

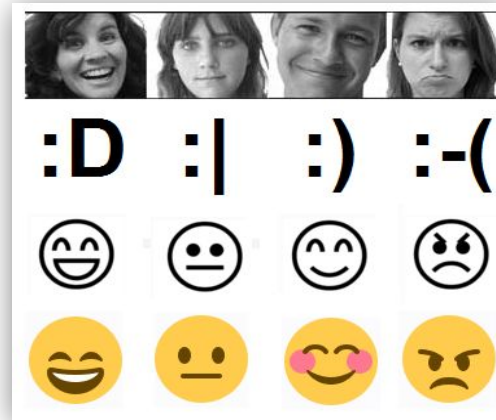
EmoTag - Towards an Emotion-Based Analysis of Emojis

Abu Awal Md Shoeb, Shahab Raji, and Gerard de Melo
Rutgers University

September 03, 2019, Varna, Bulgaria

Emojis are Ubiquitous

- A study found that half of social media text contains emojis (as of 2015)
- The same parts of the brain are activated as when we look at a real human face
- Oxford Dictionaries named “Face With Tears of Joy” 😂 its 2015 Word of the year



<http://instagram-engineering.tumblr.com/post/117889701472/emojineering-part-1-machine-learning-for-emoji>

Emoticons in mind: An event-related potential study by Churches O, Nicholls M, Thiessen M, Kohler M, Keage H (2014)

Goal: Emoji-based Lexical Resources

Problem:

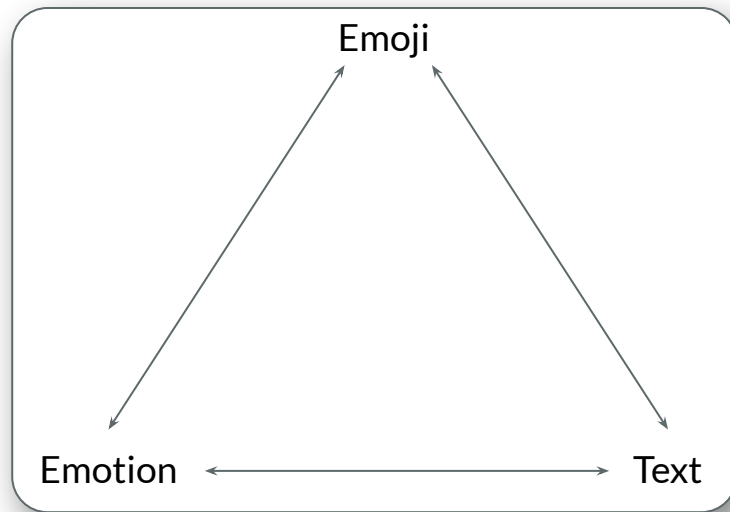
- Standard word embeddings are not interpretable
- Capture relationships among words only
- No relationships between emotion and words

What is missing:

- Interpretable Word Vectors based on emojis
- No lexicon for emoji-emotions yet

Our Approach:

- Use emoji to derive features/emotions for arbitrary words



EmoTag

Data Acquisition & Lexicons

Approach: Web Crawling

- Collected ~20M tweets over a period of 1 year
- 100 tweets per day for each of 620 most frequently used emoji
- Every single tweet contains at least one emoji

Data Cleansing

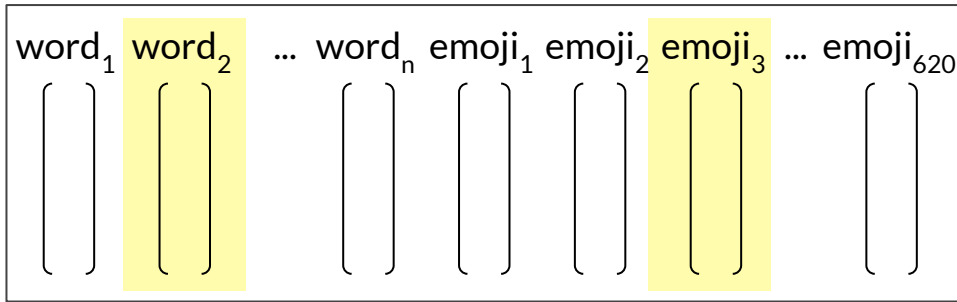
- No more than 5 tweets from an individual user

Each tweet contains

- tweet-id, text, username, date, retweets, favorites, geo-location, emoji, hashtags

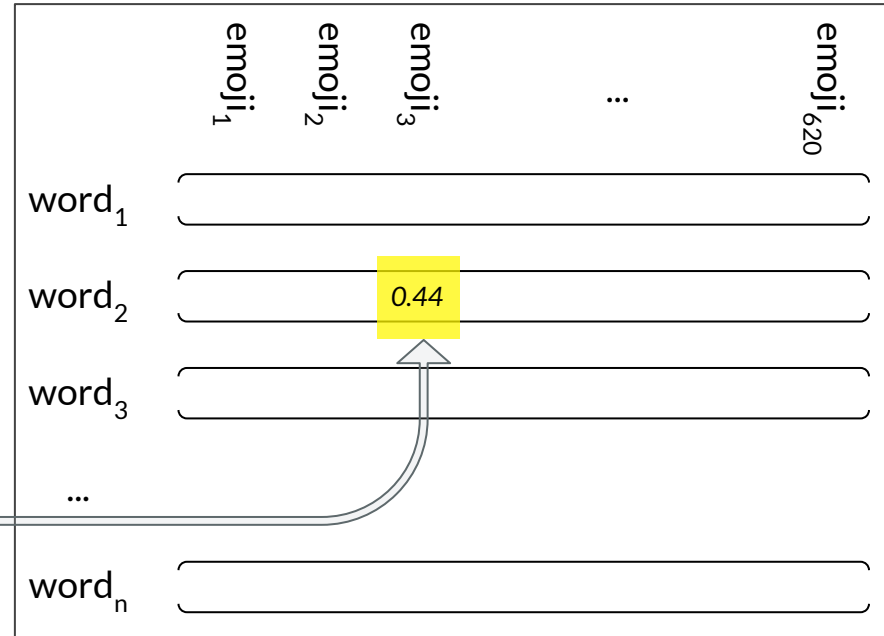
Vector Induction

Word2Vec on Tweets corpus



$$\text{Cosine_Similarity}(\text{word}_2, \text{emoji}_3) = 0.44$$

Emoji Vectors



Emoji Vector Induction

	dim ₁	dim ₂	...	dim ₃₀₀
word ₁	0.144288	-0.460809	...	0.135627
word ₂	0.442811	0.128091	...	-0.196277
...
word _n	0.128221	-0.120809	...	-0.360207
emoji ₁	0.284455	0.460809	...	-0.155627
emoji ₂	0.142886	-0.608092	...	0.356227
...
emoji ₆₂₀	0.542283	0.410809	...	0.315627

Regular Word Vectors (300D)

$\text{Sim}(\text{word}_2, \text{emoji}_1)$

	emoji ₁	emoji ₂	...	emoji ₆₂₀
word ₁	0.744283	0.608099	...	0.311627
word ₂	0.444281	0.181090	...	0.692727
...
word _n	0.612822	0.120809	...	0.336007
emoji ₁	1	0.460809	...	0.556217
emoji ₂	0.714288	1	...	0.336127
...
emoji ₆₂₀	0.154228	0.941080	...	1

Interpretable Emoji-Based Word Vectors (620D)

Evaluation of New Vectors

EmoInt – WASSA Shared Task

Task: given a tweet and an emotion X, determine the intensity or degree of emotion X felt by the speaker

- Predicts the intensity of emotions in Tweets
- Intensities are real valued scores in $[0,1]$
- Emotions: classified as anger, fear, joy, sadness

Approach: Supervised Learning Method

- **Random Forest** regressor with 800 trees
- Combines many features including the output of a CNN-LSTM network that uses **our Emoji Vectors as the word embedding**

EmoInt Results Including Other Baselines

	Methods	Anger	Fear	Joy	Sadness	Average	Dim
Interpretable	Affective Tweets	0.65	0.66	0.60	0.69	0.65	n/a
	EmoTag	0.70	0.73	0.69	0.75	0.72	620
Non-Interpretable	Random Int.	0.68	0.72	0.66	0.73	0.70	300
	word2vec	0.70	0.72	0.67	0.75	0.71	300
	GloVe	0.70	0.73	0.68	0.76	0.72	300
	GloVe Twitter	0.72	0.74	0.68	0.76	0.73	200

Pearson Correlations between Gold Score and Predicted Emotion Score for Tweets





Evaluating Sentiment & Emotion Scores

Sentiment Score Generation

Evaluating **Sentiment** of Emojis

- Prediction
 - NRC EmoLex is used to capture **sentiment words** from EmoTag
 - Find top K words (based on EmoTag Similarity Scores) for a given emoji
 - **Aggregated similarity scores** (K=3) are the final sentiment score for that emoji
- Evaluation
 - we use Sentiment of Emojis by Novak et al. as ground truth

Sentiment Score Evaluation

Unicode	Emoji	Novak Score	EmoTag Score	Description
U+1F61C		0.455	0.482	Face with stuck-out tongue and winking eye
U+1F617		0.611	0.591	Kissing face
U+1F49A		0.656	0.654	Green heart
U+1F48B		0.691	0.744	Kiss mark

Comparison of Emoji Sentiment Score

Emoji Group	Correlations
Top 100 positive emojis	0.71
Emojis with face	0.45
Monkey face emojis	0.53
Emojis with kissing	0.14

Pearson Correlations of Our Sentiment Score and Novak's Score

Emotion Score Generation

Evaluating **Emotion** of Emojis

- Prediction
 - NRC EmoLex is used to capture **emotion words** from EmoTag
 - Rank top K words (based on EmoTag Similarity Scores) for a given emoji
 - **Weighted average scores** (K=3) are the final emotion score for a given emoji
- Evaluation 1
 - Affect Intensity Lexicon from NRC is used to reproduce their score using EmoTag
 - Rank top K emojis (based on EmoTag Similarity Scores) for a given word
 - Arithmetic mean (K=10) is the final emotion scores for that word
- Evaluation 2
 - Emoji2Emotion is used to predict **Emotion Label** for Emojis

Emotion Score Evaluation 1

Emoji	Name	A	F	J	S
U+1F620 😡	Angry face	0.49	0.36	0.07	0.44
U+1F46E 👮	Police officer	0.34	0.49	0.16	0.27
U+1F492 💒	Wedding	0.09	0.14	0.63	0.25
U+1F4A9 💩	Pile of poo	0.35	0.34	0.15	0.47

Snapshot of Proposed Emotion Score for Emojis

Tokens	A	F	J	S	Avg
Common Words	0.51	0.41	0.19	0.50	0.40
All Words	0.45	0.40	0.20	0.45	0.37

Pearson Correlations of Our Score &
Gold Score for Affect Intensity Lexicon

Emotion Score Evaluation 2

Unicode	Name	Emoji	E2E Label	Anger	Joy	Sadness
U+1F612	Unamused face	😞	anger	0.418	0.119	0.333
U+1F602	Face with tears of joy	😂	joy	0.381	0.099	0.326
U+1F60D	Smiling face with heart-eyes	😍	joy	0.307	0.308	0.137
U+1F60A	Smiling face with smiling eyes	😊	joy	0.067	0.248	0.247
U+1F495	Two hearts	❤️	joy	0.172	0.383	0.142
U+1F601	Beaming face with smiling eyes	😄	joy	0.091	0.123	0.079
U+263A	Smiling face	😊	joy	0.095	0.245	0.176
U+1F604	Grinning face with smiling eyes	😄	joy	0.184	0.188	0.149
U+1F618	Face blowing a kiss	😘	joy	0.233	0.215	0.144
U+1F64F	Folded hands	🙏	joy	0.187	0.485	0.351
U+1F62D	Loudly crying face	😭	sadness	0.246	0.198	0.272
U+1F629	Weary face	😓	sadness	0.236	0.186	0.234
U+1F622	Crying face	😭	sadness	0.284	0.210	0.333

A comparison between Emoji2EMotion (E2E) and EmoTag

Conclusion: EmoTag

- It's a huge and meaningful collection of Emoji centric Tweets
- It shows how emojis and words co-occur in social media, including their connection to emotions
- It provides a unique way to create interpretable word embedding with the help of emoji



Thank You!

Contact - abu.shoeb@rutgers.edu
All resources can be found at <http://emoji.nlproc.org>

Backup

Co-Occurrences

Emoji	Description	Top Occurring Hashtags
U+1F63A 🐱	Smiling cat face with open mouth	#cat #cats #CatsOfInstagram #CatsOfTwitter #caturday #catstagram #meow #cute #kitten #kitty #pet #catlover #love #instacat #pets
U+2728 ✨	Sparkles	#LiveMe #MakeMeSmile #LiveMoment #OnAirNow #UncutATL #AmeATL #NashsNewVideo
U+1F602 😂	Face with tears of joy	#periscope #headnod #funny #trump #lol #MTVHottest #kca #gym #lmao #comedy
U+1F60D 😍	Smiling face with heart-shaped eyes	#periscope #love #kca #repost #UncutATL #BestFanArmy #TeenChoice #ameatl #MTVHottest #tl #mcm #iHeartAwards #bts #wcw #snapchat #ThatsMyGirl
U+1F600 😄	Grinning face	#joy #sunset #humanity #win #love #xmas #giveaway #happy #competition #FullScreenParty #shopping #smile #NewDeals #fun
U+1F620 😡	Angry face	#AlwaysTrump #CrimeMinisterOfPak #TheWalkingDead #FirstWorldProblems #Trump #fml #BlackLivesMatter #wtf #ImWithHer #NeverHillary #seriously #lock
U+1F637 🏠	Face with medical mask	#sick #flu #gross #help #GreysAnatomy #yuck #HillarysHealth #sos #sickday #nasty #disgusting #insomnia #cold #dying
U+1F631 😱	Face screaming in fear	#TheWalkingDead #twd #GameOfThrones #horror #pll #PrettyLittleLiars #omg #GreysAnatomy #gotham #halloween #TheExorcist #Trump #ScreamQueens #scary

Word	Emoji	Description
miss	U+1F62D 😭 U+2764 ❤️ U+1F622 🥲	loudly crying face heavy black heart crying face
happy	U+1F389 🎉 U+2764 ❤️ U+1F618 😘	party popper heavy black heart face throwing a kiss
love	U+1F60D 😍 U+1F618 😘 U+2764 ❤️	smiling face with heart-shaped eyes face throwing a kiss heavy black heart

Formation of Lexicons - An Example

Tokens		
same	1	2
to	1	2
you	1	2
keep	1	2
smiling	1	2
happy	1	2+2
hoidaze	1	2
good	0	2
morning	0	2
thursday	0	2

Same to you  keep smiling  Happy Hoidaze    

   GOOD MORNING! HAPPY THURSDAY!   

Overview of Previously Released Dataset

Paper	Year	Lang.	Manual Annotation?	# of Emoji	Source/Size	Class/Output
Sentiment of Emojis	2015	13 EUL	83 Human Annotators	751	1.6 M Tweets - only 4% has emoji	Sentiment Lexicon
Emoji2Vec	2016	English	No	1661	6088 Emoji Descriptions	Pre-trained embeddings
EmoWordNet	2018	English	DepecheMood and crowd-sourced	X	67K Terms from EWN	Emotion Lexicon
Emoji2Emotion	2018	English	500 Human annotated tweets	31+50	84777 tweets	Emoji Emotion Mapping Tech.
EmoLex	2010	English	1012	X	200 n-grams and bi-grams in 4 categories	Emotion Lexicon

There are no such huge dataset consists of frequently used emoji and text