A preliminary study on viewer sentiment analysis of social media videos mkastner@nii.ac.jp Marc A. Kastner, Shin'ichi Satoh

= Happy video?

= Sad video?

Grieving comments

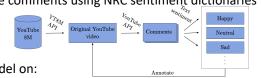
Motivation

- Purpose: Find scenes which are funny, scary, sad ...
- No existing datasets!
- Can we use user comments to cluster sentiment of videos?

Annotation expensive.

Approach

- Using videos from SNS (YouTube):
- Crawl videos + their top-n comments
- Analyze comments using NRC sentiment dictionaries



- Train model on:
- X = [Visual features + Audio features]
- Y = generated Emotion / VAD annotation

Next steps

Improve features

Dataset composition

Categories of videos

- RGB / Audio currently simple average over all frames (Switch to RNN model)
- Include audio sentiment, music mood, etc.
- Train separate models for different categories
 - Can we find per-community sentiment models?

Beauty & Fitnes:

Food & Drink

Pets & Animals

Internet & Telecor

From comments to sentiment

- The comments are direct reactions to comments
- Sentiment analysis of comments helps understanding videos
- Sentiment analysis to generate labels (majority decision) Emotion = $\{$ sad, **happy**, ... $\}$ VAD = $\{$ 0.1, 0.5, 0.3 $\}$

Experiments

- Dataset: 17,112 videos with generated Emotion/VAD from their top-100 comments
- Train separate models for each

	Val	ence	Aro	usal	Domi	nance
Features	MAE	Corr.	MAE	Corr.	MAE	Corr.
Visual	2.99	0.47	2.00	0.51	1.98	0.32
Audio	2.83	0.54	1.99	0.51	1.95	0.36
Combined	2.84	0.55	1.95	0.55	1.93	0.38

Results

Emotion

Generated emotion distribution

≡ iov

conflict

sadness

anger

■ disgust

surprise

■ fear

anticipation

- Works, but not enough data Table 2: Results for emotion estimation. for some emotions
- Dataset imbalanced

Features	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.30	0.39	0.28
Audio	0.36	0.41	0.34
Combined	0.33	0.41	0.31

Relationship Emotion <> VAD

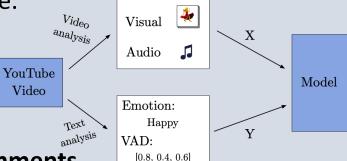
Our dataset

Used datasets

- [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
- YouTube video dataset:
- [3] YouTube- 8M: A Large-Scale Video Classification Benchmark. S. Abu-El-Haija et al., arXiv, p. 1609.08675v1 (2016).

Motivation

- Purpose: Find scenes which are funny, scary, sad ...
- Annotation expensive. No existing datasets!



Can we use user comments to cluster sentiment of videos?

From comments to sentiment

The comments are direct reactions to comments

 Sentiment analysis of comments helps understanding videos

 Sentiment analysis to generate labels (majority decision)

Emotion = $\{$ sad, **happy**, ... $\}$ VAD = $\{$ 0.1, 0.5, 0.3 $\}$

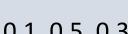
= Happy video?



Joyful comments







A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner. Shin'ichi Satoh

mkastner@nii.ac.jp

NI

Joyful comments

Grieving comments

= Happy video?

= Sad video?

Motivation

- Purpose: Find scenes which are funny, scary, sad ...
- Annotation expensive.
 No existing datasets!
 - YouTube Video

 Test

 Manualysis

 Post

 Audio

 J

 X

 M

 M

 M

 M

 M

 VAD:

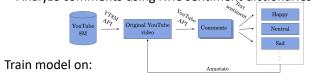
 [0.8, 0.4, 0.6]
- Can we use user comments to cluster sentiment of videos?

From comments to sentiment

- The comments are direct reactions to comments
- Sentiment analysis of comments helps understanding videos
- Sentiment analysis to generate labels (majority decision)
 Emotion = {sad, happy, ...} VAD = { 0.1, 0.5, 0.3}

<u>Approach</u>

- · Using videos from SNS (YouTube):
- Crawl videos + their top-n comments
- Analyze comments using NRC sentiment dictionaries



- X = [Visual features + Audio features]
- Y = generated Emotion / VAD annotation

Next steps

- Improve features
 - RGB / Audio currently simple average over all frames (Switch to RNN model)
 - Include audio sentiment, music mood, etc.
- Train separate models for different categories
- Can we find per-community sentiment models?

Experiments

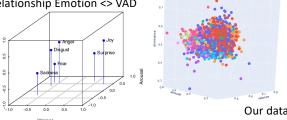
- Dataset: 17,112 videos with generated
 Emotion/VAD from their top-100 comments
- Train separate models for each

	Vale	ence	Aro	usal	Domi	nance
Features	MAE	Corr.	MAE	Corr.	MAE	Corr
Visual	2.99	0.47	2.00	0.51	1.98	0.32
Audio	2.83	0.54	1.99	0.51	1.95	0.36
Combined	2.84	0.55	1.95	0.55	1.93	0.38

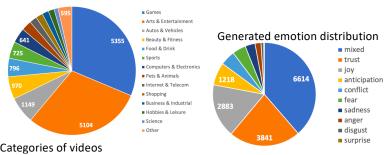
- Results
- Works, but not enough data Table 2: Reference for some emotions
- Dataset imbalanced

,,,	uatarab	ie 2. Hesuits ioi	emonon esun	nation.
	Features	Avg. Precision	Avg. Recall	Avg. F1 Score
	Visual	0.30	0.39	0.28
	Audio	0.36	0.41	0.34
	Combined	0.33	0.41	0.31

Emotion Relationship Emotion <> VAD



Dataset composition

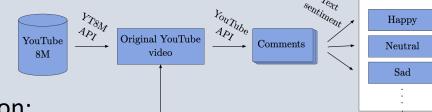


Used datasets

- Sentiment dictionaries
- [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
- YouTube video dataset:
- [3] YouTube- 8M: A Large-Scale Video Classification Benchmark. S. Abu-El-Haija et al., arXiv, p. 1609.08675v1 (2016).

<u>Approach</u>

- Using videos from SNS (YouTube):
 - Crawl videos + their top-n comments
 - Analyze comments using NRC sentiment dictionarie



- Train model on:
 - X = [Visual features + Audio features]
 - Y = generated Emotion / VAD annotation

Experiments

- Dataset: 17,112 videos with generated
 Emotion/VAD from their top-100 comments
 - Train separate models for each

Table 1: Results for VAD estimation.						
	Vale	ence	Aro	usal	Domi	nance
Features	MAE	Corr.	MAE	Corr.	MAE	Corr.
Visual Audio Combined	2.99 2.83 2.84	0.47 0.54 0.55	2.00 1.99 1.95	0.51 0.51 0.55	1.98 1.95 1.93	0.32 0.36 0.38

- Results
 - Works, but not enough data for some emotions
 - Dataset imbalanced

5	Table 2: Results for emotion estimation.						
d	Features	Avg. Precision	Avg. Recall	Avg. F1 Score			
u	Visual	0.30	0.39	0.28			
	Audio	0.36	0.41	0.34			
	Combined	0.33	0.41	0.31			

A preliminary study on viewer sentiment analysis of social media videos Marc A. Kastner. Shin'ichi Satoh mkastner@nii.ac.jp

NII

Joyful comments

Grieving comments

= Happy video?

= Sad video?

<u>Motivation</u>

- Purpose: Find scenes which are funny, scary, sad ...
- Annotation expensive.

 No existing datasets!

 Video analysis Audio Audio Model

 Visual Audio Model

 X Audio Model

 Emotion:
- Can we use user comments to cluster sentiment of videos?

From comments to sentiment

- The comments are direct reactions to comments
- Sentiment analysis of comments helps understanding videos
- Sentiment analysis to generate labels (majority decision)
 Emotion = {sad, happy, ...} VAD = { 0.1, 0.5, 0.3}

<u>Approach</u>

- Using videos from SNS (YouTube):
- Crawl videos + their top-n comments
- Analyze comments using NRC sentiment dictionaries



- Train model on:
- X = [Visual features + Audio features]
- Y = generated Emotion / VAD annotation

Experiments

- Dataset: 17,112 videos with generated Emotion/VAD from their top-100 comments
- Train separate models for each

Table 1. Results for VIID estimation.						
	Vale	ence	Aro	usal	Domi	nance
Features	MAE	Corr.	MAE	Corr.	MAE	Corr.
Visual	2.99	0.47	2.00	0.51	1.98	0.32
Audio	2.83	0.54	1.99	0.51	1.95	0.36
Combined	2.84	0.55	1.95	0.55	1.93	0.38

Results

disgustsurprise

- Works, but not enough for some emotions
- Dataset imbalanced

511	uata rable 2: Results for emotion estimation.						
	Features	Avg. Precision	Avg. Recall	Avg. F1 Score			
	Visual	0.30	0.39	0.28			
	Audio	0.36	0.41	0.34			
	Combined	0.33	0.41	0.31			

Next steps

Improve features

Dataset composition

Categories of videos

- RGB / Audio currently simple average over all frames (Switch to RNN model)
- Include audio sentiment, music mood, etc.
- Train separate models for different categories
 - Can we find per-community sentiment models?

Emotion Relationship Emotion <> VAD | Supprise | Suppr

| Sames | Arts & Entertainment | Autos & Vehicles | Beauty & Fitness | Food & Drink | Sports | Computers & Electronics | Pets & Animals | Internet & Telecom | Shopping | Business & Industrial | Hobbies & Leisure | Sediness | Sadness | S

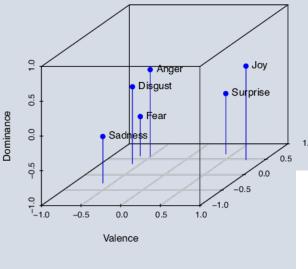
Used datasets

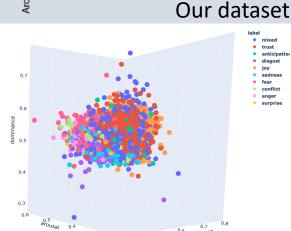
- Continuent distinueries
- [1] Crowdsourcing a Word-Emotion Association Lexicon, S. M. Mohammad and P. Turney, Computational Intelligence, 29 (3), 436-465, 2013
- [2] Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. S. M. Mohammad. ACL 2018
- YouTube video dataset:
- [3] YouTube- 8M: A Large-Scale Video Classification Benchmark. S. Abu-El-Haija et al., arXiv, p. 1609.08675v1 (2016).

Next steps

- Improve features
 - RGB / Audio currently simple average over all frames (Switch to RNN model)
 - Include audio sentiment, music mood, etc.
- Train separate models for different categories
 - Can we find per-community sentiment models?

Emotion Relationship Emotion <> VAD





A preliminary study on viewer sentiment analysis of social media videos mkastner@nii.ac.jp Marc A. Kastner. Shin'ichi Satoh



Joyful comments

Grieving comments

= Happy video?

= Sad video?

Motivation

- No existing datasets!
- Can we use user comments to cluster sentiment of videos?

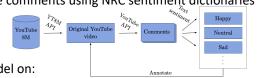
Purpose: Find scenes which are funny, scary, sad ... Annotation expensive.

From comments to sentiment

- The comments are direct reactions to comments
- Sentiment analysis of comments helps understanding videos
- Sentiment analysis to generate labels (majority decision) Emotion = $\{\text{sad}, \text{happy}, ...\}$ VAD = $\{0.1, 0.5, 0.3\}$

Approach

- Using videos from SNS (YouTube):
- · Crawl videos + their top-n comments
- Analyze comments using NRC sentiment dictionaries



- Train model on:
- X = [Visual features + Audio features]
- Y = generated Emotion / VAD annotation

Experiments

- Dataset: 17,112 videos with generated Emotion/VAD from their top-100 comments
- Train separate models for each

Table 1: Results for VAD estimation.						
	Vale	ence	Aro	usal	Domi	nance
Features	MAE	Corr.	MAE	Corr.	MAE	Corr.
Visual	2.99	0.47	2.00	0.51	1.98	0.32
Audio	2.83	0.54	1.99	0.51	1.95	0.36
Combined	2.84	0.55	1.95	0.55	1.93	0.38

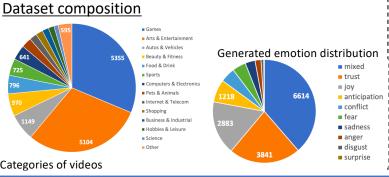
- Results
- Works, but not enough dataTa for some emotions
- Dataset imbalanced

ζII	Uala Table 2: Results for emotion estimation.						
	Features	Avg. Precision	Avg. Recall	Avg. F1 Score			
	Visual	0.30	0.39	0.28			
	Audio	0.36	0.41	0.34			
	Combined	0.33	0.41	0.31			

Next steps

- Improve features
 - RGB / Audio currently simple average over all frames (Switch to RNN model)
 - Include audio sentiment, music mood, etc.
- Train separate models for different categories
 - Can we find per-community sentiment models?

Emotion Relationship Emotion <> VAD Our dataset

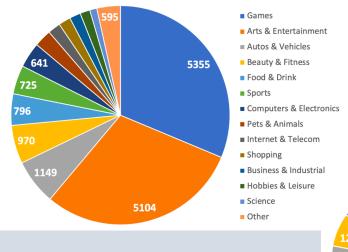


Used datasets

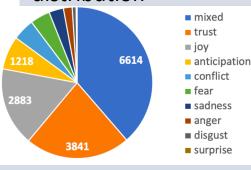
- [1] Crowdsourcing a Word-Emotion Association Lexicon, S. M. Mohammad and P. Turney, Computational Intelligence, 29 (3),
- [2] Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. S. M. Mohammad. ACL
- YouTube video dataset:
- [3] YouTube- 8M: A Large-Scale Video Classification Benchmark. S. Abu-El-Haija et al., arXiv, p. 1609.08675v1 (2016).

Dataset composition

Categories of videos



Generated emotion distribution



Used datasets

- Sentiment dictionaries
- [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.
- YouTube video dataset:
 - [3] YouTube- 8M: A Large-Scale Video Classification Benchmark. S. Abu-El-Haija et al., arXiv, p. 1609.08675v1 (2016).