Understanding Emotions:
A Dataset of Tweets to Study Interactions between Affect Categories

Saif M. Mohammad and Svetlana Kiritchenko
National Research Council Canada

Emotions

- Central to how we make sense of the world
- Commonplace and familiar, yet complex and nuanced
- There is a lot we do not know
  - how to categorize emotions
  - how the mind represents emotions
  - the relationships between different emotions or affect categories

Psychological Models of Emotions:

- basic emotions models (Plutchik, Ekman, etc.)
- valence, arousal, dominance model

An large majority of past work has focused on one model or another.

We annotate data for both:

- valence, arousal, and dominance
- basic emotions (such as anger, fear, and joy)

We Annotate

- tweets for the emotions of people that posted the tweets
  - emotions that can be inferred from the text of the tweet
  - tweets are self-contained, widely used, public posts, and tend to be rich in emotions

- for these affect dimensions
  - Current work: anger, fear, joy, sadness, and valence
  - Future work: arousal and dominance

- for coarse classes as well as fine-grained real-valued scores indicating the intensity of emotion

Tasks

First introduced in the WASSA-2017 Shared Task:
Emotion Intensity in Tweets

1. Emotion Intensity Regression (EI-reg):

Given a tweet and an emotion E,

determine the intensity of E that best represents the mental state of the tweeter

- a real-valued score between 0 (least E) and 1 (most E)

Natural language applications benefit from knowing both the class of emotion and its intensity

- E.g., useful for commercial customer satisfaction system to distinguish between significant frustration or anger vs. instances of minor inconvenience

Tasks

1. Emotion Intensity Regression (EI-reg):
   Given a tweet and an emotion E,
   determine the intensity of E that best represents the mental state of the tweeter
   - a real-valued score between 0 (least E) and 1 (most E)

2. Emotion Intensity Ordinal Classification (EI-oc):
   Given a tweet and an emotion E,
   classify the tweet into one of four ordinal classes of intensity of E that best represents the mental state of the tweeter;
   - not angry, slightly angry, moderately angry, very angry

Tasks (continued)

3. Valence (Sentiment) Regression (V-reg):

   Given a tweet,

   determine the intensity of sentiment or valence (V) that best represents the mental state of the tweeter

   - a real-valued score between 0 (most negative) and 1 (most positive)

4. Valence Ordinal Classification (V-oc):

   Given a tweet,

   classify it into one of seven ordinal classes of valence (sentiment intensity) that best represents the mental state of the tweeter

   - very negative, moderately negative, slightly negative, neutral or mixed, slightly positive, moderately positive, very positive

Tasks (continued)

All five tasks part of SemEval-2018 Task 1: Affect in Tweets

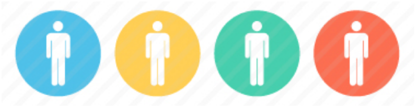
5. Emotion Classification (E-c):

Given a tweet,
classify it into one, or more, of twelve given categories
that best represent the mental state of the tweeter.

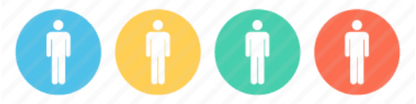
- anger (also includes annoyance, rage)
- anticipation (also includes interest, vigilance)
- disgust (also includes disinterest, dislike, loathing)
- fear (also includes apprehension, anxiety, terror)
- joy (also includes serenity, ecstasy)
- love (also includes affection)
- optimism (also includes hopefulness, confidence)
- pessimism (also includes cynicism, no confidence)
- sadness (also includes pensiveness, grief)
- surprise (also includes distraction, amazement)
- trust (also includes acceptance, liking, admiration)
- neutral or no emotion

■ Plutchik emotions
■ other

Motivation

Human annotations of tweets for emotions 

- For use by automatic systems:
  - that detect emotions in tweets
  - other emotion related tasks such as detecting stance, personality traits, well-being, cyber-bullying, etc.



- To draw inferences about people:
  - to understand emotions, or how we convey emotions through language

Research Questions

- which emotions often present together in tweets?
- how reliably can we order tweets as per emotion intensity?
- how do the intensities of the three negative emotions relate to each other?
- how do the intensities of the basic emotions relate to valence, arousal, and dominance?

Collect Tweets using Query Terms

For each emotion,

- we select 50 to 100 related terms from the *Roget's Thesaurus*
  - associated with that emotion at different intensity levels
    - for anger: *angry, mad, frustrated, annoyed, peeved, irritated, miffed, fury,* and so on
    - for sadness: *sad, devastated, sullen, down, crying, dejected, heartbroken, grief*, and so on

- emojis that are associated with the four emotions
- emoticons such as :), :(, and :D that are indicative of happiness and sadness
- synonyms of the emotion words in a word-embeddings space created from tweets

Presence of terms does not guarantee an emotion or a certain intensity of the emotion.

- Overall, the set is relatively more likely to be conveying emotions

Tweets

- Polled the Twitter API for tweets that included the query terms
  - discarded retweets and tweets with urls

- For about 10% of the tweets:
  - Removed the trailing emoticon, emoji, or hashtagged query term

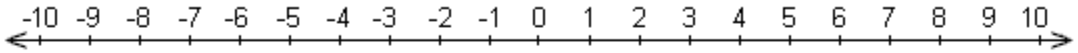
That jerk stole my photo on Tumblr #grrr #angry

⬇

That jerk stole my photo on Tumblr #grrr

Affect in Tweets Dataset: Overview

- Emotion intensity datasets: sampled from the collected tweets
- Valence dataset: selected a subset of tweets from each of the four emotion intensity datasets
- Emotion classification dataset: selected all the tweets from the four emotion intensity datasets



How to capture fine-grained emotion intensity reliably? A harder task!

Humans are not good at giving real-valued scores:

- difficult to maintain consistency across annotators
- difficult for an annotator to be self consistent
- scale region bias

Intensity Annotations

Best–Worst Scaling (Louviere & Woodworth, 1990)**:**
Give k terms and ask which is most X, and which is least X
(*k is usually 4 or 5*)

- preserves the comparative nature
- keeps the number of annotations down to about 2N
- leads to more reliable, less biased, more discriminating annotations (Kiritchenko and Mohammad, 2017, Cohen, 2003)

Example BWS Annotation Instance: for emotion intensity from tweets

Speaker 1: These days I see no light. Nothing is working out #depressed
Speaker 2: The refugees are the ones running from terror.
Speaker 3: Tim is sad that the business is not going to meet expectations.
Speaker 4: Too many people cannot make ends meet with their wages.

Q1. Which of the four speakers is likely to be the MOST SAD (or having a mental state most inclined towards sadness)

Q2. Which of the four speakers is likely to be the LEAST SAD (or having a mental state least inclined towards sadness)

Ran Annotations on CrowdFlower

Dataset	Scheme	Location	Item	#Items	#Annotators	MAI	#Q/Item	#Annotat.
English								
E-c	categorical	World	tweet	11,090	303	7	2	174,356
EI-reg								
anger	BWS	USA	4-tuple of tweets	2,780	168	4	2	27,046
fear	BWS	USA	4-tuple of tweets	2,750	220	4	2	26,908
joy	BWS	USA	4-tuple of tweets	2,790	132	4	2	26,676
sadness	BWS	USA	4-tuple of tweets	2,744	118	4	2	26,260
V-reg	BWS	USA	4-tuple of tweets	5,134	175	4	2	49,856
Total								**331,102**

Q = Questions
Location = Location of annotators
MAI = Minimum (and Median) Annotations per Item

Ran Annotations on CrowdFlower

Dataset	Scheme	Location	Item	#Items	#Annotators	MAI	#Q/Item	#Annotat.
English								
E-c	categorical	World	tweet	11,090	303	7	2	174,356
EI-reg								
anger	BWS	USA	4-tuple of tweets	2,780	168	4	2	27,046
fear	BWS	USA	4-tuple of tweets	2,750	220	4	2	26,908
joy	BWS	USA	4-tuple of tweets	2,790	132	4	2	26,676
sadness	BWS	USA	4-tuple of tweets	2,744	118	4	2	26,260
V-reg	BWS	USA	4-tuple of tweets	5,134	175	4	2	49,856
Total								**331,102**

Q = Questions
Location = Location of annotators
MAI = Minimum (and Median) Annotations per Item

Annotation Aggregation

- Emotion classification labels:
  - If more than k of the seven people indicated that a certain emotion applies, then that label was chosen

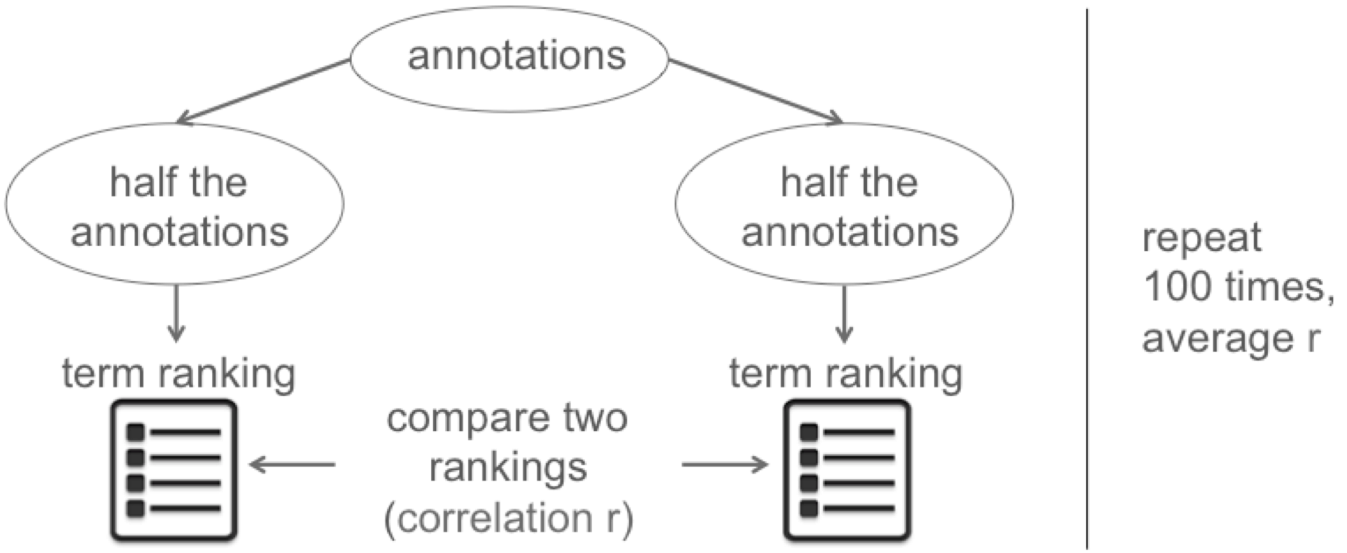
- Intensity scores:
  - counting method (Orme, 2009)

    score(w) = (#mostE(w) - #leastE(w)) / #annotations(w)

    the scores are re-scaled to be in the interval:
    0 (lowest emotion intensity)
    to 1 (highest emotion intensity)

Reliability (Reproducibility) of Annotations

Average split-half reliability (SHR): a commonly used approach to determine consistency (Kuder and Richardson, 1937; Cronbach, 1946)

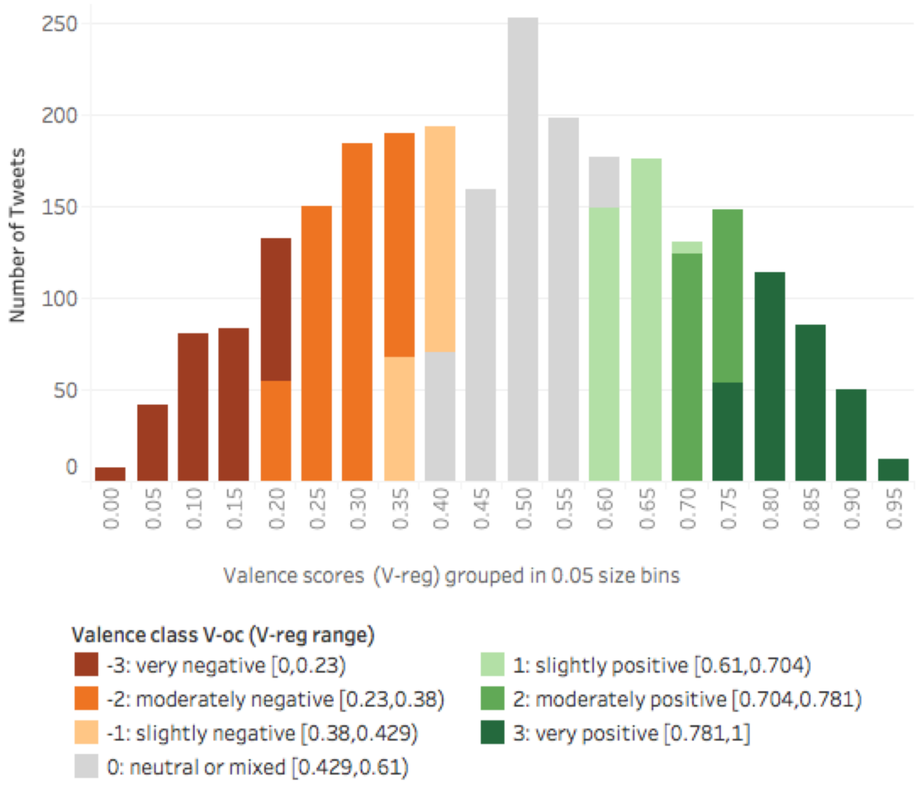


Split-Half Reliability: Emotion Intensity Annotations

Emotion	Spearman Corr. (r)	Pearson Corr. (ρ)
anger	0.89	0.90
fear	0.84	0.85
joy	0.90	0.91
sadness	0.82	0.83
valence	0.92	0.92

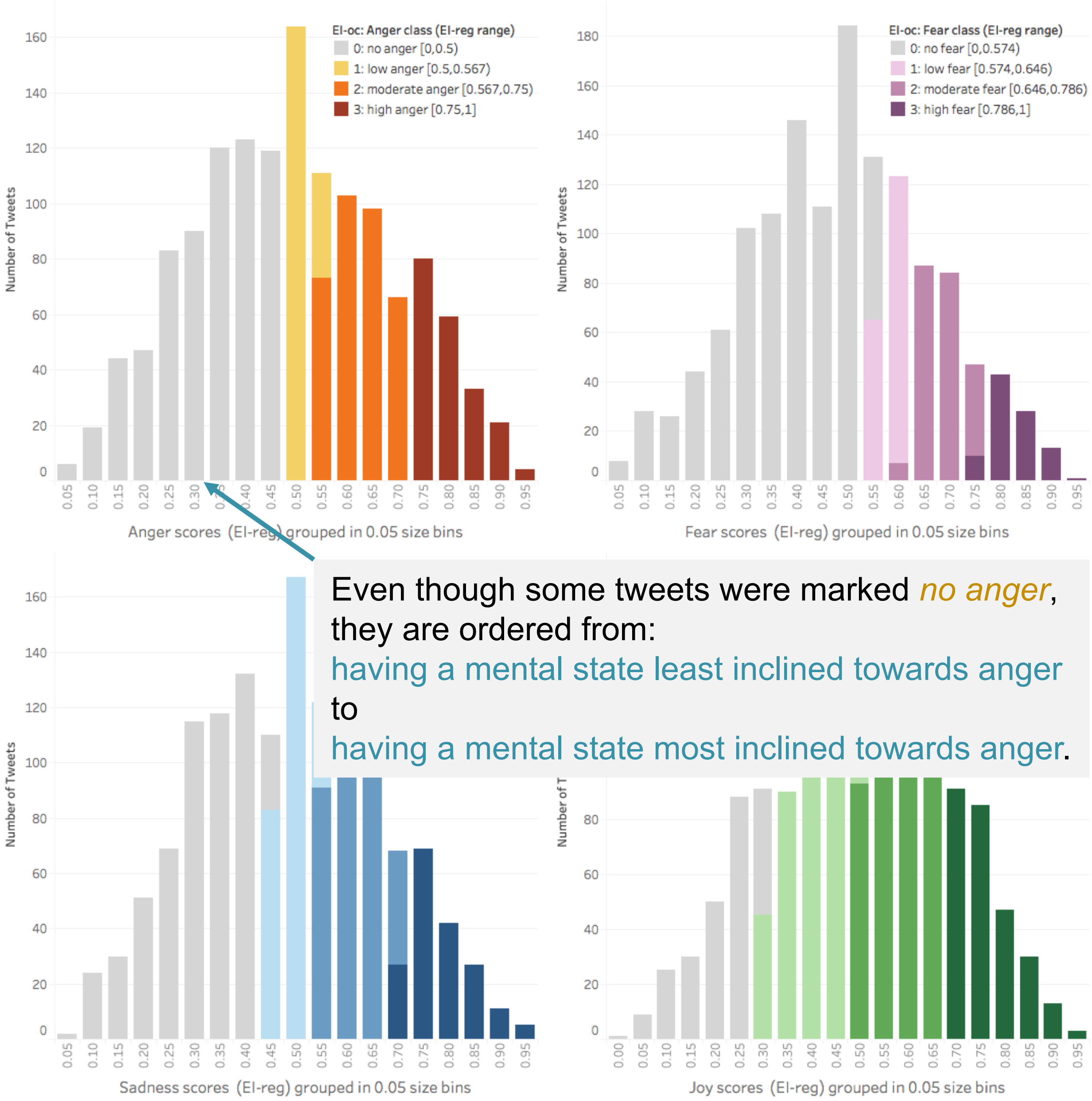
High correlation numbers indicate a high degree of reproducibility.

Distribution: Valence score (V-reg) and Valence class (V-oc)



The boundaries between valence classes were manually identified by the authors.

Emotion Intensity and Class Distributions



Even though some tweets were marked *no anger*, they are ordered from:
having a mental state least inclined towards anger
to
having a mental state most inclined towards anger.

SemEval-2018 Affect in Tweets Dataset

Dataset	Train	Dev	Test	Total
English				
E-c	6,838	886	3,259	10,983
EI-reg, EI-oc				
anger	1,701	388	1,002	3,091
fear	2,252	389	986	3,627
joy	1,616	290	1,105	3,011
sadness	1,533	397	975	2,905
V-reg, V-oc	1,181	449	937	2,567

A tweet in any training or development set does not occur in any test set.

Relationships Between Affect Dimensions

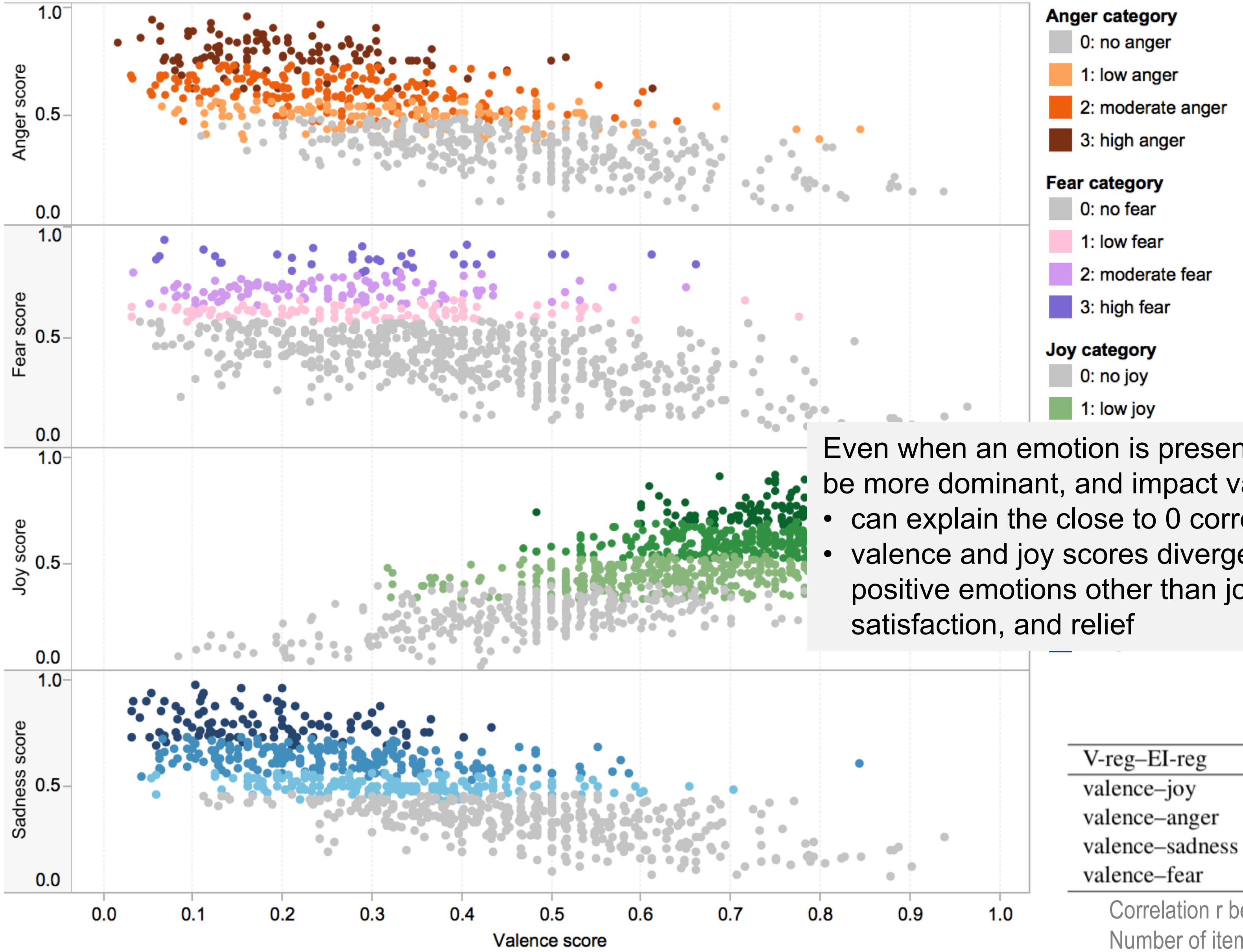
Co-occurrence of Emotions (from E-c data)

If there is anger, then there is an 81% chance there is disgust as well.

Emotion	anger	anticipat..	disgust	fear	joy	love	optimism	pessimis..	sadness	surprise	trust
anger	1.00	0.07	0.81	0.15	0.09	0.01	0.07	0.11	0.35	0.04	0.01
anticipation	0.18	1.00	0.19	0.16	0.47	0.08	0.45	0.07	0.16	0.12	0.11
disgust	0.80	0.07	1.00	0.17	0.08	0.01	0.05	0.14	0.40	0.04	0.01
fear	0.31	0.13	0.37	1.00	0.13	0.02	0.13	0.19	0.35	0.04	0.03
joy	0.08	0.17	0.08	0.06	1.00	0.29	0.62	0.02	0.10	0.06	0.09
love	0.03	0.09	0.02	0.02	0.94	1.00	0.66	0.01	0.07	0.05	0.11
optimism	0.08	0.20	0.06	0.07	0.78	0.26	1.00	0.03	0.09	0.05	0.13
pessimism	0.34	0.09	0.45	0.27	0.08	0.01	0.07	1.00	0.71	0.03	0.01
sadness	0.43	0.08	0.49	0.20	0.13	0.03	0.10	0.28	1.00	0.03	0.01
surprise	0.25	0.31	0.25	0.14	0.49	0.11	0.29	0.06	0.20	1.00	0.07
trust	0.06	0.31	0.04	0.09	0.69	0.26	0.78	0.02	0.07	0.07	1.00

For a pair of emotions, *i* and *j*, the cells show the proportion of tweets labeled with both emotions *i* and *j*, out of all the tweets annotated with emotion *i*. Darker shades are used for higher proportions.

- highly contrasting emotions (love – disgust) have low scores
- pairs of emotions with scores greater than 0.5: anger – disgust, disgust – anger, love – joy, love – optimism, etc.
- for love and joy, the association is markedly stronger only in one direction



Even when an emotion is present, another emotion could be more dominant, and impact valence:

- can explain the close to 0 correlation with fear
- valence and joy scores diverge when tweets convey positive emotions other than joy such as optimism, satisfaction, and relief

V-reg–EI-reg	all data	the emotion present
valence–joy	0.79 (607)	0.65 (496)
valence–anger	-0.73 (598)	-0.40 (282)
valence–sadness	-0.73 (603)	-0.47 (313)
valence–fear	-0.60 (600)	-0.09 (175)

Correlation r between valence and the emotions. Number of items is shown in brackets.

Relationship between valence and emotion intensity

Pearson Correlation between Pairs of Negative Emotions

EI-reg–EI-reg	all data	both emotions present
fear–sadness	0.64 (668)	0.09 (174)
anger–sadness	0.62 (616)	0.08 (224)
anger–fear	0.51 (599)	-0.13 (124)

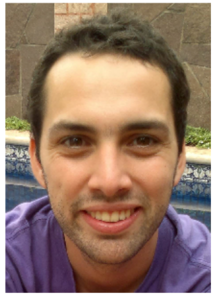
The scores are much closer to 0, when considering only those tweets where both emotions are present.

SemEval-2018 Task 1: Affect in Tweets

https://competitions.codalab.org/competitions/17751

Tasks: Inferring likely affectual state of the tweeter

- emotion intensity regression and ordinal classification
- sentiment intensity regression and ordinal classification
- emotion classification task



Felipe José Bravo Márquez

Mohammad Salameh

Svetlana Kiritchenko

English, Arabic, and Spanish Tweets

75 Team (~200 participants)

Includes a separate evaluation component for inappropriate biases in the systems.

Summary

- We created a new Affect in Tweets dataset:
  - more than 11,000 tweets
  - annotated for four basic emotions and valence
  - annotated for coarse classes and for fine-grained real-valued scores of intensity

- Useful for:
  - training and testing supervised machine learning algorithms (SemEval-2018 Task 1)
  - understanding emotions and relations between affect categories

Resources Available at: www.saifmohammad.com

- Affect in Tweets Data
- Sentiment and emotion lexicons
- Links to shared tasks
- Interactive visualizations

Saif M. Mohammad
saif.mohammad@nrc-cnrc.gc.ca

Emotions Evoked by Art

WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art. Saif M. Mohammad and Svetlana Kiritchenko. In *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, May 2018, Miyazaki, Japan.

Art and Emotions

- Art is imaginative human creation meant to evoke an emotional response
- Large amounts of art are now online
  - With title, painter, style, year, etc.
  - Not labeled for emotions evoked
- Useful:
  - Ability to search for paintings evoking the desired emotional response
  - Automatically detect emotions evoked by paintings
  - Automatically transform (or generate new) paintings
  - Identify what makes paintings evocative

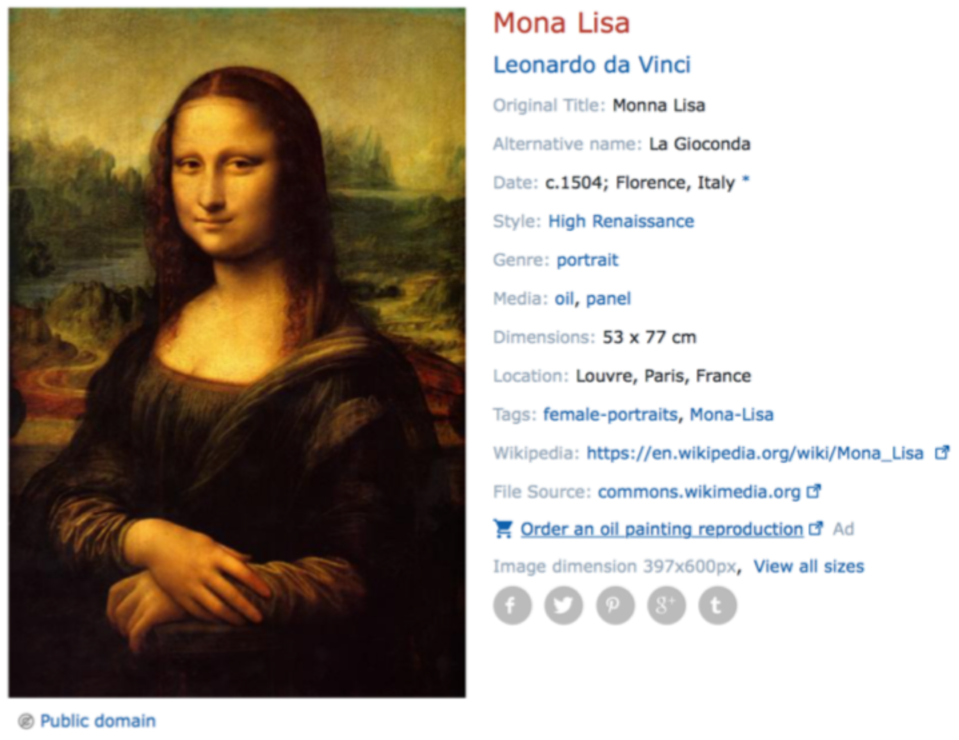


Figure 1: WikiArt.org's page for the *Mona Lisa*.

WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art

- ~4K pieces of art (mostly paintings)
- From four styles: *Renaissance Art, Post-Renaissance Art, Modern Art,* and *Contemporary Art*
- 20 categories: Impressionism, Expressionism, Cubism, Figurative art, Realism, Baroque,…
- Annotated for emotions evoked, amount liked, does it depict a face.

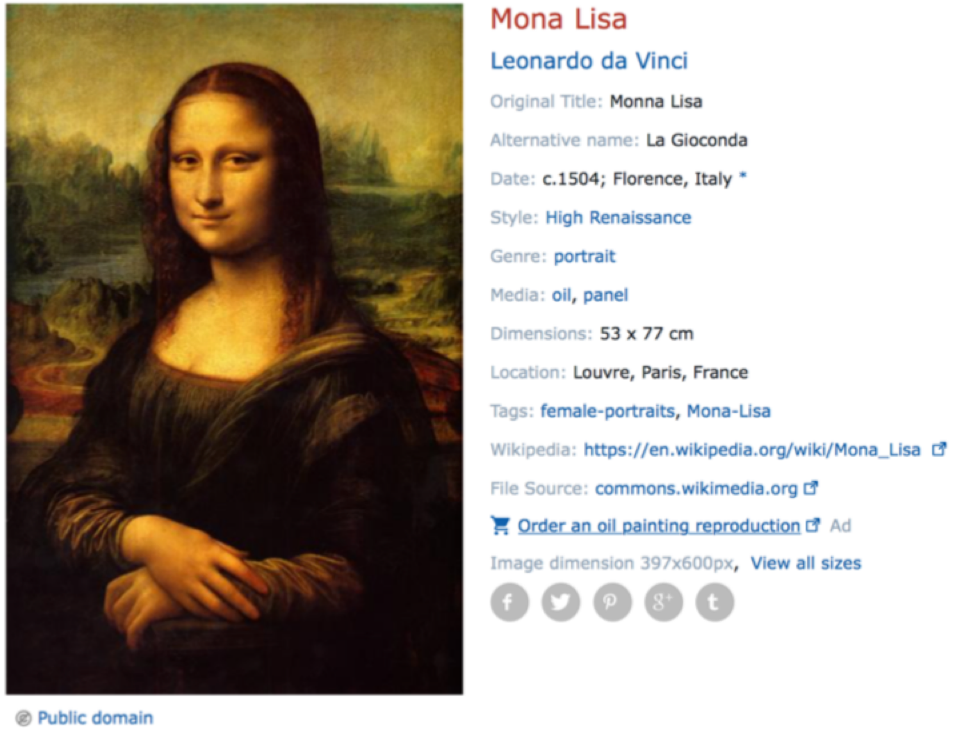
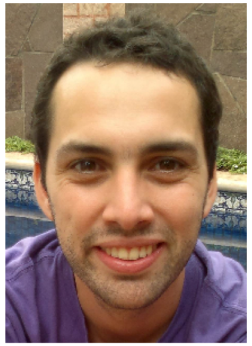


Figure 1: WikiArt.org's page for the *Mona Lisa*. In the WikiArt Emotions Dataset, the *Mona Lisa* is labeled as evoking happiness, love, and trust; its average rating is 2.1 (in the range of -3 to 3).

Emotion is any conscious experience characterized by intense mental activity and a high degree of pleasure or displeasure. Scientific discourse has drifted to other meanings and there is no consensus on a definition.

-- Wikipedia



Felipe José Bravo Márquez

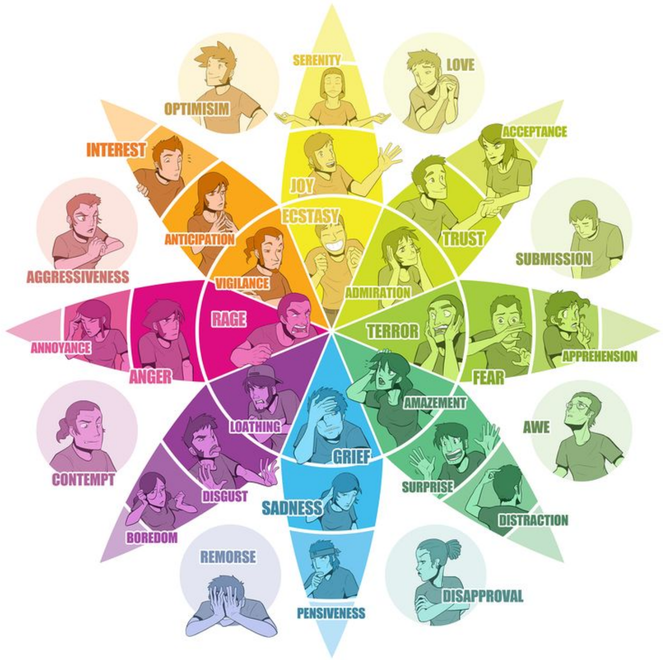
Emotion Intensity in Tweets

Paper:
WASSA-2017 Shared Task on Emotion Intensity. Saif M. Mohammad and Felipe Bravo-Marquez. In *Proceedings of the EMNLP 2017 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA)*, September 2017, Copenhagen, Denmark.

Psychological Theories of Basic Emotions

- Paul Ekman, 1971: Six Basic Emotions
- Plutchik, 1980: Eight Basic Emotions
- And many others

In this work, we focus on four emotions common to most theories: anger, fear, joy, and sadness.



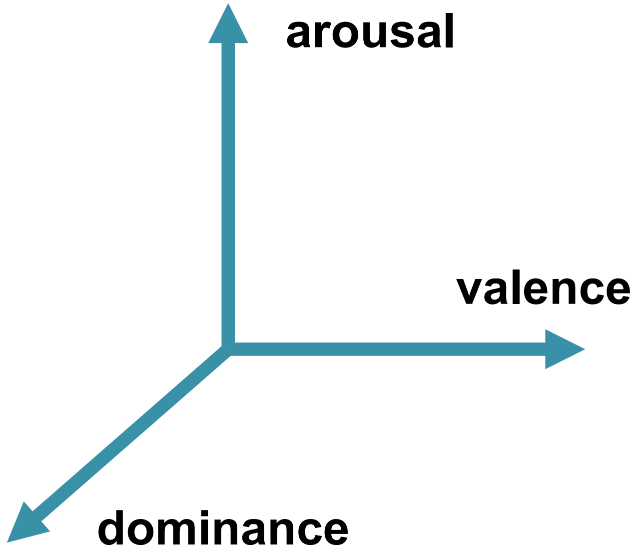
Plutchik's Emotion Wheel
Image credit: Julia Belyanevych

Circumplex Model of Emotions (Russell, 1980)

Primary dimensions of affectual adjectives

- valence: positive/pleasure – negative/displeasure
- arousal: active/stimulated – sluggish/bored
- dominance: powerful/strong – powerless/weak

Emotion is point in the multi-dimensional space



WASSA-2017 Shared Task: Emotion Intensity in Tweets

Task:

Given a tweet and an emotion X, determine intensity of emotion X felt by the speaker,

- a real-valued score between 0 and 1
  - 1: the speaker is feeling the maximum amount of emotion X
  - 0: the speaker is feeling the least amount of emotion X

Data

  - Annotated sentences using BWS

Task website:

http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html