Multimodal Emotion Recognition in Response to Videos

통계학과 서석현

Introduction

- Affective self-reporting might be held in doubt, ~ … 자기 자신을 평가하는 것은 문제가 될 수 있다.
  -> 자기는 용감하다고 판단했지만 실제로는 겁에 질렸을 수 있다.

- multimedia 에 관련된 감정은 크게 3 가지로 연구되어 왔다.

1. Estimating emotions from multimedia content - Affective Video content representation and modeling (A. Hanjalic and L-Q. Xu, IEEE Trans Multimedia vol 7.)

2. Recognizing emotions induced by videos - Exploiting Facial Expressions for Affective Video Summerisation (H. Jobo, Proc ACM ~)

3. detect topical relevance or summarizing videos - Looking at the Viewer Analysing Facial Activity to detect personal highlights of multimedia Contents(H. Jobo, Multimedia Tools and applications)

- 논문에서는 EEG 신호와 Eye gaze 데이터를 사용해서 감정 인식을 했다.

- Irie et al은 토픽 모델링을 사용해서 영화의 음성 혹은 단어를 통해 감정을 인식하려고 했다.

감정 인식에 대한 대표적인 선행연구

- Using Noninvasive Wearable Computers to recognize Human Emotions from Physiological Signals (C.L. Lisette and F. Nasoz, Applied signal Processing vol 2004) -> 6개의 감정들을 동영상을 보고 Classification 진행함. 장점: 좋은 성능, 단점: 매우 감정적인 부분을 segment를 하여 보여줬다.

- Takahashi는 EEG를 통해 41.7의 정확도를 보이는 모델을 완성하였다. 하지만 feature level fusion 과 peripheral 한 신호는 성능 개선에 도움을 주지 못하였다.

- peripheral 시그널을 사용해 relevance vector machine을 사용한 Soleymani의 연구도 있었다.

- arousal, valence를 가지고 분류를 했더니 성능이 어느정도 좋아졌다.

- startle stimuli (random noise sounds)로 자극하고 분류를 하였더니 성능이 77.5로 좋아졌다.

- The Pupil as a Measure of Emotional Arousal and Autonomic Activation(M.M. Bradley, Psychophysiology vol 45) - 1000회 이상 인용

- Pupil size variation as an Indication of Affective Processing(T. partial, Human-computer studies, vol 59) - 600회 이상 인용

- Potential application은

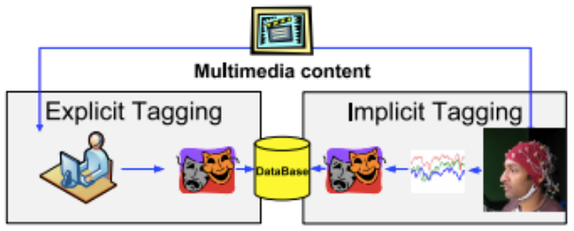

Fig. 1. Implicit affective tagging versus explicit tagging scenarios. The analysis of the bodily responses replaces the direct interaction between user and the computer. Therefore, users do not have to be distracted for tagging the content.

Material and Methods

- preliminary study에서는 2분 짜리 동영상을 155개 만들었다. 각 영상들에는 10개의 annotation들이 있었고 50명 이상의 피험자가 있었다.
- Emotion keyword(arousal, valence)에 대해서 9개 척도가 존재했다. 거기서 각각 많은 척도를 받은 14개 영상을 선택되었고 3개는 온라인에서(joy, happiness, disgust)와 3개는 기상일보 영상을 가져왔다.
- 영상 20개의 시간은 34.9 ~ 117초이며 M = 81.4, SD = 22.5초였다.
- facial video, audio, vocal expression, eye gaze, physicological signal이 녹화되었다.
- Tobii, Biosemi active 2 system
- ECG, EEG(32), galvanic skin response, respiration amplitude, skin temperature
- peripheral, vocal, facial X, EEG, pupillary response and gaze distance O
- 30명의 피험자 in Imperial College London, 17명의 여성 피험자, 13명의 남성 피험자, 19 ~40(m = 26.06, sd = 4.39), 다양한 배경
- 6명의 데이터는 품질이 좋지 못하여 24명으로 분석함. http://mahnob-db.eu

TABLE 1
The Video Clips and Their Sources

Code	Emotion Labels	Video clips sources
1	Act., Unp.	Hannibal
2	Act., Unp.	The Pianist
3	Med., Pls.	Mr. Bean's holiday
4	Act., Neu.	Ear worm (blip.tv)
5	Med., Neu.	Kill Bill VOL I
6	Med., Pls.	Love actually
7	Med., Pls.	Mr. Bean's holiday
8	Cal., Pls.	The thin red line
9	Med., Neu.	The shining
10	Med., Pls.	Love actually
11	Act., Unp.	The shining
12	Med., Unp.	Gangs of New York
13	Act., Unp.	Silent hill
14	Med., Unp.	The thin red line
15	Cal., Neu.	AccuWeather New York weather report (youtube.com)
16	Act., Unp.	American history X
17	Cal., Neu.	AccuWeather Detroit weather report (youtube.com)
18	Act., Pls.	Funny cats (youtube.com)
19	Cal., Neu.	AccuWeather Dallas weather report (youtube.com)
20	Act., Pls.	Funny (blip.tv)

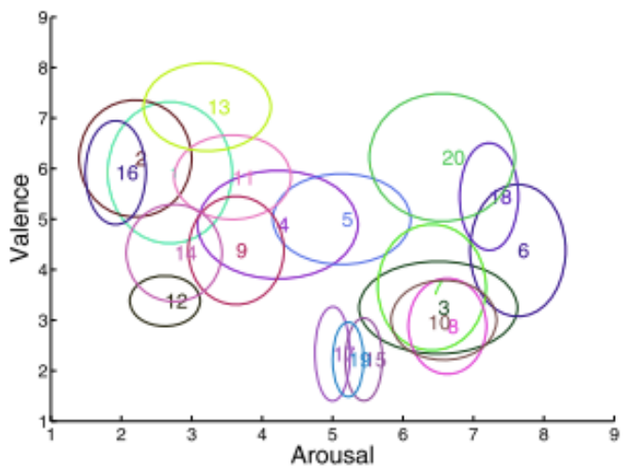

Fig. 3. Stimulus videos are shown in the valence-arousal plane. The center of the ellipses represents the mean arousal and valence and the horizontal and vertical radius represents the standard deviation of the online assessments. The clip codes are printed at the center of each ellipse.

Material and Methods

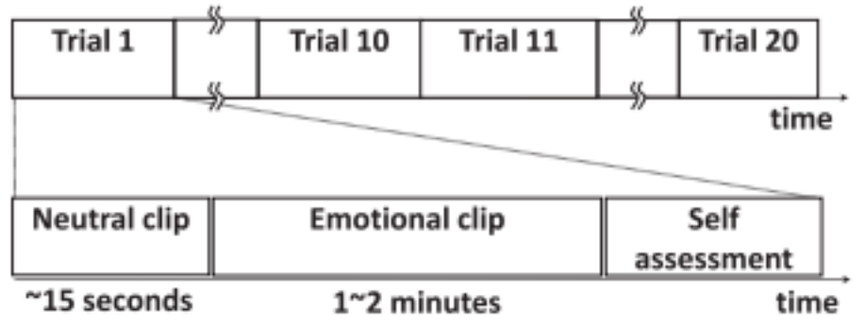

Fig. 5. Each trial started by a 15 s neutral clip and continued by playing one emotional clip. The self-assessment was done at the end of each trial. There were 20 trials in each session of the experiment.