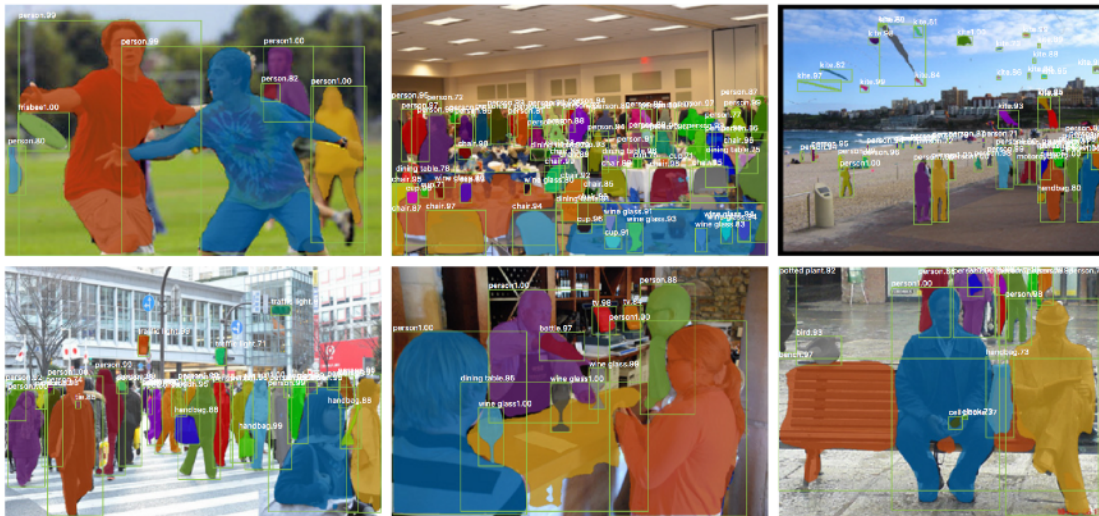

Neutral (Angry) Happy (Happy) Sad (Angry) Surprise (Fear)

Fear (Fear) Disgust (Disgust) Angry (Angry) Contempt (Happy)

Non-face (Surprise) Uncertain (Sad) None (Fear) None (Happy)

Image source: [http://mohammadmahoor.com/affectnet/]

Image source: [He et al., Mask R-CNN, ICCV 2017]

Anger	0.00007
Contempt	0.01314
Disgust	0.00021
Fear	0.00001
Happiness	0.43183
Neutral	0.55356
Sadness	0.00104
Surprise	0.00013

Image source: Microsoft Emotion Recognition
https://thenextweb.com/microsoft/2015/11/11/take-that-inside-out/
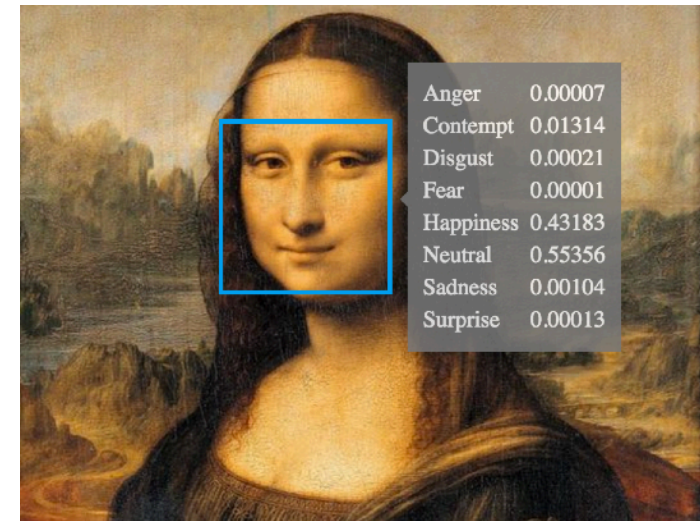
Recent Contributions: A Comparison

- Classification on Same dataset

- Performance get improved monthly

- Balance between recognition performance & model size

- This comparison is for general image classification problem, but CNNs are able to perform *Transfer Learning*, i.e. it can directly apply to emotion inference
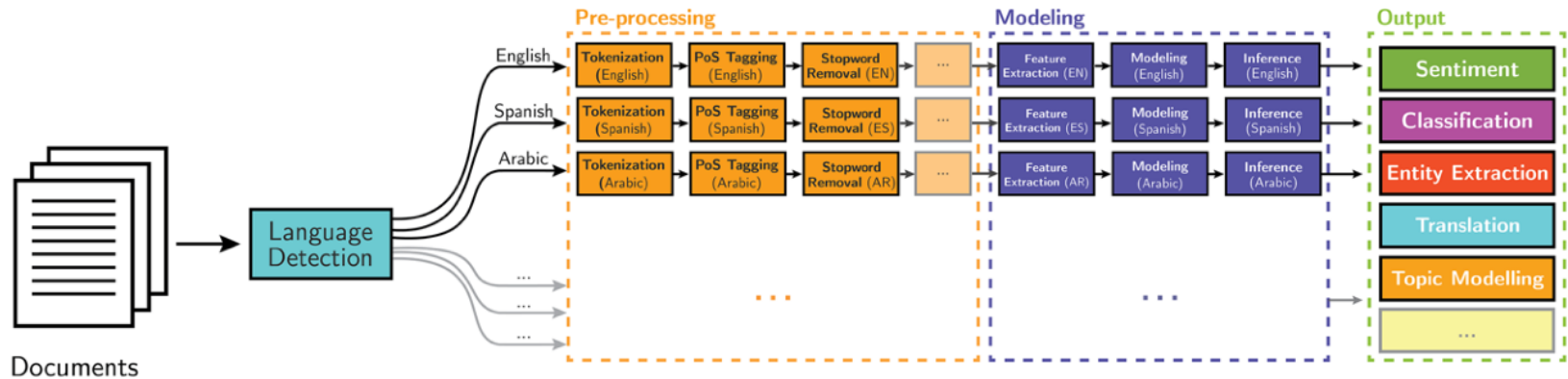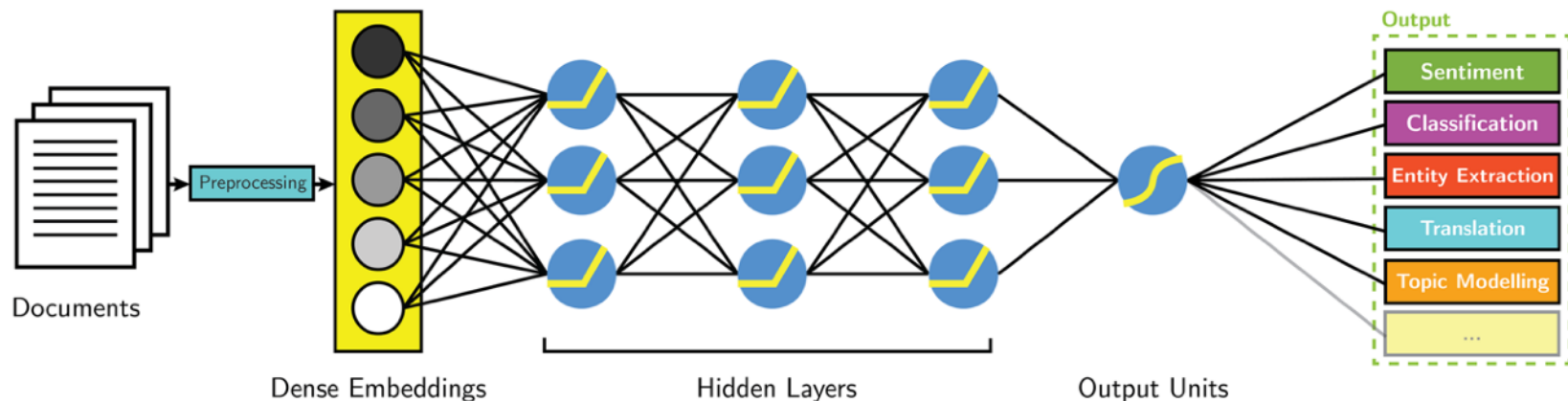
Error

80
65
50
35
20

Traditional

AlexNet   SqueezeNet   MobileNet

VGG   ShuffleNet

Xception

2010 2011 2012 2013 2014 2015 2016 2017
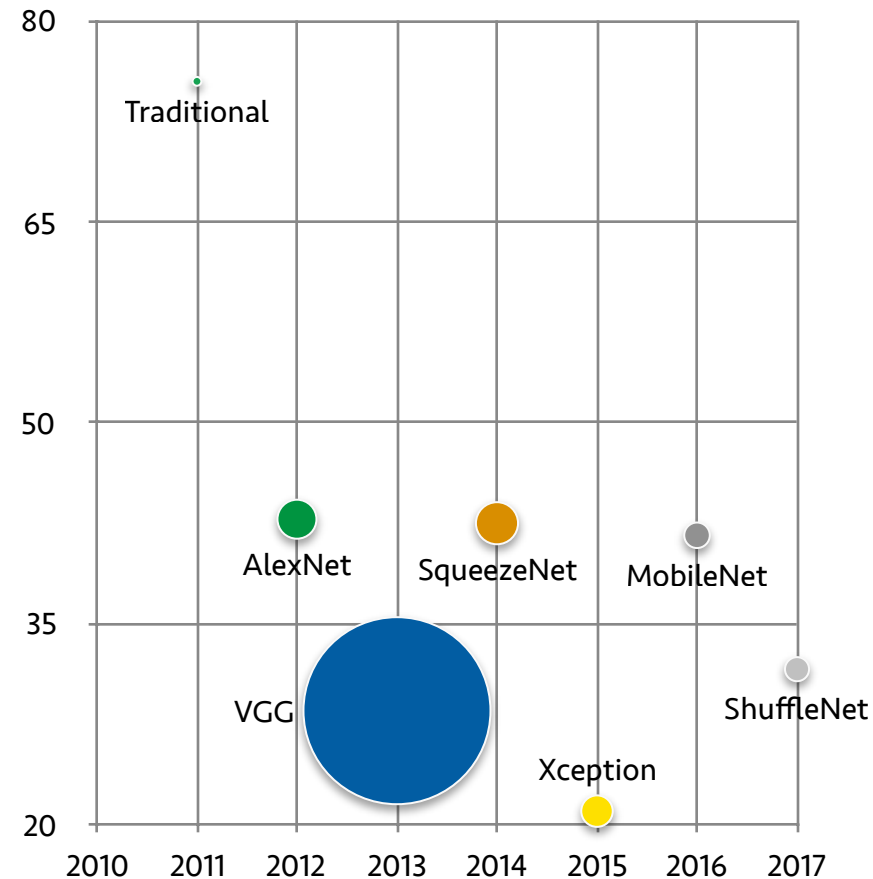
*The size of a circle represents computation complexity.

Speech Aspects

Voice-based Methods & NN Models

Classical Methods: Language Engineering

Neural Networks

Voice-based Methods

- Voice-based Emotion Inference can be decomposed to two steps:
  - Step 1: Speech to Text
  - Step 2: Text Sentiment Analysis[Rajalakshmi et al. 2017][Zhang et al. Feb. 2018]

- For directly analyses from tones: NN Models include this case
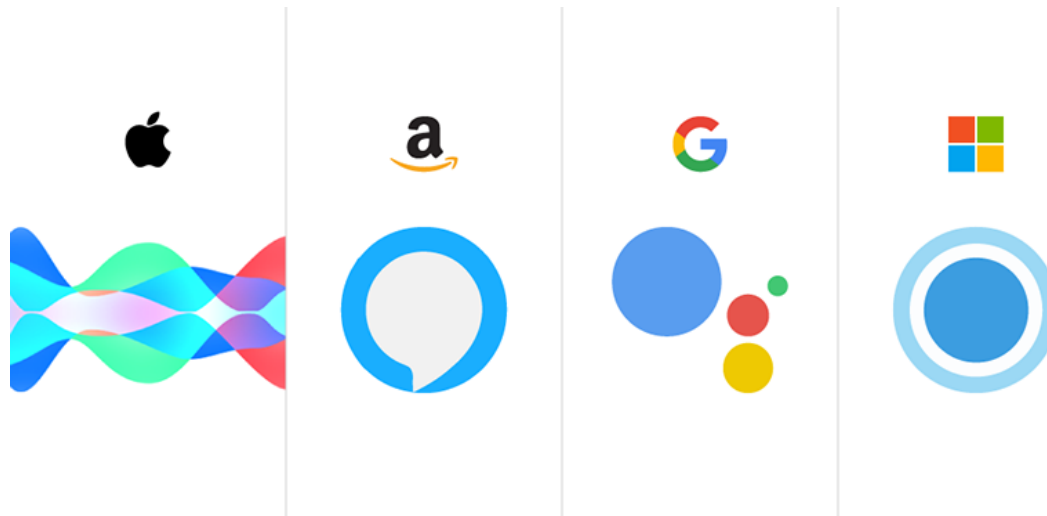
Result:

3%

Actual label for sample text:
negative

a fugitive on the run a bit like the incredible hulk tv series without the shirt ripping jimmy crosses the mob in an entirely contrived way and goes on the run and in an entirely contrived manner finds himself working at a catholic reform school have you noticed an oft used description in the last sentence entirely contrived is the answer let me repeat for the hard of thinking that this is an entirely contrived film where everything relies on coincidence another problem i had was the reform school run by the church it s far too compassionate and kind i m led to believe these type of <OOV> make alcatraz look like a country club i m saying this is a fact but when the head priest looks like the spitting image of donald <OOV> you do feel there s a large amount of sugar <OOV> going on br br to be honest despite the ridiculous plot twists etc wanted <OOV> t really a bad thriller though it s a terribly good one either i never really had the urge to switch it off no matter how contrived it became which is an under hand compliment to the movie

Result source: Keras examples

Commercial Success

- Voice Assistants have became consumer products
- Massive **human-level performance** language **APIs** provided by platforms

- FYI: STT is hard to optimize, only focus on sentiment analysis

Icon source: lazear@dribbble

Interaction Aspects

Interaction-based Methods

- Touch interaction and device motions (gyroscope, accelerometer, etc.) are the commonly used;

- Promising results presented by traditional feature engineering

- 99.9% papers only consider three emotions and only in a specific context.

- Be aware the reliability.

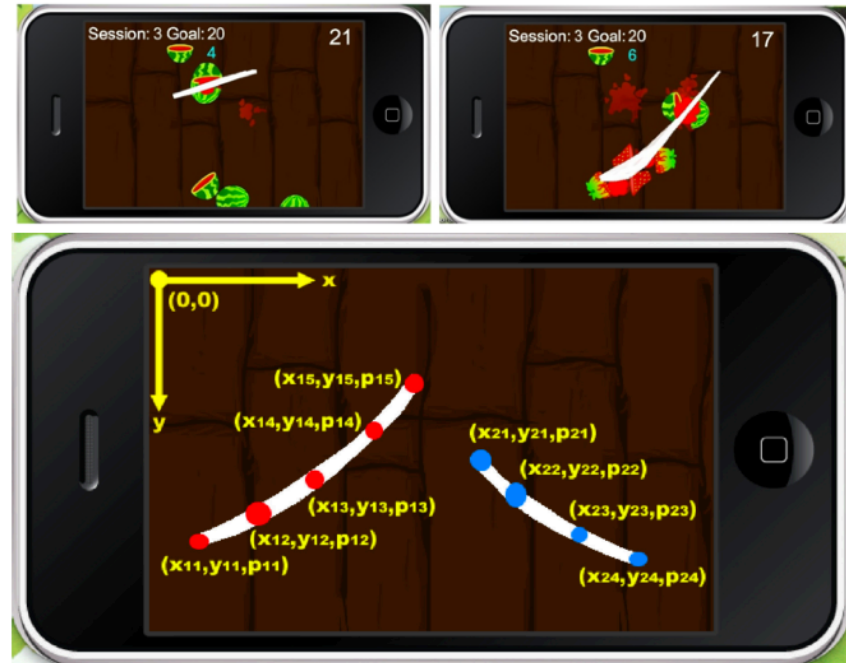| Features (>8) of Touch Interaction |
|---|
| Deviation in number of strikes |
| Deviation in number of taps |
| Mode of strike length |
| Average of strike length |
| …. |

Negative   Neutral   Positive

Image source: [Gao et al., What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?, TOCHI 2012]

Multimodal

Multimodal Method

- 99% papers based on sensors fusion method consider facial+speech data;

- All channels fusion is not discovered by researches.

Applications

Emotion-aware System

Video source: https://youtu.be/ldi1NCpe2Aw

Emotion-aware System

Sensing → Express ← User / Any UI

- Two typical application in mobile HCI:
    - Case 1: Spoken Dialogue System
        - Example: Siri

    - Case 2: Adaptive GUI
        - Example: Input Keyboard

    - General idea: dynamically adapting user interfaces based on user's emotions, research is rare
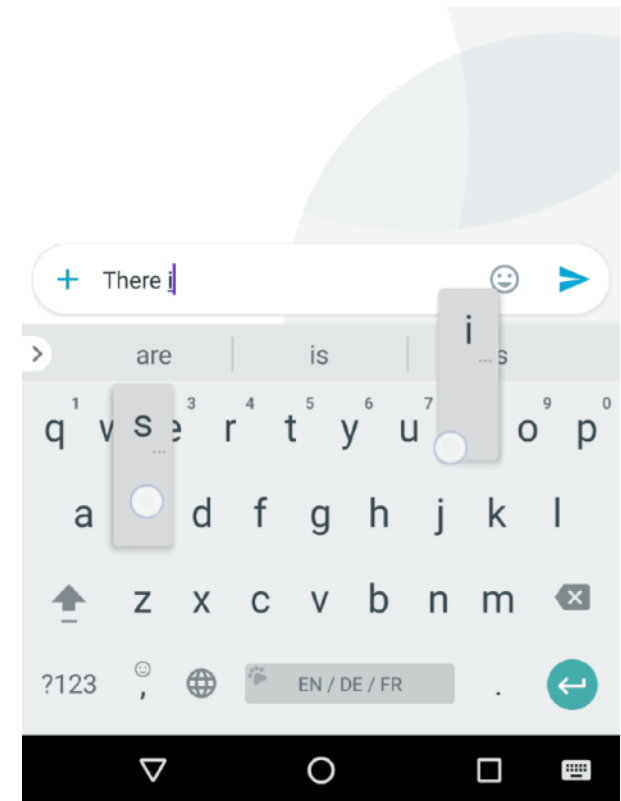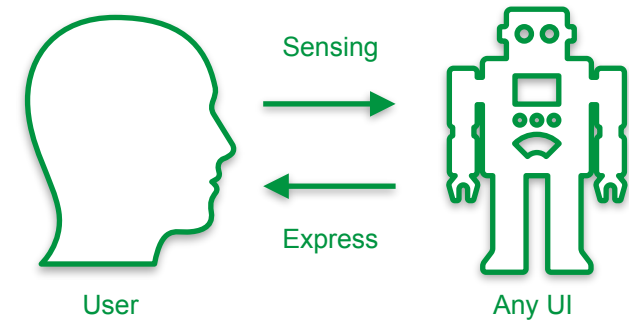

Image source: https://medium.com/google-design/human-centered-machine-learning-a770d10562cd

Challenges

Challenge: Continues Understanding

- Emotions are not just state;
- Emotions influences each other and transform to others continuously;

Image source: https://stanchew.wordpress.com/2012/04/23/a-map-of-human-emotions/

Challenge: Impermeable Emotions

- Impermeable emotions can not be labeled

    - e.g. *I am jealous of...;*

    - Some research defense this argument and claims impermeable emotions are trivial and not interested.
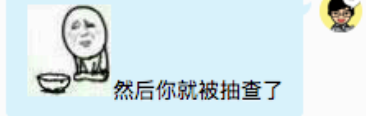

We were talking about buy a MacBook from Apple Store with educational discount… But this guy isn't a student anymore

My colleague bought a MacBook Pro with 20% education discount, he wasn't examined…

I'll do the same next time

Then you get caught, GG

Challenge: Impermeable Emotions II

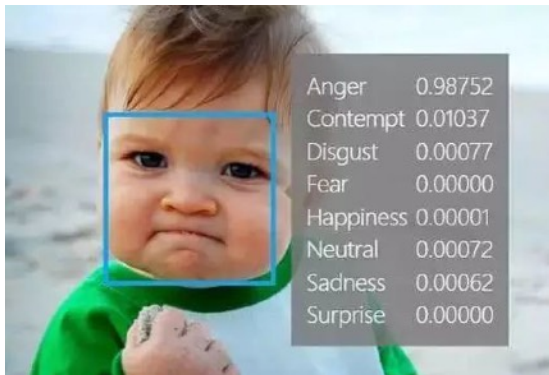- Impermeable emotions has culture difference [Markus et al. 1991]

"Sadness, please allow"

请悲允

Image source: Microsoft Emotion Recognition   Image source: Google Image Search

Summary

- Facial & voice channel is the most important channel over all channels, and Neural Networks are recently advances for emotion recognition (>100 different emotions);

- Typical applications of mobile emotion recognition consider Emotion-aware UI;

- Emotion-based HCI research is rare (Design principle, User Testing, etc.);

- Emotion Inference is a challenging problem & ***may not*** bring success.

References

- Mollahosseini et al. ***AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild***. IEEE Transactions on Affective Computing 2017.

- He et al. ***Mask R-CNN***, ICCV 2017.

- Howard et al. ***MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications***, 2017

- Howard et al. ***Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification***, *Detection and Segmentation*, 2018

- Zhang et al. ***Deep Learning for Sentiment Analysis: A Survey***, 2018

- Markus et al. ***Culture and self: Implications for Cognition, Emotion, and Motivation***, 1991