

On modelling viewer sentiment of social media videos for attractiveness computing

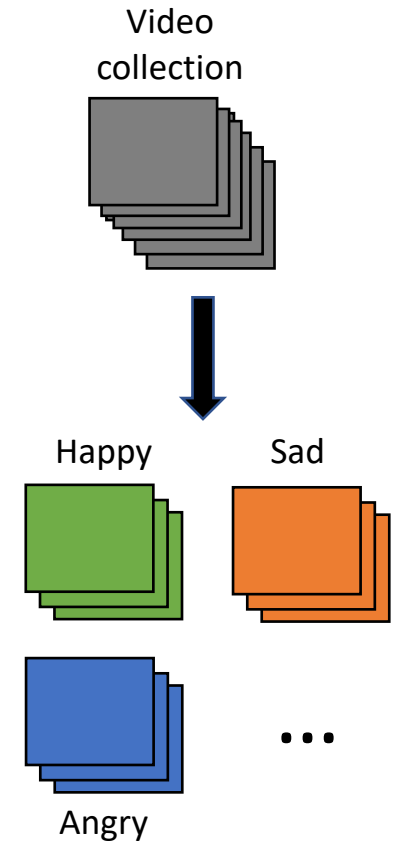
Marc A. Kastner, Shin'ichi Satoh
National Institute of Informatics

 @mkasu

 mkastner@nii.ac.jp

Viewer sentiment analysis of videos

- Try to find perceptual based clustering of scenes
 - Which might differ from genres
 - Content-based clustering also avoids annotation bias
- Goals
 - Find “funny scenes” “scary scenes” etc.
 - *Create something like Sentibank for videos*



Motivation

- Applications:
 - Video recommendations
 - Video retrieval
- Could be improved by being user-centric
 - Include viewer sentiment to include perception of a video
 - Two videos with the same genre might yield opposite sentiment
 - A “related video” recommendation should be *perceived* the same, not just have similar meta-data

Related research


- Visual Sentiment
 - SentiBank
 - Detectors for “funny cat” vs. “cute cat”
 - (Mostly) targeting images
 - Strong supervision
- Video emotion research
 - Does not analyze the viewer but emotion of somebody inside video
 - Not directly connected to sentiment: A prank video of somebody laughing might create an angry sentiment in viewer.

Problem: Sentiment annotations


- How to get annotations for viewer sentiment?
 - E.g. annotate scenes with “funny” “scary” “sad” ...
 - No sentiment annotations
 - Also not much emotion/sentiment research done with YouTube datasets yet
 - Annotation expensive
- Idea: Use user reactions for weak supervision, instead
 - Analyze user comments with text sentiment techniques

Weak supervision


- Use viewer comments to model viewer sentiment
 - E.g. sad comments => sad sentiment

 Love Humanity 8 hours ago
O God gives heaven to your father.


 2  REPLY

 zohaib 2 hours ago
Sorry for ur loss it's heartbreaking when u lose ur father something u can never ever get over peace 🙏 !!!


  REPLY

 Tim D 6 hours ago
I am sorry brother.....


 1  REPLY

 AkSevda 35Un1977 8 hours ago
R.i.p 🥺🙏🥺

 2  REPLY


 LoveToHearUSing 6 hours ago
Sending love and prayers from Germany



  REPLY


 goldgurme 8 hours ago
RiP doctor




 3  REPLY


- Happy comments => happy sentiment

 Lily Silva 1 month ago
Congrats you guys I love you 🎉🥳❤️❤️


 1   REPLY

 Kimberly Mirelez 1 month ago
I'm sooo happy for the both of you ! Thank you for sharing this amazing moment with us 💕💕


 1   REPLY

 Melanie Cifuentes 1 month ago
Omg YAY! Congratulations 🎉❤️


 1   REPLY

 Life with Loeras 1 month ago
The beginning made me tear up !!!!! 🥺 congrats !!!! 💕💕💕

 1   REPLY

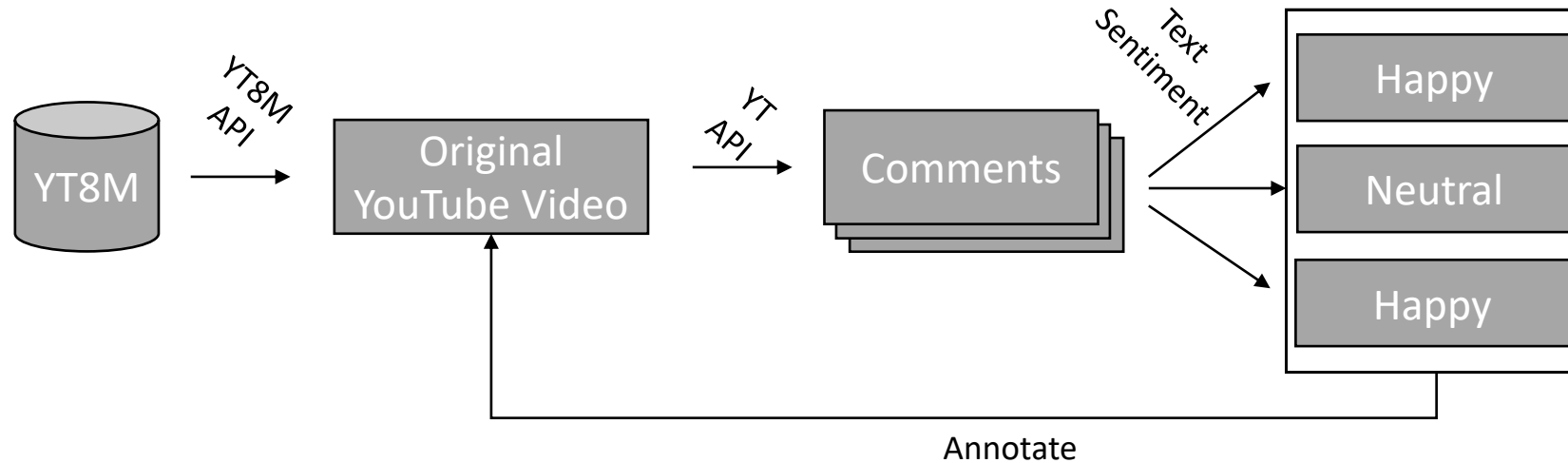
 Life with Loeras 1 month ago
I'm commenting while watching lol!!!

 1   REPLY

 Melanie Cifuentes 1 month ago
Omg YAY! Congratulations 🎉❤️

 1   REPLY

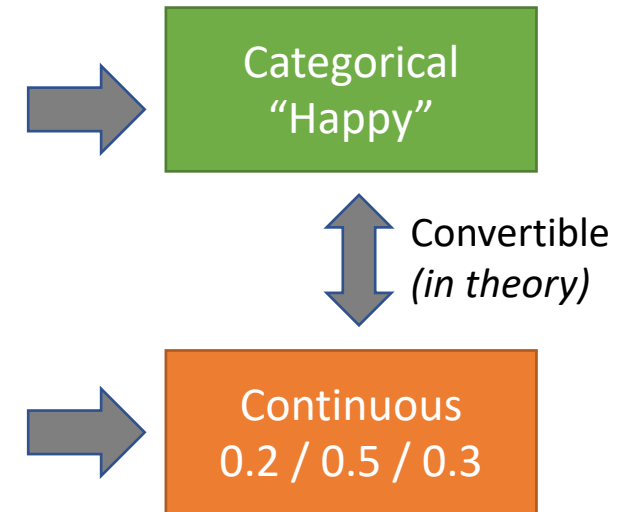
Weak supervision



- Crawl YouTube dataset
 - Get comments through API
 - Analyze text sentiment
 - Determine sentiment label for video

Text sentiment analysis

- Two dictionaries for word sentiment:
 - NRC Word-Emotion Association Lexicon [1] (EmoLex)
 - 14,182 English words with annotations for 10 classes
 - Emotions:
Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust
 - Sentiment:
Positive, Negative
 - NRC Valence, Arousal, Dominance Lexicon [2]
 - 20,007 English words with granular scores between [0,1]
- Both dictionaries provide machine-translated multi-language annotations

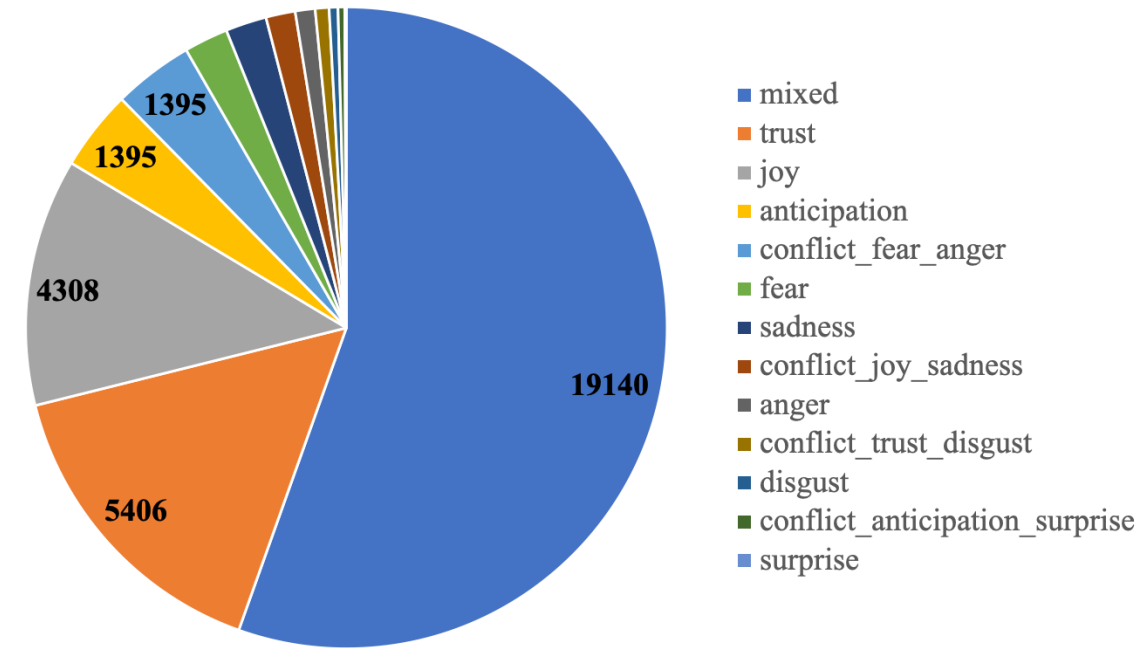


[1] Crowdsourcing a Word-Emotion Association Lexicon, S. M. Mohammad and P. Turney, Computational Intelligence, 29 (3), 436-465, 2013

[2] Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. S. M. Mohammad. ACL 2018.

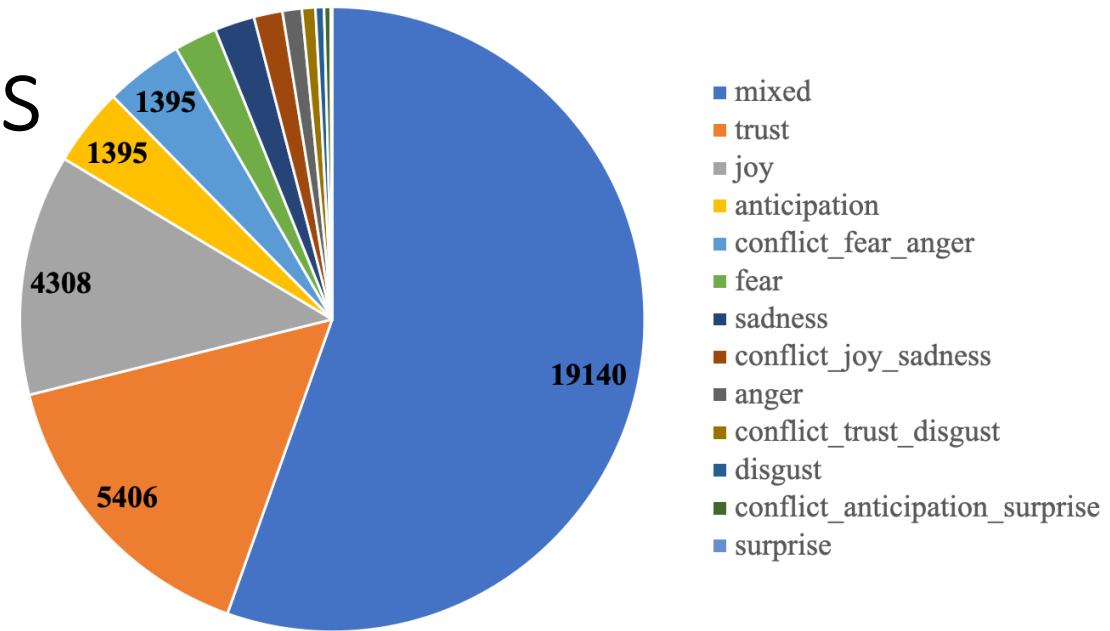
Dataset

- Crawled approx. 100 comments each for 34,518 YT videos
 - Analyzed all comments with word sentiment lexicons
 - Word-Emotion: *Majority vote*
 - VAD scores: *Average over all words/comments*
- Finding:
 - SNS data *very noisy*
 - For majority of videos no easy majority decision
 - But quite some videos actually can get a majority decision!



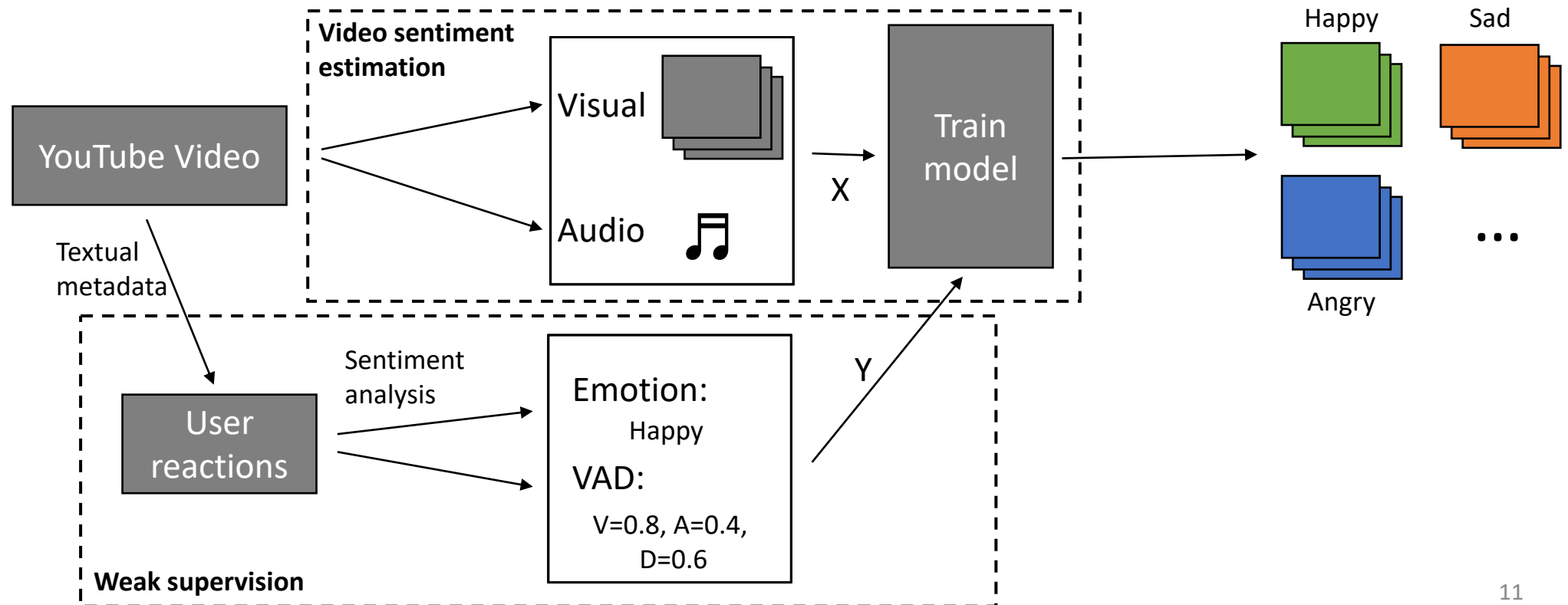
Adding new emotion labels

- “Mixed”:
 - Comments have no clear emotion attached
 - Three or more trends in comments make it hard to do a “majority decision”
- “Conflicting”:
 - There are opposite trends
 - For example, half of comments “sad” and half of comments “happy”
 - Detecting conflicting trends might be interesting for news videos
- For the evaluations
 - For now, ignore “mixed” emotion videos
 - Analyze videos with clear emotion or conflicting emotion



Model

- Train audio-visual features to classify the viewer sentiment annotations retrieved by weak supervision



Preliminary experiment

- Train Random Forest on video-level features
 - Visual: Inception-V3 pre-trained on ImageNet
 - Audio: VGG-inspired audio model
- Experiments
 - Regress V/A/D
 - Trained separately towards V-A-D scores in the interval of [0,100]
 - Classify emotion
 - 12 classes: Anger, Anticipation, Disgust, Fear, Joy, Sadness, Surprise, Trust, Conflicting_* (x4)
- Dataset
 - Training: 12,302 videos
 - Testing: 3,076 videos

Experiments: VAD (Left) / Emotion (Right)

(For interval [0, 100])

Valence	Features	MAE	Correlation
	Visual	3.19	0.57
	Audio	3.06	0.59
	Both	2.96	0.63

Arousal	Features	MAE	Correlation
	Visual	2.13	0.47
	Audio	2.06	0.52
	Both	2.04	0.54

Dominance	Features	MAE	Correlation
	Visual	2.06	0.30
	Audio	2.03	0.34
	Both	2.02	0.36

	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.45	0.48	0.38
Audio	0.44	0.51	0.43
Both	0.43	0.50	0.40

Emotion

Considerations

- Weak supervised labels
 - Need to be evaluated
 - Compare to small-scale crowd-sourced annotation?
- VAD
 - Seems to work quite well actually despite naïve model
- Emotion
 - Very imbalanced and noisy

Currently running... (Future work)

- Improve model
 - Fuse with features from visual sentiment analysis + audio mood
 - Loss function idea: Use a triplet loss
 - In triplet loss, usually the idea is to give “easily mistaken” negative samples
 - Use training mechanism focusing on “conflicting” labels for this

Currently investigating... (Future work)

- Multi-modal weak supervision for violence detection[1]
 - They use weak supervision of large video data from YouTube
 - From: Video-level annotations -> 6 violence classes
 - To: Frame-level detection of violence
 - A similar approach could work for emotion classes
- Try Bayesian Network[2]

[1] Wu et al. Not only Look, but also Listen: Learning Multimodal Violence Detection under Weak Supervision. ECCV 2020

[2] Matin et al. Hey Human, If your Facial Emotions are Uncertain, You Should Use Bayesian Neural Networks! ECCV 2020.

Conclusion

- Creating an audiovisual -> video sentiment model for SNS content
 - Using user comments as weak supervision
- Performance of VAD regression shows promising results even for naïve model
 - More data and better features should give good performance
- Weak supervision might need crowd-sourced evaluation

 @mkasu

 mkastner@nii.ac.jp