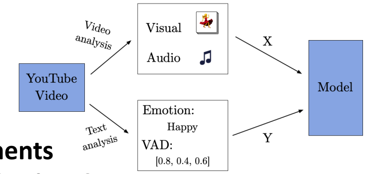
A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner, Shin'ichi Satoh
mkastner@nii.ac.jp

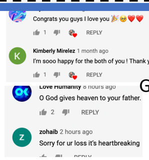
Motivation

- 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



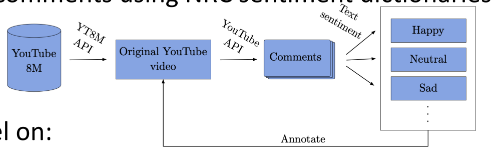
Joyful comments = Happy video?

Grieving comments = Sad video?

- 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}

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.

	Valence		Arousal		Dominance	
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 for some emotions
  - Dataset imbalanced

Table 2: Results for emotion estimation.

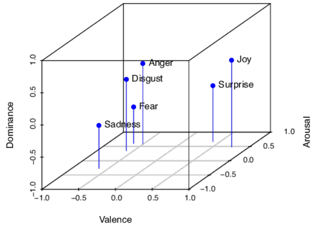
Features	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.30	0.39	0.28
Audio	**0.36**	**0.41**	**0.34**
Combined	0.33	**0.41**	0.31

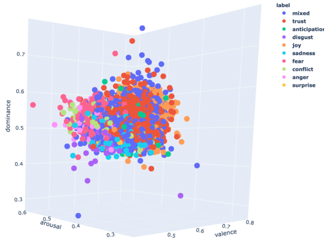
Next steps

- 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

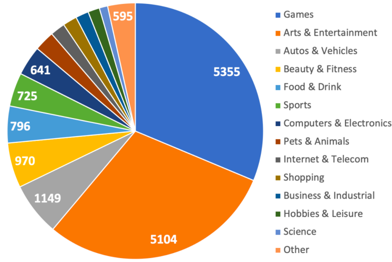
Relationship Emotion <> VAD





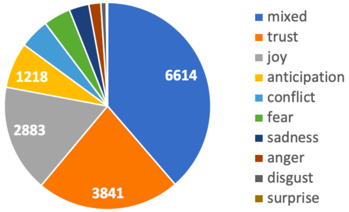
Our dataset

Dataset composition



Categories of videos

Generated emotion distribution

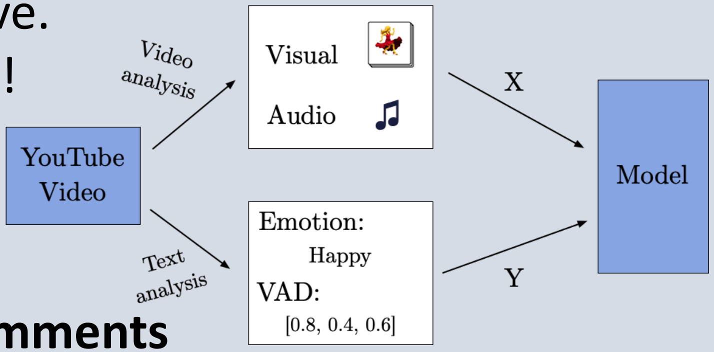


Used datasets

- 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).

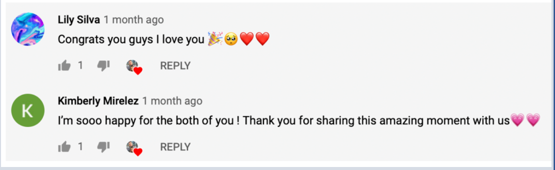
Motivation

- 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

Joyful comments = Happy video?



- The comments are direct reactions to comments
  - Sentiment analysis of comments helps understanding videos

Grieving comments = Sad video?

- Sentiment analysis to generate labels (majority decision)
  Emotion = {sad, **happy**, …}   VAD = { 0.1, 0.5, 0.3}

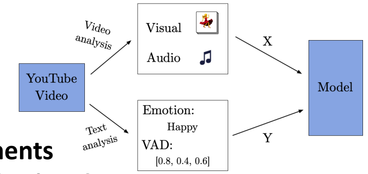
A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner, Shin'ichi Satoh
mkastner@nii.ac.jp

NII

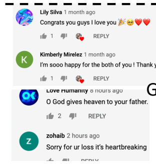
Motivation

- 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



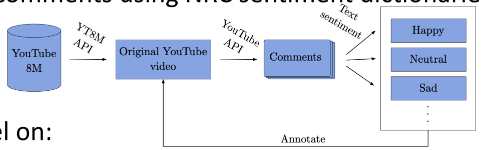
Joyful comments = Happy video?

Grieving comments = Sad video?

- 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}

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.

	Valence		Arousal		Dominance	
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 for some emotions
  - Dataset imbalanced

Table 2: Results for emotion estimation.

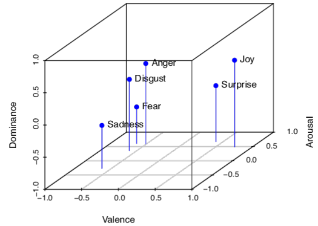
Features	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.30	0.39	0.28
Audio	**0.36**	**0.41**	**0.34**
Combined	0.33	**0.41**	0.31

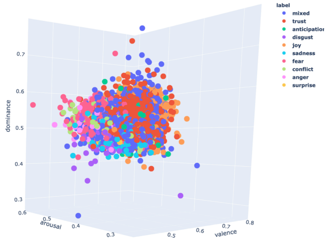
Next steps

- 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

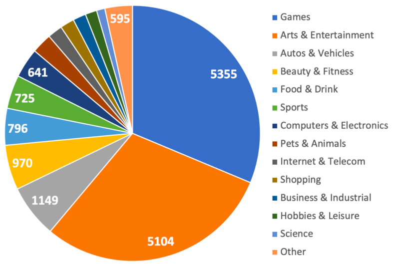
Relationship Emotion <> VAD





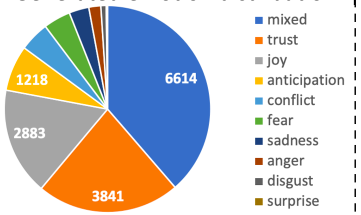
Our dataset

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).

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

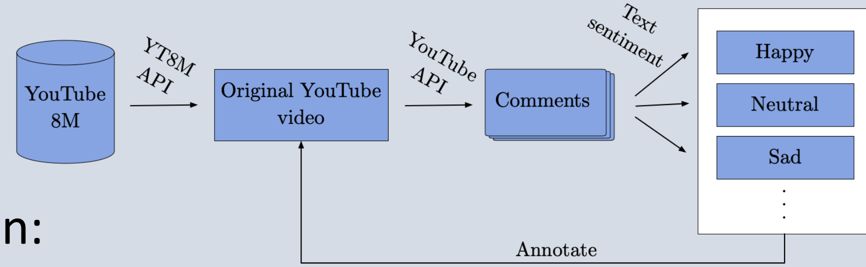
Table 1: Results for VAD estimation.

	Valence		Arousal		Dominance	
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 for some emotions
  - Dataset imbalanced

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



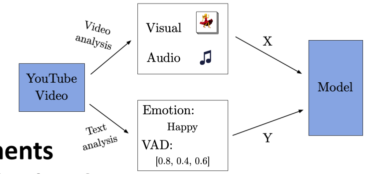
A preliminary study on viewer sentiment analysis of social media videos

Marc A. Kastner, Shin'ichi Satoh
mkastner@nii.ac.jp

NII

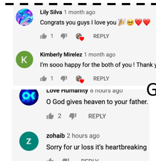
Motivation

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



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

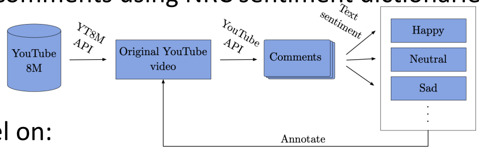
From comments to sentiment



- 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}

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

Table 1: Results for VAD estimation.

	Valence		Arousal		Dominance	
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**

Table 2: Results for emotion estimation.

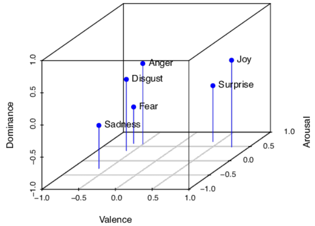
Features	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.30	0.39	0.28
Audio	**0.36**	**0.41**	**0.34**
Combined	0.33	**0.41**	0.31

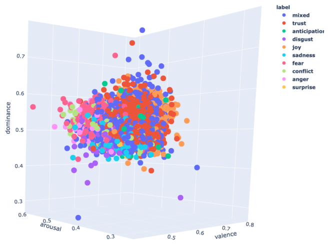
Next steps

- 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

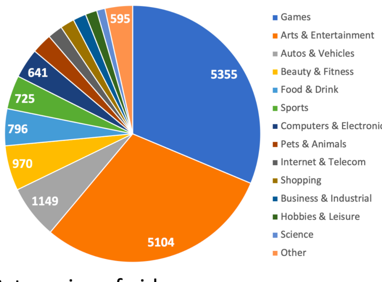
Relationship Emotion <> VAD





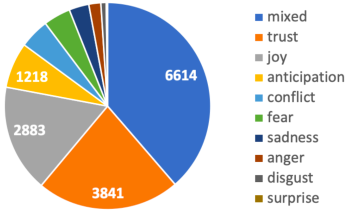
Our dataset

Dataset composition



Categories of videos

Generated emotion distribution



Used datasets

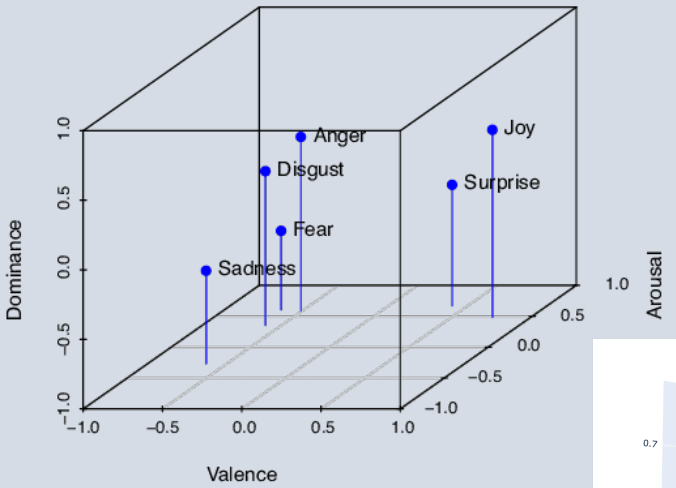
- Sentiment dictionaries
  - [1] Crowdsourcing a Word-Emotion Association Lexicon, S. M. Mohammad and P. Turney, Computational Intelligence, 29 (3), 436-465, 2013
  - [2] Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. S. M. Mohammad. ACL 2018.
- YouTube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

Next steps

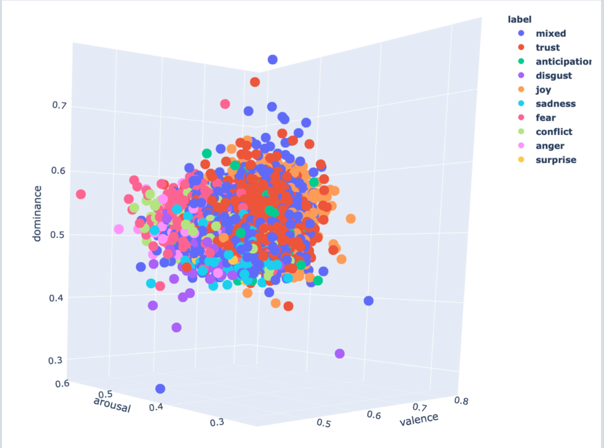
- Improve features
  - RGB / Audio currently simple average over all frames (Switch to RNN model)
  - Include audio sentiment, music mood, etc.
- Train separate models for different categories
  - ***Can we find per-community sentiment models?***

Emotion

Relationship Emotion <> VAD



Our dataset

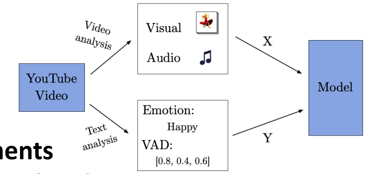


A preliminary study on viewer sentiment analysis of social media videos

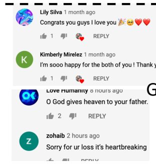
Marc A. Kastner, Shin'ichi Satoh
mkastner@nii.ac.jp
NII

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

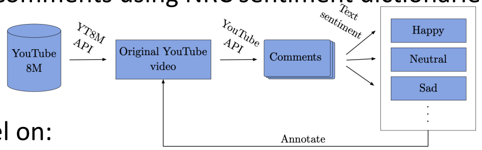


Joyful comments = Happy video?
Grieving comments = Sad video?

- 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}

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.

	Valence		Arousal		Dominance	
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 for some emotions
  - Dataset imbalanced

Table 2: Results for emotion estimation.

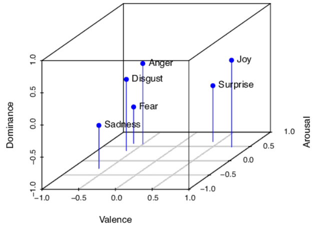
Features	Avg. Precision	Avg. Recall	Avg. F1 Score
Visual	0.30	0.39	0.28
Audio	**0.36**	**0.41**	**0.34**
Combined	0.33	**0.41**	0.31

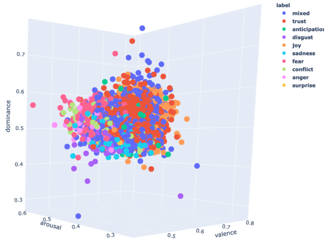
Next steps

- 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

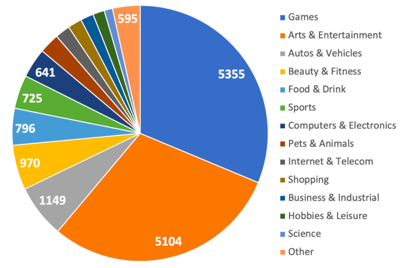
Relationship Emotion <> VAD





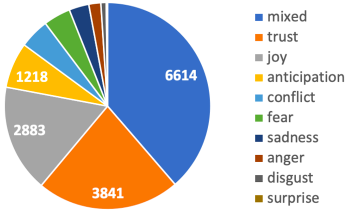
Our dataset

Dataset composition



Categories of videos

Generated emotion distribution

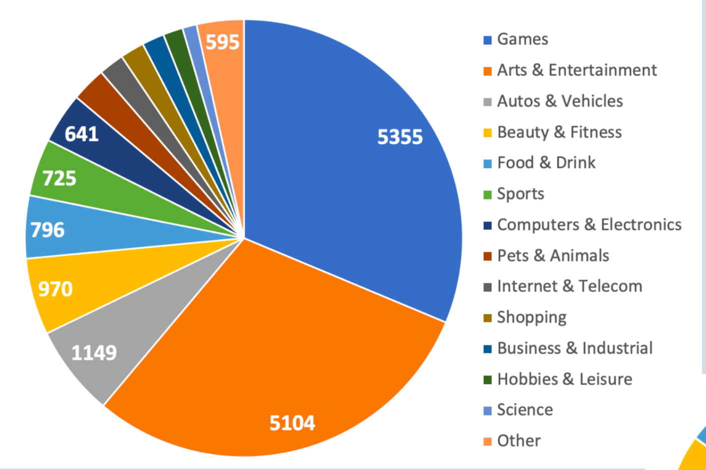


Used datasets

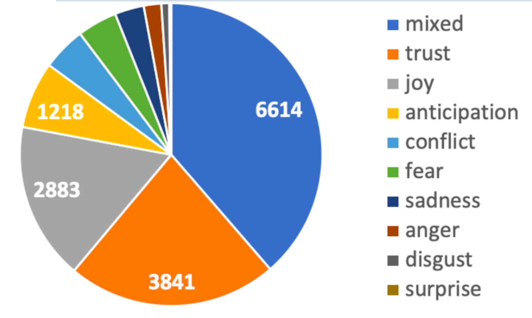
- Sentiment dictionaries
  - [1] Crowdsourcing a Word-Emotion Association Lexicon, S. M. Mohammad and P. Turney, Computational Intelligence, 29 (3), 436-465, 2013
  - [2] Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. S. M. Mohammad. ACL 2018.
- YouTube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

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).