Word Affect Intensities

Saif M. Mohammad

National Research Council Canada

Affect or Emotion

- Refers to concepts such as joy, sadness, fear, guilt, etc.
- Commonplace and familiar, yet complex and nuanced

(For this talk, I will treat the terms *affect* and *emotion* as being synonymous.)

Word-Emotion Associations

Words have associations with emotions:

- violence and shout are associated with anger
- shudder and public speaking are associated with fear
- yummy and vacation are associated with joy
- loss and crying are associated with sadness

Goal: Capture word-emotion associations (out of context binary values)

NRC Emotion Lexicon

- Word—emotion associations for 14,200 English words
- Available at: www.saifmohammad.com

Word-Emotion Intensities

Words can be associated with different intensities (or degrees) of an emotion.

- *outrage* is associated with a greater degree of anger (more anger) than *irritate*
- *sobbing* is associated with more sadness than *sigh*

Goal of this work: Capture word-emotion intensities (out of context scores)

Motivation

Human annotations of words for emotion intensities 

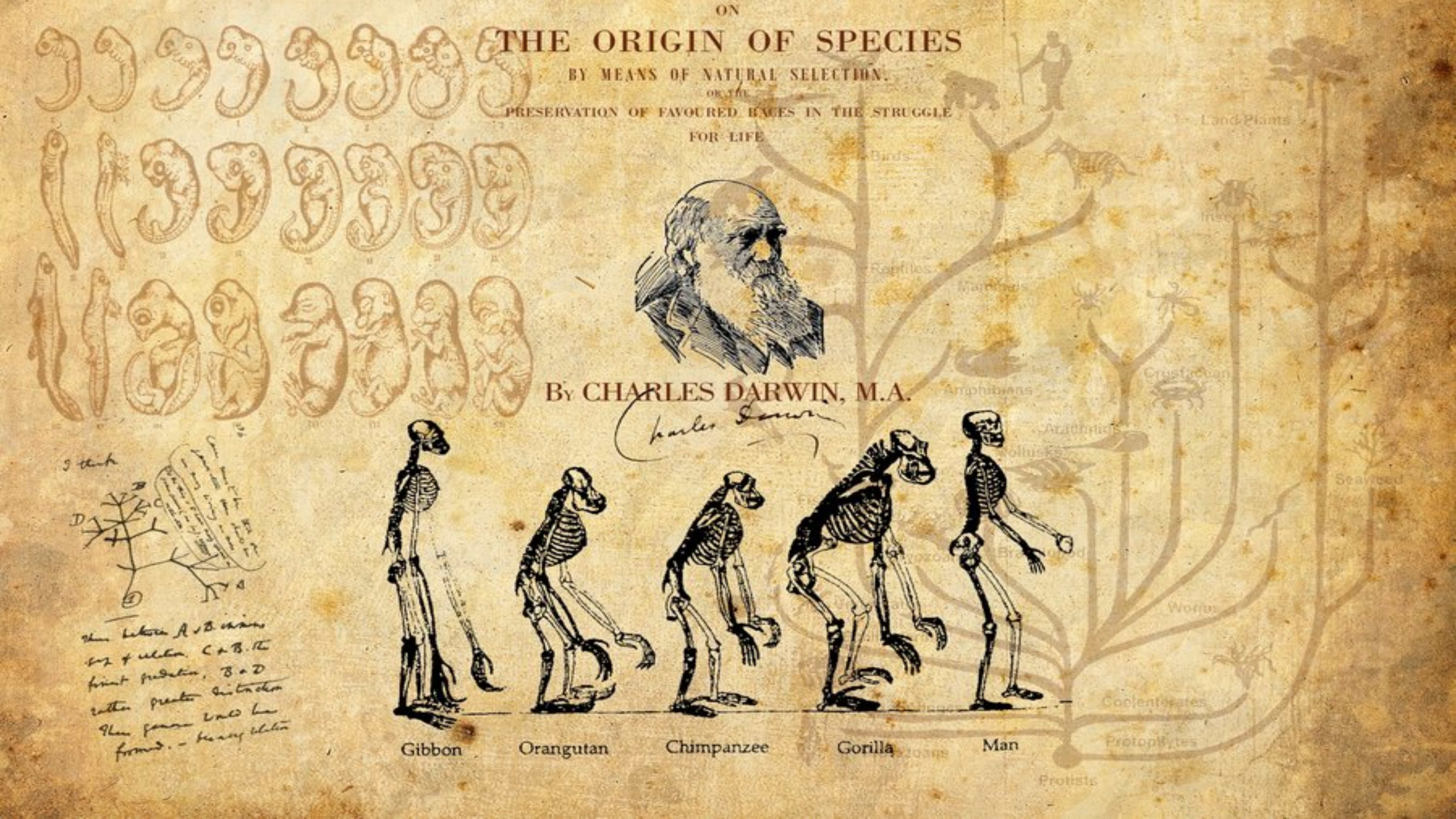
- For use by automatic systems: 
  - predicting emotion intensities of words
  - predicting sentiment and emotions of sentences, tweets, etc.
  - detecting stance, personality traits, well-being, cyber-bullying, etc.

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

Research Questions

- how reliably can we order words as per emotion intensity?
- how do the intensities of different emotions relate to each other?
- how do the intensities of the basic emotions relate to valence, arousal, and dominance?
  - do the words for anger fall predominantly in a particular region of the valence--arousal--dominance space?

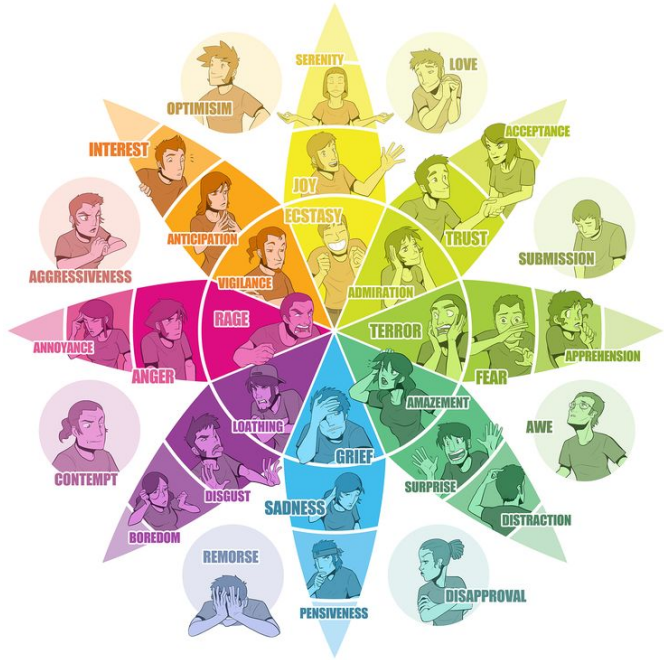
Which Emotions?


ON
THE ORIGIN OF SPECIES
BY MEANS OF NATURAL SELECTION,
OR THE
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE
By CHARLES DARWIN, M.A.
I think
Gibbon
Orangutan
Chimpanzee
Gorilla
Man
Land Plants
Birds
Reptiles
Amphibians
Seaweed
Worms
Coelenterates
Protists

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

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

Comparative Annotations

Paired Comparisons (Thurstone, 1927; David, 1963)**:**

If X is the property of interest (positive, useful, etc.),

give two terms and ask which is more X

- less cognitive load
- helps with consistency issues
- requires a large number of annotations
  - order N^2, where N is number of terms to be annotated

Best–Worst Scaling (BWS)

- The annotator is presented with four words (say, A, B, C, and D) and asked:
  - which word is associated with the most anger
  - which word is associated with the least anger

- By answering just these two questions, five out of the six inequalities are known
  - For e.g.:
    - If A: most anger
    - and D: least anger, then we know:

$$A > B, A > C, A > D, B > D, C > D$$

Best–Worst Scaling

- 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)

Building the NRC Affect Intensity Lexicon

Term Selection

Since most words do not convey a particular emotion to a marked degree, annotating all words for all emotions is sub-optimal.

Selected:

- words in the NRC Emotion Lexicon that are marked as associated with anger, fear, joy, and sadness
  - the NRC Emotion Lexicon includes terms that occur frequently in the Google n-gram corpus
- words that co-occurred more often than chance with #anger, #fear, #joy, and #sadness in a tweets corpus (Hashtag Emotion Corpus) (Mohammad, 2012)
  - these words are likely to be associated with the corresponding emotions

Ran BWS Annotations on CrowdFlower

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
anger	1,483	USA	4-tuple of words	2,966	119	4	2	12,212
fear	1,765	USA	4-tuple of words	3,530	82	4	2	14,129
joy	1,268	USA	4-tuple of words	2,536	76	4	2	10,365
sadness	1,298	USA	4-tuple of words	2,596	76	4	2	10,429
Total	**5,814**							**47,135**

The annotation tasks were approved by the National Research Council Canada's Institutional Review Board.

Ran BWS Annotations on CrowdFlower

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
anger	1,483	USA	4-tuple of words	2,966	119	4	2	12,212
fear	1,765	USA	4-tuple of words	3,530	82	4	2	14,129
joy	1,268	USA	4-tuple of words	2,536	76	4	2	10,365
sadness	1,298	USA	4-tuple of words	2,596	76	4	2	10,429
Total	**5,814**							**47,135**

The annotation tasks were approved by the National Research Council Canada's Institutional Review Board.

Ran BWS Annotations on CrowdFlower

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
anger	1,483	USA	4-tuple of words	2,966	119	4	2	12,212
fear	1,765	USA	4-tuple of words	3,530	82	4	2	14,129
joy	1,268	USA	4-tuple of words	2,536	76	4	2	10,365
sadness	1,298	USA	4-tuple of words	2,596	76	4	2	10,429
Total	**5,814**							**47,135**

The annotation tasks were approved by the National Research Council Canada's Institutional Review Board.

Ran BWS Annotations on CrowdFlower

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
anger	1,483	USA	4-tuple of words	2,966	119	4	2	12,212
fear	1,765	USA	4-tuple of words	3,530	82	4	2	14,129
joy	1,268	USA	4-tuple of words	2,536	76	4	2	10,365
sadness	1,298	USA	4-tuple of words	2,596	76	4	2	10,429
Total	**5,814**							**47,135**

The annotation tasks were approved by the National Research Council Canada's Institutional Review Board.

Ran BWS Annotations on CrowdFlower

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
anger	1,483	USA	4-tuple of words	2,966	119	4	2	12,212
fear	1,765	USA	4-tuple of words	3,530	82	4	2	14,129
joy	1,268	USA	4-tuple of words	2,536	76	4	2	10,365
sadness	1,298	USA	4-tuple of words	2,596	76	4	2	10,429
Total	**5,814**							**47,135**

The annotation tasks were approved by the National Research Council Canada's Institutional Review Board.

Quality Control

About 5% of the data was annotated internally beforehand (by the author)

- These gold questions are interspersed with other questions
- If one gets a gold question wrong, they are immediately notified of it
  - feedback to improve task understanding
- If one's accuracy on the gold questions falls below 70%,
  - they are refused further annotation
  - all of their annotations are discarded

Mechanism to avoid malicious or random annotations

Affect Intensity Lexicon: Example entries

Highest anger intensity:

outraged	0.964
brutality	0.959
hatred	0.953

Highest fear intensity:

torture	0.984
terrorist	0.972
horrific	0.969

Lowest anger intensity:

sisterhood	0.015
musical	0.011
tree	0.000

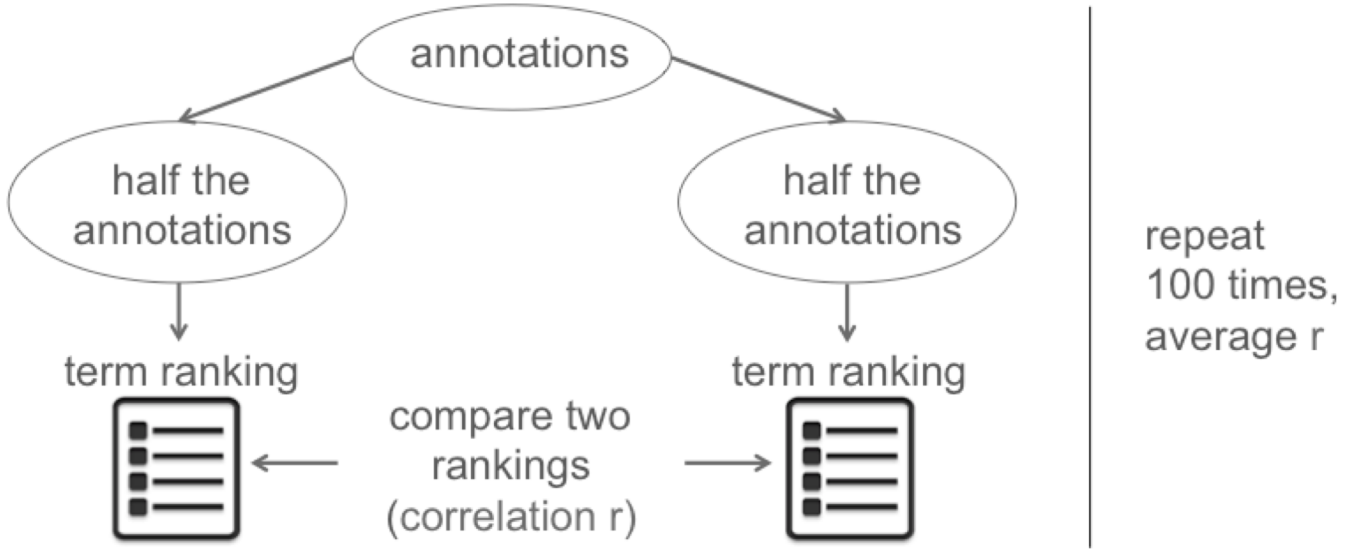
Lowest fear intensity:

volunteer	0.031
lines	0.031
romance	0.031

Scores are in the range 0 (lowest intensity) to 1 (highest 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: Affect Intensity Annotations

Emotion	Spearman Corr. (r)	Pearson Corr. (ρ)
anger	0.906	0.912
fear	0.910	0.912
joy	0.925	0.924
sadness	0.904	0.909

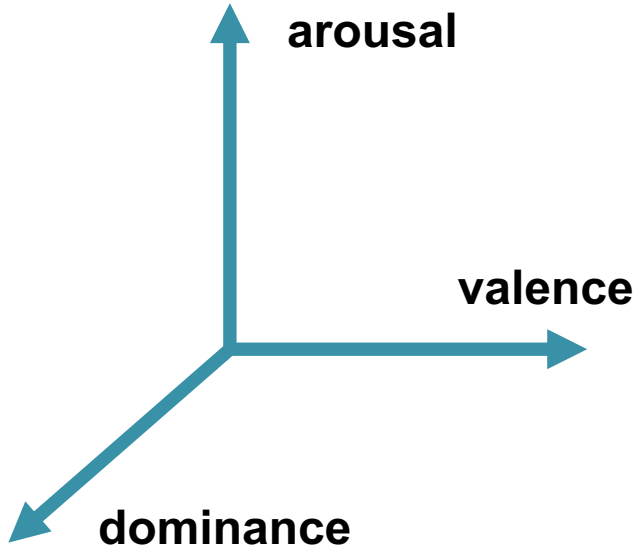
High correlation numbers indicate a high degree of reproducibility.

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



ACL-2018 Paper

Obtaining Valence, Arousal, and Dominance Annotations
(with BWS)

Dataset	#words	Location of Annotators	Annotation Item	#Items	#Annotators	MAI	#Q/Item	#Best–Worst Annotations
valence	20,007	worldwide	4-tuple of words	40,014	1,020	6	2	243,295
arousal	20,007	worldwide	4-tuple of words	40,014	1,081	6	2	258,620
dominance	20,007	worldwide	4-tuple of words	40,014	965	6	2	276,170
Total								**778,085**

Includes:

- Terms from the NRC Emotion Lexicon
- Terms from the Warriner et al. (2013) VAD lexicon
- And some other sources

ACL-2018 Paper

Example Entries in the VAD Lexicon

Dimension	Word	Score↑	Word	Score↓
valence	*love*	1.000	*toxic*	0.008
	happy	1.000	*nightmare*	0.005
	happily	1.000	*shit*	0.000
arousal	*abduction*	0.990	*mellow*	0.069
	exorcism	0.980	*siesta*	0.046
	homicide	0.973	*napping*	0.046
dominance	*powerful*	0.991	*empty*	0.081
	leadership	0.983	*frail*	0.069
	success	0.981	*weak*	0.045

Scores are in the range 0 (lowest V/A/D) to 1 (highest V/A/D).

Split-Half Reliability Scores for the VAD Annotations

Annotations	# Terms	# Annotations	V	A	D
Warriner et al. (2013)	13,915	20 per term	0.914	0.789	0.770

Split-Half Reliability Scores for the VAD Annotations

Annotations	# Terms	# Annotations	V	A	D
Warriner et al. (2013)	13,915	20 per term	0.914	0.789	0.770
Ours (Warriner terms)	13,915	6 per tuple	0.952	0.905	0.906

ACL-2018 Paper

Split-Half Reliability Scores for the VAD Annotations

Annotations	# Terms	# Annotations	V	A	D
Warriner et al. (2013)	13,915	20 per term	0.914	0.789	0.770
Ours (Warriner terms)	13,915	6 per tuple	0.952	0.905	0.906
Ours (all terms)	20,007	6 per tuple	0.950	0.899	0.902

Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. Saif M. Mohammad. In *Proceedings of* the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, July 2018.

Relationships Between Affect Dimensions

Average VAD Scores for the Four Emotions

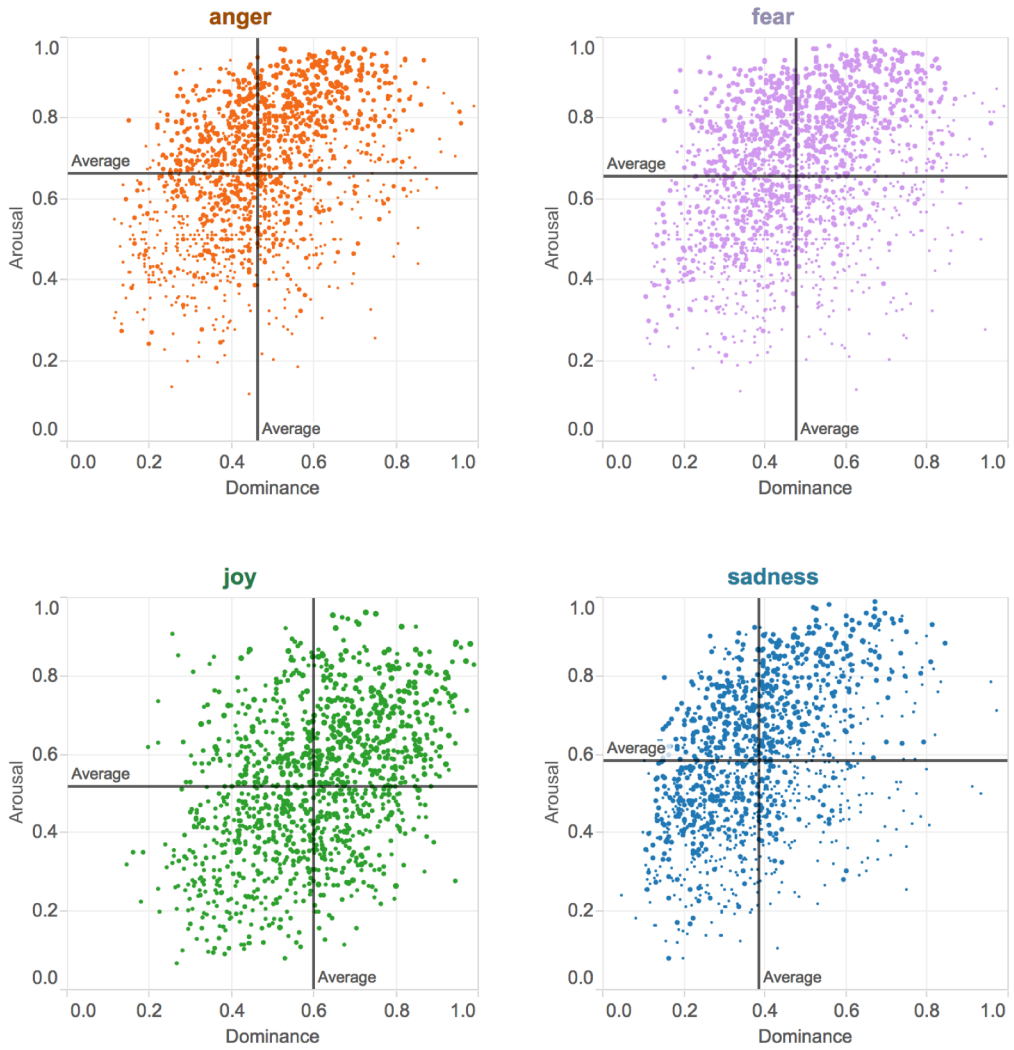
Emotion	Avg. Valence	Avg. Arousal	Avg. Dominance
anger	0.26	0.66	0.46
fear	0.29	0.66	0.48
joy	0.77	0.52	0.60
sadness	0.24	0.58	0.38

The cells are in shades of green with the darkness proportional to the score.

- on average, joy words have higher valence scores (are more positive) than the anger, fear, and sadness words
- joy words also have somewhat lower arousal scores (are more passive) and higher dominance score
- anger and fear have a very similar profile of average VAD scores

Dominance–Arousal scatter plots for words associated with the four emotions

- Negative emotion words belong to an overlapping range of arousal and dominance scores
- The range of scores overlaps markedly with the joy words as well



The size of the point is proportional to the intensity of the emotion.

Summary

- We created the NRC Affect Intensity Lexicon:
  - has entries for close to 6,000 English words
  - has fine-grained real-valued scores of intensity for four basic emotions
  - showed that the annotations are reliable (high split-half reliability scores)

- Useful for:
  - studying the relationships between affect dimensions
    - especially, when used in combination with the NRC VAD Lexicon
  - training and testing automatic algorithms that predict word-emotion intensities
  - a wide variety of sentiment and emotion related tasks
    - in recent tweet tasks (EmoInt2017, SemEval-2018 Task 1):
      the top teams have made substantial use of the Affect Intensity Lexicon

Resources Available at: www.saifmohammad.com

- NRC Affect Intensity Lexicon
- NRC Valence, Arousal, and Dominance Lexicon
- Other lexicons and emotion-labeled tweets datasets
- Links to shared tasks
- Interactive visualizations

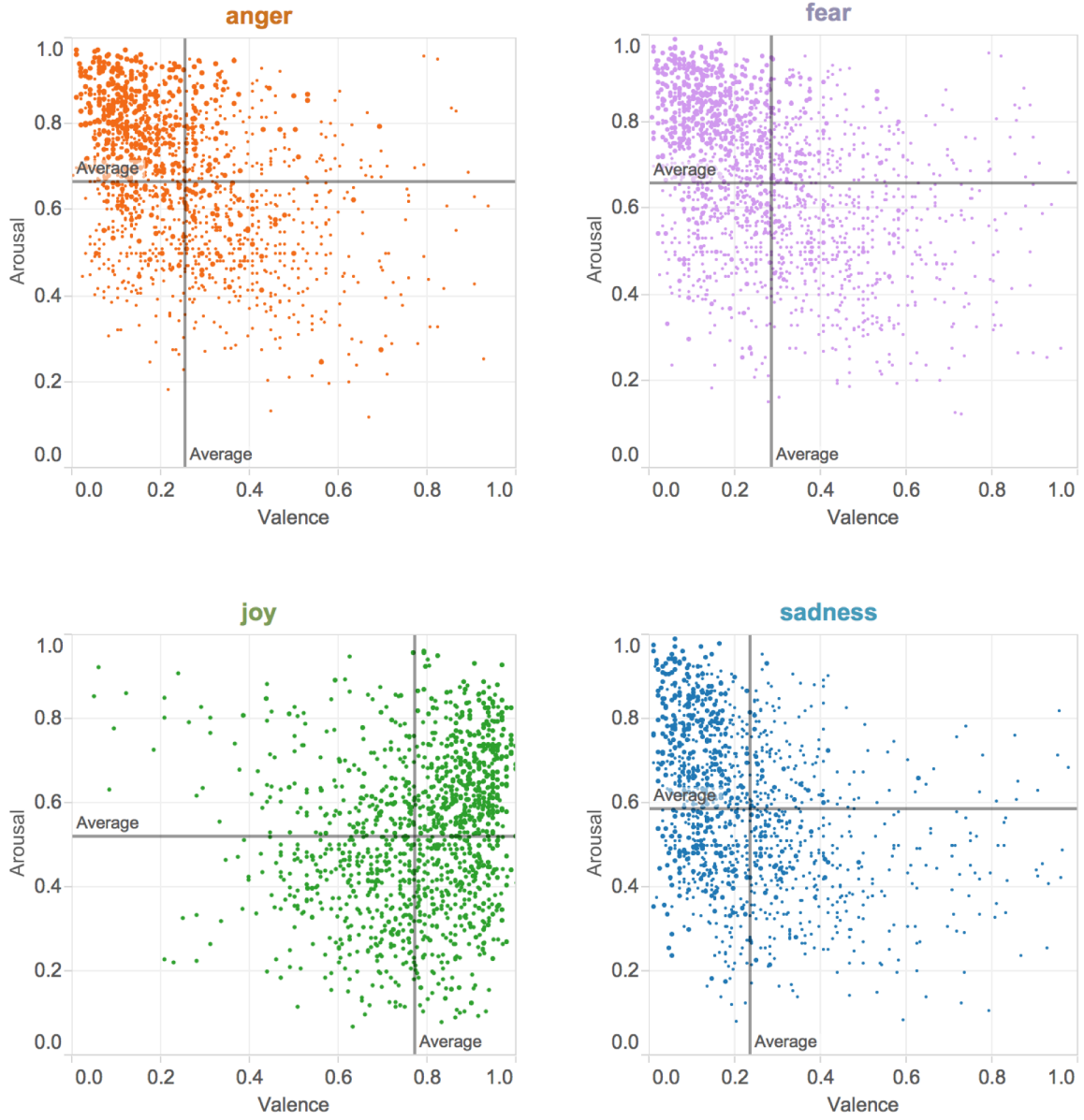
Saif M. Mohammad
saif.mohammad@nrc-cnrc.gc.ca

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

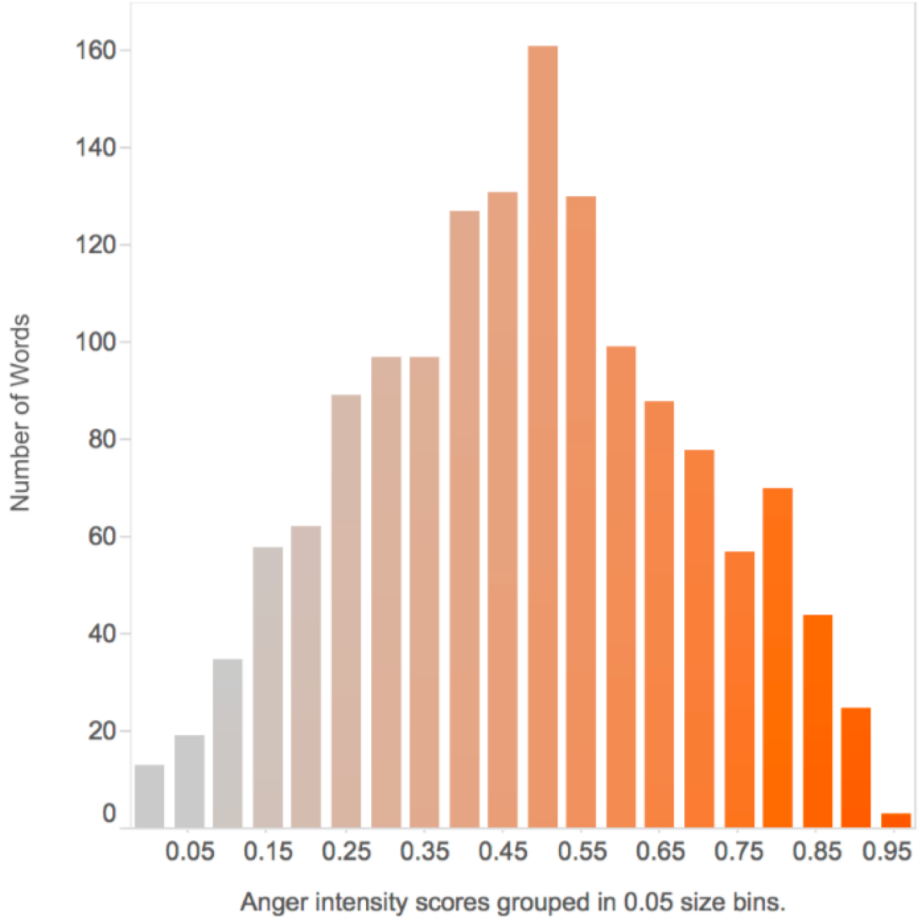
Valence–Arousal scatter plots for words associated with the four emotions

- Joy stands out in terms of high valence
- Points for the negative emotions are in a large common overlapping space



The size of the point is proportional to the intensity of the emotion.

A histogram of word–anger intensities



Anger intensity scores are grouped in bins of size 0.05. The colors of the bars go from gray to orange in increasing order of affect intensity.

Word-Emotion Associations

Words are associations with emotions:

- attack and public speaking are associated with fear
- yummy and vacation are associated with joy
- loss and crying are associated with sadness

Words are associated with high/low valence, arousal, and dominance:

- yummy and vacation are associated with high valence
- thief and infected are associated with low valence
- excited and public speaking are associated with high arousal
- siesta and depression are associated with low arousal
- superhero and influential are associated with high dominance
- sick and unemployed are associated with low dominance

Best–Worst Scaling

- Each of these BWS questions can be presented to multiple annotators.
- We can obtain real-valued scores for all the terms using a simple counting method (Orme, 2009)

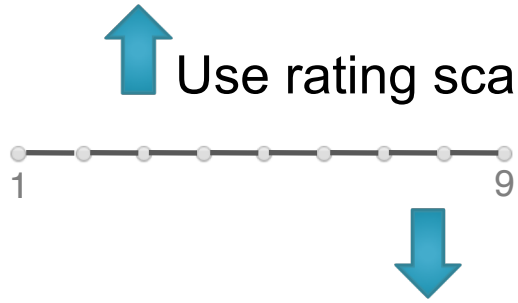
  score(w) = (#best(w) - #worst(w)) / #annotations(w)

  the scores range from:
  -1 (least association with anger)
  to 1 (most association with anger)

  - the scores can then be used to rank all the terms

Split-Half Reliability Scores for Past Work on Valence, Arousal, and Dominance (VAD) Annotations

Annotations	# Terms	# Annotations	V	A	D
Warriner et al. (2013)	13,915	20 per term	0.914	0.789	0.770



Other earlier work on creating valence, arousal, dominance lexicons:

- Affective Norms for English Words (ANEW)

Comparative Annotations

Paired Comparisons (Thurstone, 1927; David, 1963):

If X is the property of interest (positive, useful, etc.),
give two terms and ask which is more X

Need a method that preserves the comparison aspect, without greatly increasing the number of annotations needed.

Possible solution:

Best–Worst Scaling (Louviere & Woodworth, 1990):
(a.k.a. Maximum Difference Scaling or MaxDiff)

Affect Intensity Lexicon: Examples entries

Word	Anger	Word	Fear	Word	Joy	Word	Sadness
outraged	0.964	*horror*	0.923	*sohappy*	0.868	*sad*	0.844
brutality	0.959	*horrified*	0.922	*superb*	0.864	*suffering*	0.844
satanic	0.828	*hellish*	0.828	*cheered*	0.773	*guilt*	0.750
hate	0.828	*grenade*	0.828	*positivity*	0.773	*incest*	0.750
violence	0.742	*strangle*	0.750	*merrychristmas*	0.712	*accursed*	0.697
molestation	0.742	*tragedies*	0.750	*bestfeeling*	0.712	*widow*	0.697
volatility	0.687	*anguish*	0.703	*complement*	0.647	*infertility*	0.641
eradication	0.685	*grisly*	0.703	*affection*	0.647	*drown*	0.641
cheat	0.630	*cutthroat*	0.664	*exalted*	0.591	*crumbling*	0.594
agitated	0.630	*pandemic*	0.664	*woot*	0.588	*deportation*	0.594
defiant	0.578	*smuggler*	0.625	*money*	0.531	*isolated*	0.547
coup	0.578	*pestilence*	0.625	*rainbow*	0.531	*unkind*	0.547
overbearing	0.547	*convict*	0.594	*health*	0.493	*chronic*	0.500
deceive	0.547	*rot*	0.594	*liberty*	0.486	*injurious*	0.500
unleash	0.515	*turbulence*	0.562	*present*	0.441	*memorials*	0.453
bile	0.515	*grave*	0.562	*tender*	0.441	*surrender*	0.453
suspicious	0.484	*failing*	0.531	*warms*	0.391	*beggar*	0.422
oust	0.484	*stressed*	0.531	*gesture*	0.387	*difficulties*	0.421
ultimatum	0.439	*disgusting*	0.484	*healing*	0.328	*perpetrator*	0.359
deleterious	0.438	*hallucination*	0.484	*tribulation*	0.328	*hindering*	0.359

Table 2: Example entries for four emotions in the NRC Affect Intensity Lexicon. For each emotion, the table shows every 100th and 101st entry, when ordered by decreasing emotion intensity.