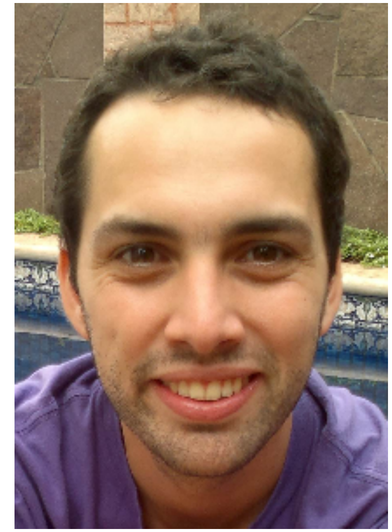
Emotion Intensities in Tweets

Saif M. Mohammad
National Research Council Canada

Felipe Bravo-Marquez
University of Waikato



Felipe Bravo-Marquez
University of Waikato

Postdoctoral research fellow in the Machine Learning Group at the University of Waikato.

Acknowledgments: We thank Svetlana Kiritchenko for helpful discussions. We thank Samuel Larkin for helping with data collection.

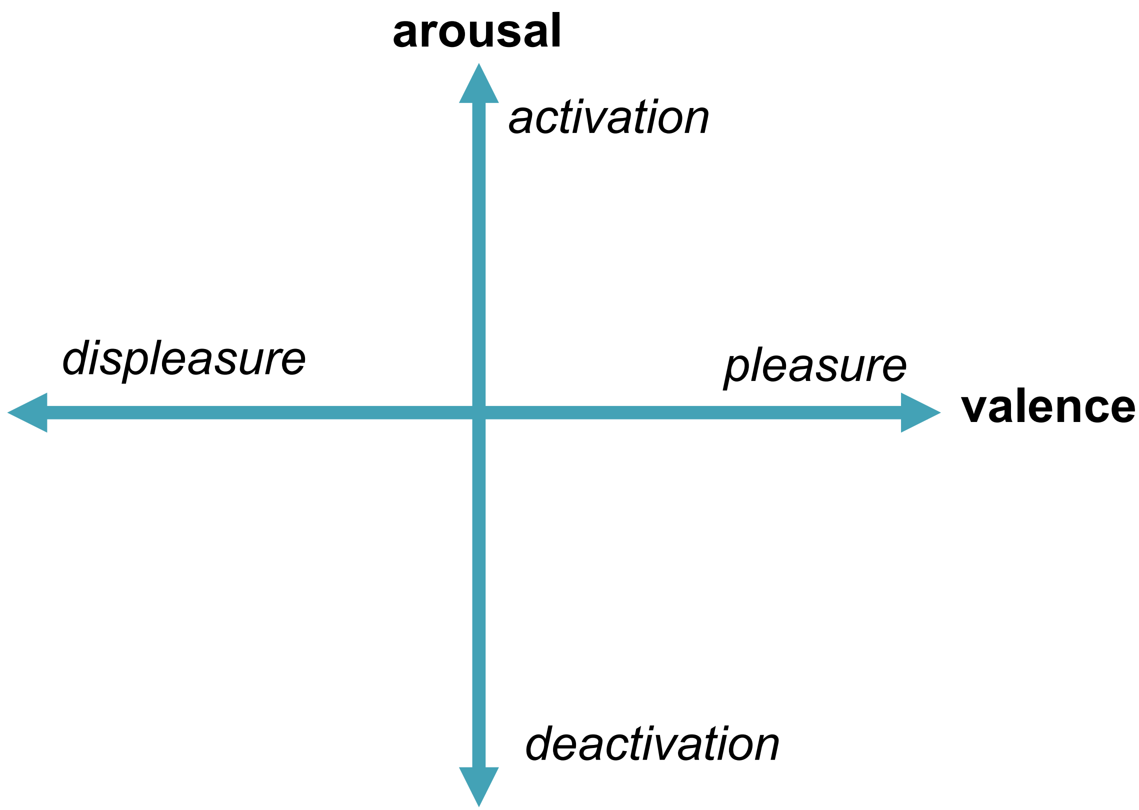
Emotions


Relaxed
Worry
Hurt
Fear
Courage
Delight
Joy
Surprised
Relieved
Anxiety
Content
Affection
Pleasure
Powerlessness
Pride
Annoyance
Shame
Envy
Embarrassment
Stress
Sadness
Contempt
Excitement
Empathy
Anger
Amusement
Helplessness
Love
Satisfaction
Friendliness
Disappointment
Irritation
Happiness
Boredom
Guilt
Despair
Elation
Politeness
Calm
Frustration
Disgust
Doubt
Serene
Hope
Interest
Tension
Trust
Shock

Dimensional Model of Emotions

- small number of dimensions
- emotion is point in the multi-dimensional space

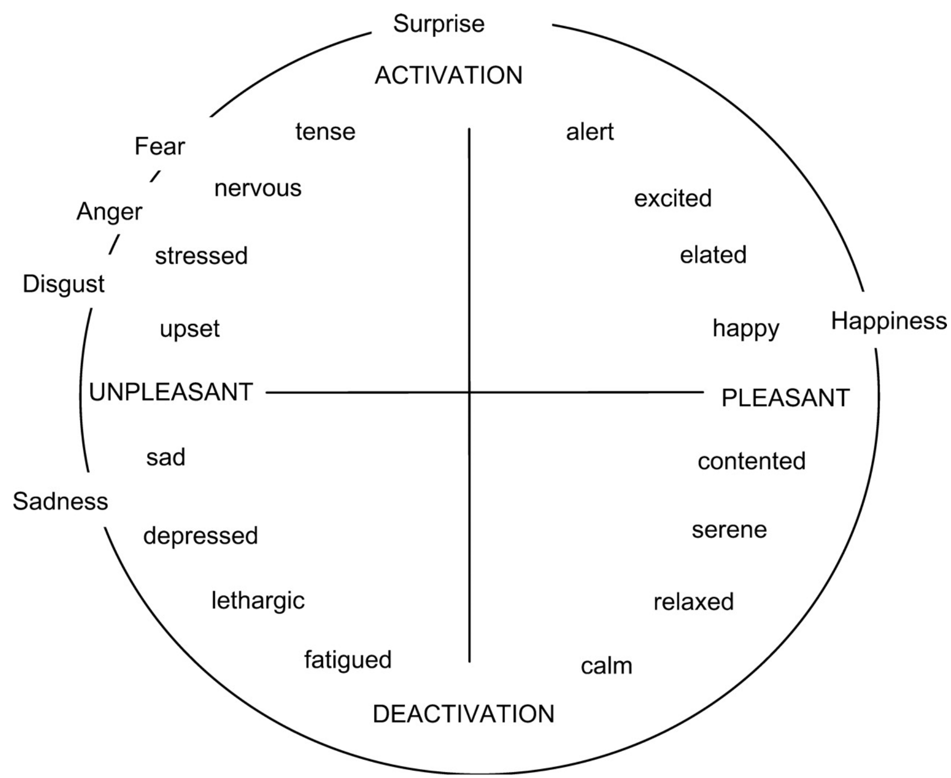
Circumplex Model (Russell, 1980)



Dimensional Model of Emotions

- small number of dimensions
- emotion is point in the multi-dimensional space

Circumplex Model (Russell, 1980)



Categorical Model of Emotions

- a handful of basic emotions

Ekman (1971): 6 basic emotions, Plutchik (1980): 8 basic emotions

- anger

not angry at all → *extremely angry*

Intensity of Anger

- joy

not happy at all → *extremely happy*

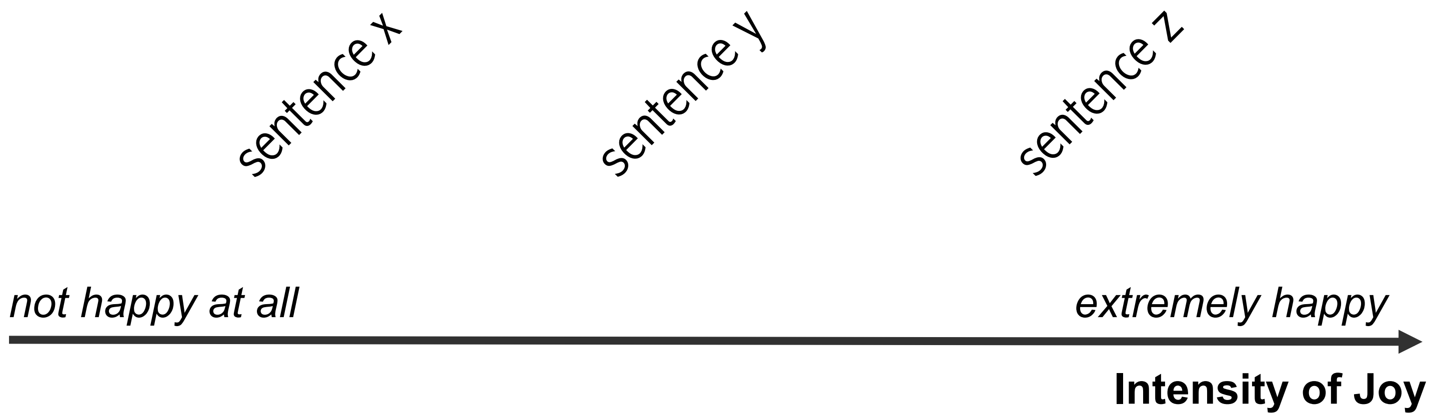
Intensity of Joy

- …

National Research Council Canada

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

We use language to communicate not only the category of of the emotion but also the intensity.



Here, intensity refers to the degree or amount of an emotion such as anger, sadness, or joy.

Why model emotion intensity?

- Natural language applications benefit from knowing both the class of emotion and its intensity
- Commercial customer satisfaction system
  - significant frustration or anger vs. instances of minor inconvenience

However, most work on automatic emotion detection has focused on categorical classification:

- built models for presence of anger, joy, sadness, etc.
- lack of data annotated for intensity

Challenges in Annotating Emotion Intensity

- Respondents are presented with greater cognitive load
- Particularly hard to ensure consistency
  - both across responses by different annotators, and
  - within the responses produced by an individual annotator

We present work on detecting and analyzing fine-grained emotion intensities from tweets.

- manually
- automatically

Emotion Intensity Task

Given:

- a tweet
- an emotion X (anger, fear, joy, or sadness)

Task: determine the intensity or degree of emotion X felt by the speaker—a real-valued score between 0 and 1.

- A score of 1 means that the speaker feels the highest amount of emotion X.
- A score of 0 means that the speaker feels the lowest amount of emotion X.

(There are other ways in which the intensity task can be framed.)

DATA

- compiling tweets
- annotating for emotion intensity
- analysis

Query Terms

For each emotion X,

- 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, antagonism,* and so on.
    - for sadness: *sad, devastated, sullen, down, crying, dejected, heartbroken, grief, weeping*, and so on.

Tweets

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

- Created a subset of the remaining tweets by:
  - selected at most 50 tweets per query term
  - selected at most 1 tweet for every tweeter–query term combination

Hashtags in Tweets

- To study the impact of emotion word hashtags on the intensity of the tweet
  - identified tweets that had a query term in hashtag form towards the end of the tweet

    This mindless support of a demagogue needs to stop. #racism #angry
    Hashtag Query Term Tweet (HQT Tweet)

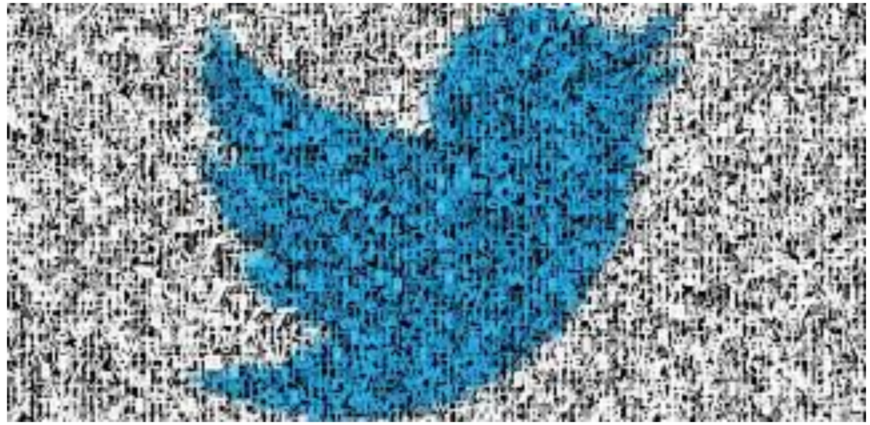
  - created copies of these tweets and then removed the hashtag query terms from the copies

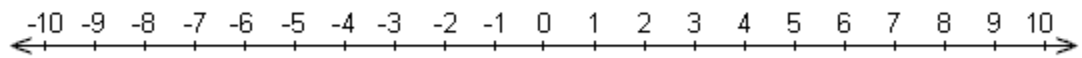
    This mindless support of a demagogue needs to stop. #racism
    No Query Term Tweet (NQT Tweet)

Master Tweets Set

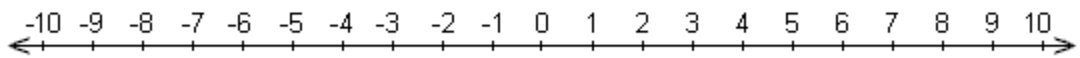
Includes 7,097 tweets:

- 1030 Hashtag Query Term Tweets (HQT Tweets)
- 1030 No Query Term Tweets (NQT Tweets)
- 5037 remaining tweets





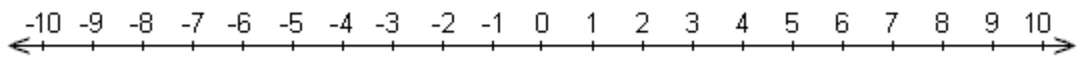
How to capture fine-grained affect intensity associations reliably?



How to capture fine-grained affect intensity associations reliably?

Ranking Jelly Bean Flavours

- Black Pepper
- Booger
- Dirt
- Earthworm
- Earwax
- Rotten Egg
- Sausage
- Soap
- …



How to capture fine-grained affect intensity associations reliably?

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

Comparative Annotations

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

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 annotations
  - less biased and more discriminating (Kiritchenko and Mohammad, 2017, Cohen, 2003)

Best–Worst Scaling (BWS)

with example from Kiritchenko et al. 2014

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

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

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

Our Example BWS Annotation Instance:
for tweet emotion intensity

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

Q2. Which of the four speakers is likely to be the LEAST SAD

Best–Worst Scaling

- Multiple sets of 2N 4-tuples generated randomly
  - that set chosen which maximizes tuple diversity
  - each item is seen in ~8 different 4-tuples
  - no pair of items occurs in more than one 4-tuple
- Each of the 4-tuples presented to 3 annotators
- A real-valued score for all the terms is determined from the BWS annotations (Orme, 2009)

$$score(t) = \%best(t) - \%worst(t)$$

the scores linearly transformed to the 0 to 1 range:
0 (lowest emotion intensity)
to 1 (highest emotion intensity)

  - the scores can then be used to rank the tweets

Tweet Emotion Intensity (TEI) Dataset

- ~7K tweets: 1500 to 2200 tweets per emotion
  - anger, fear, joy, sadness
- For machine learning experiments
  - about 50% of the tweets in the training set
  - about 5% in the development set
  - about 45% in the test set

Interactive Visualization: Tweet Emotion Intensity Dataset

http://saifmohammad.com/WebPages/TweetEmotionIntensity-dataviz.html

% by Emotion

Emotion	
anger	23.97%
fear	31.73%
joy	22.70%
sadness	21.60%
Grand Total	100.00%

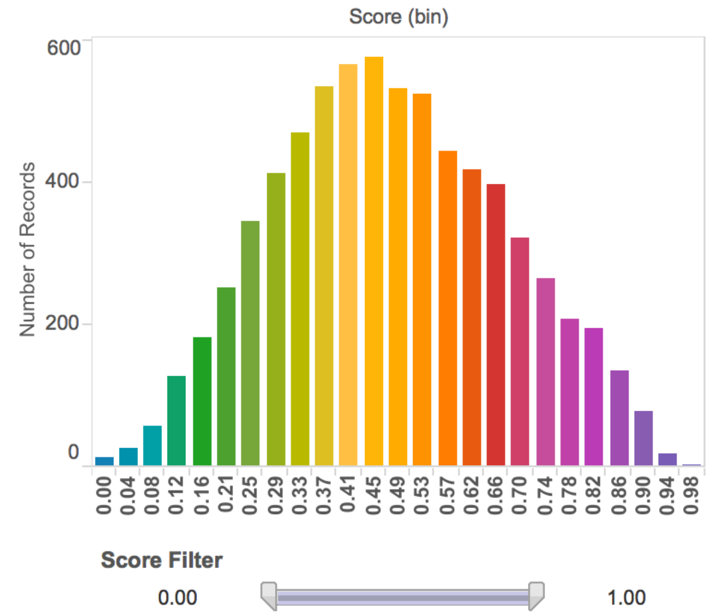
% by train-dev-test

Testflag	
train	50.91%
dev	4.82%
test	44.27%
Grand Total	100.00%

% by emotion, train-dev-test

Emotion	Testflag	
anger	train	50.38%
	dev	4.94%
	test	44.68%
	Total	100.00%
fear	train	50.93%
	dev	4.88%
	test	44.18%
	Total	100.00%
joy	train	51.09%
	dev	4.59%
	test	44.32%
	Total	100.00%
sadness	train	51.27%
	dev	4.83%
	test	43.90%
	Total	100.00%

Intensity Histogram



Tweets

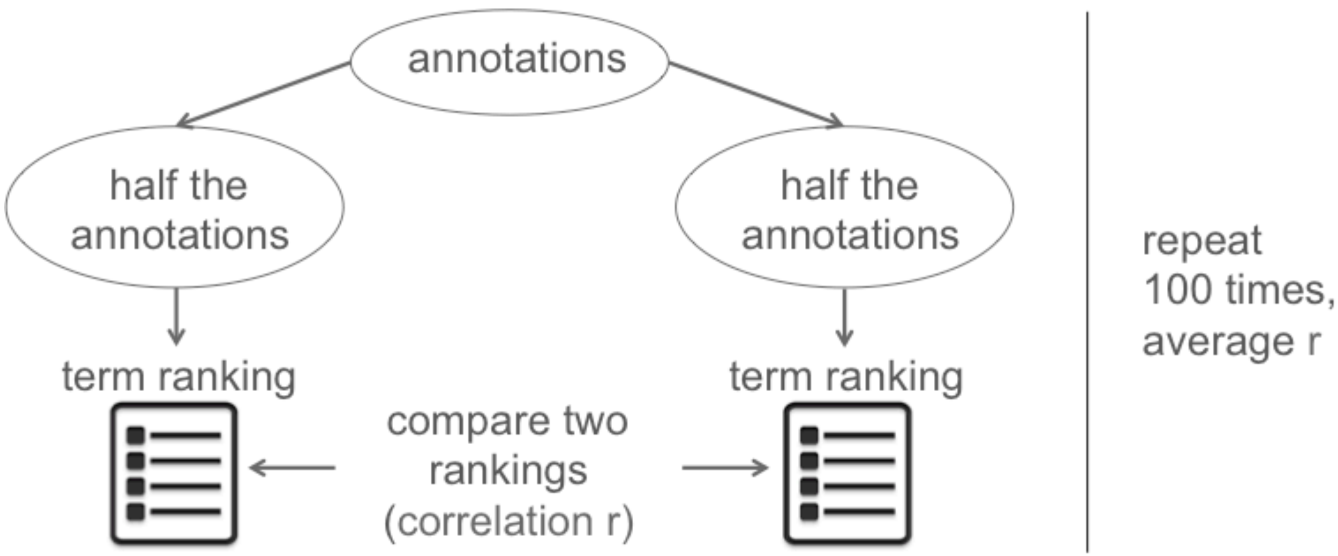
Id	Tweet	Emotion	Score
10000	How the fu*k! Who the heck! moved my fridge!... should I knock the landlord door. #angry #mad ##	anger	0.94
10001	So my Indian Uber driver just called someone the N word. If I wasn't in a moving vehicle I'd have jumped out #disgusted	anger	0.90
10002	@DPD_UK I asked for my parcel to be delivered to a pick up store not my address #fuming #poorcustomerservice	anger	0.90
10003	so ef whichever butt wipe pulled the fire alarm in davis bc I was sound asleep #pissed #angry #upset #tired #sad #tired #h..	anger	0.90
10004	Don't join @BTCare they put the phone down on you, talk over you and are rude. Taking money out of my acc willynilly! #f..	anger	0.90

Measuring Quality of Annotations

- Less useful: standard inter-annotator agreement measures
  - when a tuple has two items that are close in emotion intensity
  - the disagreement is a useful signal for BWS
- More useful: a measure of reproducibility of the end result
  - repeat annotations
  - involve multiple respondents
  - if similar intensity rankings (and scores) are produced
    - one can be confident that the scores capture the true emotion intensities.

Reliability (Reproducibility) of Annotations

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



Average SHR for the TEI Dataset

Emotion	Spearman ρ	Pearson r
anger	0.779	0.797
fear	0.845	0.850
joy	0.881	0.882
sadness	0.847	0.847

For fear, joy, and sadness datasets:

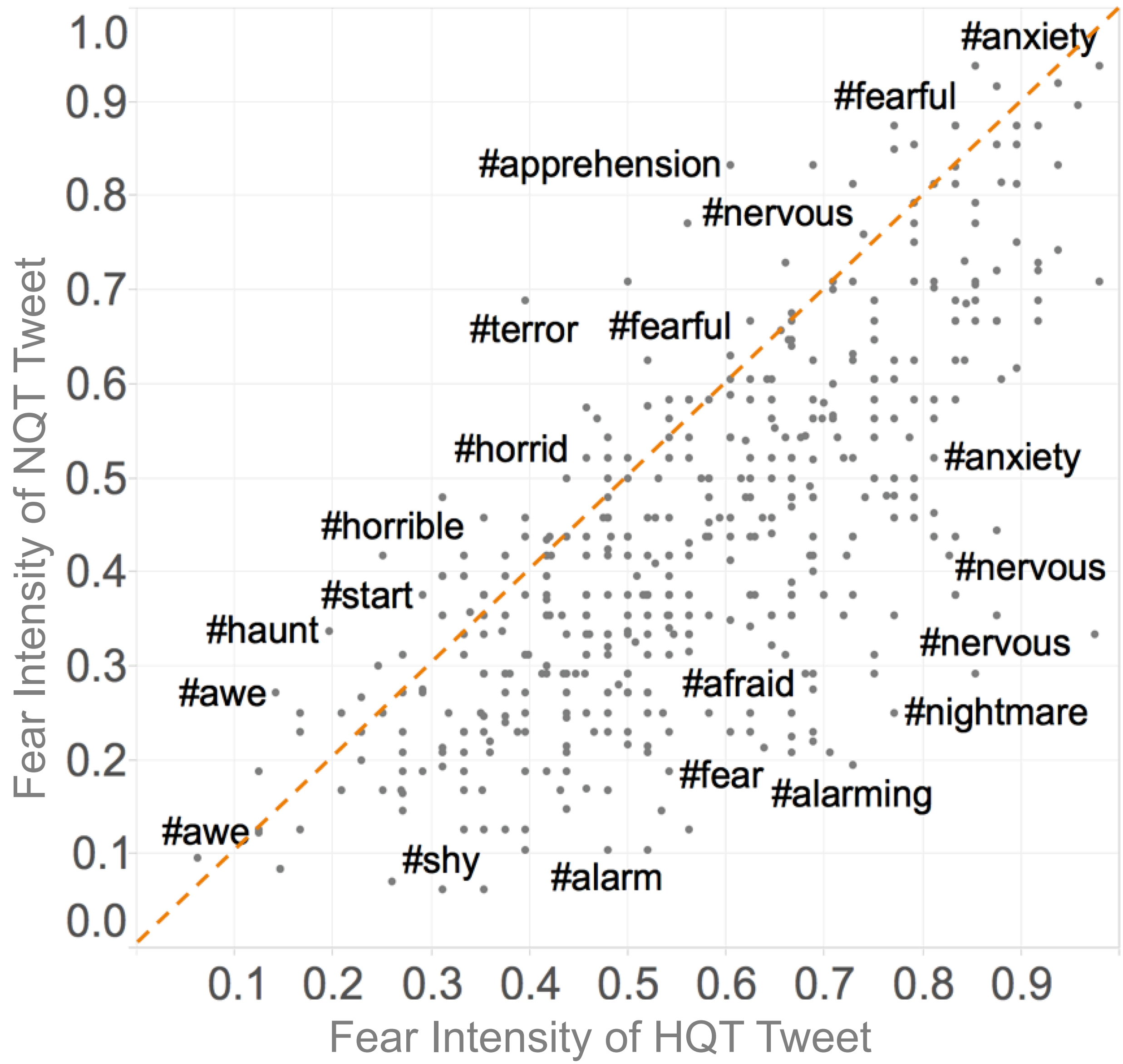
- r between 0.84 and 0.88, indicating a high degree of reliability
- the correlations are slightly lower for anger

Impact of Emotion Word Hashtags
on Emotion Intensity

This mindless support of a demagogue needs to stop. #racism #grrr #angry

Emotion	No. of HQT–NQT Tweet Pairs	% Tweets Pairs Drop	% Tweets Pairs Rise	% Tweets Pairs None	Average Emotion Intensity Score HQT tweets	Average Emotion Intensity Score NQT tweets	Average Emotion Intensity Score Drop	Average Emotion Intensity Score Rise
anger	282	76.6	19.9	3.4	0.58	0.48	0.15	0.07
fear	454	86.1	13.9	4.4	0.57	0.43	0.18	0.07
joy	204	71.6	26.5	1.9	0.59	0.50	0.15	0.09
sadness	90	85.6	11,1	3.3	0.65	0.49	0.19	0.05
All	1030	78.6	17.8	3.6	0.58	0.47	0.17	0.08

The impact of removal of emotion word hashtags on the emotion intensities of tweets (NQT-HQT subset of our dataset).



The scatter plot of fear intensity of HQT tweet vs. corresponding NQT tweet. As per space availability, some points are labeled with the rele- vant hashtag.

Impact of Emotion Word Hashtags
on Emotion Intensity

- Emotion word hashtags are often *not* redundant with the rest of tweet in terms of emotion intensity
- Often these hashtags increase emotion intensity
- Complex interplay between the text and the hashtag
  - if the rest of the tweet clearly indicates an emotion:
    - small change in the perceived emotion intensity
  - if the rest of the tweet is under-specified in terms of the emotion of the speaker:
    - marked increase in the perceived emotion intensity

AUTOMATICALLY DETERMINING TWEET EMOTION INTENSITY

- benchmark regression system and analysis
- quantifying similarity of emotions

AffectiveTweets Package for Weka

https://github.com/felipebravom/AffectiveTweets

- Provides a collection of filters for extracting features for sentiment analsyis and other related tasks
- Includes features used in:
  - Kiritchenko et al. (2014): sentiment analysis
  - Mohammad et al. (2017): stance detection

We use AffectiveTweets and Train Weka regression models

- LibLinear SVM regression

Features

- word n-grams (WN): presence or absence of word n-grams from n = 1 to n = 4
- character n-grams (CN): presence or absence of character n-grams from n = 3 to n = 5
- word embeddings (WE): an average of the word embeddings of all the words in a tweet
  - negative sampling skip-gram model implemented in Word2Vec (Mikolov et al., 2013)
  - word vectors are trained from ten million English tweets taken from the Edinburgh Twitter Corpus (Petrović et al., 2010)
  - window size: 5
  - number of dimensions: 400.

Features (continued)

- Affect Lexicons (L):
  - the number of words in the tweet matching each class are counted
  - sum individual scores for each class

	Twitter	Annotation	Scope	Label
AFINN (Nielsen, 2011)	Yes	Manual	Sentiment	Numeric
BingLiu (Hu and Liu, 2004)	No	Manual	Sentiment	Nominal
MPQA (Wilson et al., 2005)	No	Manual	Sentiment	Nominal
NRC Affect Intensity Lexicon (NRC-Aff-Int) (Mohammad, 2017)	Yes	Manual	Emotions	Numeric
NRC Word-Emotion Assn. Lexicon (NRC-EmoLex) (Mohammad and Turney, 2013)	No	Manual	Emotions	Nominal
NRC10 Expanded (NRC10E) (Bravo-Marquez et al., 2016)	Yes	Automatic	Emotions	Numeric
NRC Hashtag Emotion Association Lexicon (NRC-Hash-Emo) (Mohammad and Kiritchenko, 2015)	Yes	Automatic	Emotions	Numeric
NRC Hashtag Sentiment Lexicon (NRC-Hash-Sent) (Mohammad et al., 2013)	Yes	Automatic	Sentiment	Numeric
Sentiment140 (Mohammad et al., 2013)	Yes	Automatic	Sentiment	Numeric
SentiWordNet (Esuli and Sebastiani, 2006)	No	Automatic	Sentiment	Numeric
SentiStrength (Thelwall et al., 2012)	Yes	Manual	Sentiment	Numeric

Evaluation

- Pearson correlation coefficient (r)
  - scores produced by the automatic system on the test sets vs. the gold intensity scores
  - -1 (perfectly inversely correlated) to 1 (perfectly correlated)
  - a score of 0 indicates no correlation
- Spearman rank correlations were inline with the results obtained using Pearson

Results: r

system output vs gold

	anger	fear	joy	sad.	avg.
Individual feature sets					
word ngrams (WN)	0.42	0.49	0.52	0.49	0.48
char. ngrams (CN)	0.50	0.48	0.45	0.49	0.48
word embeds. (WE)	0.48	0.54	0.57	0.60	0.55
all lexicons (L)	**0.62**	**0.60**	**0.60**	**0.68**	**0.63**
Individual Lexicons					
AFINN	0.48	0.27	0.40	0.28	0.36
BingLiu	0.33	0.31	0.37	0.23	0.31
MPQA	0.18	0.20	0.28	0.12	0.20
NRC-Aff-Int	0.24	0.28	0.37	0.32	0.30
NRC-EmoLex	0.18	0.26	0.36	0.23	0.26
NRC10E	0.35	0.34	0.43	0.37	0.37
NRC-Hash-Emo	**0.55**	**0.55**	**0.46**	0.54	**0.53**
NRC-Hash-Sent	0.33	0.24	0.41	0.39	0.34
Sentiment140	0.33	0.41	0.40	0.48	0.41
SentiWordNet	0.14	0.19	0.26	0.16	0.19
SentiStrength	0.43	0.34	**0.46**	**0.61**	0.46
Combinations					
WN + CN + WE	0.50	0.48	0.45	0.49	0.48
WN + CN + L	0.61	0.61	0.61	0.63	0.61
WE + L	**0.64**	0.63	**0.65**	**0.71**	**0.66**
WN + WE + L	0.63	**0.65**	**0.65**	0.65	0.65
CN + WE + L	0.61	0.61	0.62	0.63	0.62
WN + CN + WE + L	0.61	0.61	0.61	0.63	0.62

National Research Council Canada

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Results: r
system output vs gold

- Features from affect lexicons: strongest single feature category
- NRC-Hash-Emo: best single lexicon (Mohammad, 2012; Mohammad and Kiritchenko, 2015)
- WE + L: best overall results

	anger	fear	joy	sad.	avg.
Individual feature sets					
word ngrams (WN)					0.48
char. ngrams (CN)					0.48
word embeds. (WE)					0.55
all lexicons (L)					**0.63**
Individual Lexicons					
AFINN					0.36
BingLiu					0.31
MPQA					0.20
NRC-Aff-Int					0.30
NRC-EmoLex					0.26
NRC10E					0.37
NRC-Hash-Emo					**0.53**
NRC-Hash-Sent					0.34
Sentiment140					0.41
SentiWordNet					0.19
SentiStrength					0.46
Combinations					
WN + CN + WE					0.48
WN + CN + L					0.61
WE + L					**0.66**
WN + WE + L					0.65
CN + WE + L					0.62
WN + CN + WE + L					0.62

Results: r

system output vs gold

- Features from affect lexicons: strongest single feature category
- NRC-Hash-Emo: best single lexicon (Mohammad, 2012; Mohammad and Kiritchenko, 2015)
- WE + L: best overall results

	anger	fear	joy	sad.	avg.
Individual feature sets					
word ngrams (WN)	0.42	0.49	0.52	0.49	0.48
char. ngrams (CN)	0.50	0.48	0.45	0.49	0.48
word embeds. (WE)	0.48	0.54	0.57	0.60	0.55
all lexicons (L)	**0.62**	**0.60**	**0.60**	**0.68**	**0.63**
Individual Lexicons					
AFINN	0.48	0.27	0.40	0.28	0.36
BingLiu	0.33	0.31	0.37	0.23	0.31
MPQA	0.18	0.20	0.28	0.12	0.20
NRC-Aff-Int	0.24	0.28	0.37	0.32	0.30
NRC-EmoLex	0.18	0.26	0.36	0.23	0.26
NRC10E	0.35	0.34	0.43	0.37	0.37
NRC-Hash-Emo	**0.55**	**0.55**	**0.46**	0.54	**0.53**
NRC-Hash-Sent	0.33	0.24	0.41	0.39	0.34
Sentiment140	0.33	0.41	0.40	0.48	0.41
SentiWordNet	0.14	0.19	0.26	0.16	0.19
SentiStrength	0.43	0.34	**0.46**	**0.61**	0.46
Combinations					
WN + CN + WE	0.50	0.48	0.45	0.49	0.48
WN + CN + L	0.61	0.61	0.61	0.63	0.61
WE + L	**0.64**	0.63	**0.65**	**0.71**	**0.66**
WN + WE + L	0.63	**0.65**	**0.65**	0.65	0.65
CN + WE + L	0.61	0.61	0.62	0.63	0.62
WN + CN + WE + L	0.61	0.61	0.61	0.63	0.62

Similarity of Emotions

Are all emotion pairs equally similar/ dissimialr?

Some emotions are closer to each other than others?

- which ones?
- can we quantify this?

Similarity of Emotions: In Language

Hypothesis: Some emotion pairs may have similar manifestions in language

- for example, similar words and expressions are used when expressing/describing both emotions

Experiment: We quantify this similarity of linguistic manifestation by using the Tweet Emotion Intensity dataset and the following experiment

- train regression system
  - with features WN + WE + L
  - on the training data for one emotion
- evaluate predictions on the test data for a different emotion

Correlation Results (r)

	Test On			
Train on	anger	fear	joy	sadness
anger	0.63	0.37	-0.37	0.45
fear	0.46	0.65	-0.39	0.63
joy	-0.41	-0.23	0.65	-0.41
sadness	0.39	0.47	-0.32	0.65

- The correlations are asymmetric
- All of the emotion pairs are correlated at least to some extent
  - diagonal: using training data for same emotion as test data
  - positive r: negative emotion with negative emotion
  - negative r: positive emotion with negative emotion
  - highest r: learning from fear and predicting sadness scores
    - r = 0.63 is close to the upper-bound (r = 0.65)

WASSA- 2017 Shared Task on Emotion Intensity

- The competition was organized on a CodaLab website: http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html
- Baseline system, AffectiveTweets package, released: https://github.com/felipebravom/AffectiveTweets
- Twenty-two teams participated.
  - Best system: ensemble of deep learning models (r = 0.74)
  - Top 3 teams: feature vector from the AffectiveTweets package

More details in the task-description paper (Mohammad and Bravo-Marquez, 2017)

Results

Our publicly released baseline: 0.66

Team Name	r avg. (rank)
1. Prayas	0.747 (1)
2. IMS	0.722 (2)
3. SeerNet	0.708 (3)
4. UWaterloo	0.685 (4)
5. IITP	0.682 (5)
6. YZU NLP	0.677 (6)
7. YNU-HPCC	0.671 (7)
8. TextMining	0.649 (8)
9. XRCE	0.638 (9)
10. LIPN	0.619 (10)
11. DMGroup	0.571 (11)
12. Code Wizards	0.527 (12)
13. Todai	0.522 (13)
14. SGNLP	0.494 (14)
15. NUIG	0.494 (14)
16. PLN PUCRS	0.483 (16)
17. H.Niemtsov	0.468 (17)
18. Tecnolengua	0.442 (18)
19. GradAscent	0.426 (19)
20. SHEF/CNN	0.291 (20)
21. deepCybErNet	0.076 (21)
Late submission	
∗ SiTAKA	0.631

	Team																					
Features	1	2	3	4	5	6	7	8	9	*	10	11	12	13	14	15	16	17	18	19	20	21
N-grams				✓									✓									
CN													✓									
WN				✓									✓			✓						
Word Embeddings	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓	✓	✓	✓				✓		
Glove			✓	✓	✓	✓	✓	✓		✓				✓		✓				✓		
Emoji Vectors			✓	✓																		
Word2Vec	✓	✓	✓	✓																		
Other								✓					✓		✓							
Sentence Embeddings																						
CNN	✓	✓				✓	✓	✓		✓					✓						✓	✓
LSTM	✓	✓			✓	✓	✓	✓						✓		✓				✓		
Other				✓												✓				✓	✓	
Affective Lexicons		✓	✓	✓	✓	✓		✓	✓	✓				✓				✓	✓	✓		
AFINN	✓	✓	✓		✓			✓														
ANEW		✓																				
BingLiu	✓	✓	✓		✓			✓	✓													
Happy Ratings		✓																				
Lingmotif																			✓			
LIWC																	✓					
MPQA	✓	✓	✓		✓			✓														
NRC-Aff-Int	✓		✓	✓				✓														
NRC-EmoLex	✓	✓	✓	✓	✓			✓	✓													
NRC-Emoticon-Lex	✓		✓	✓				✓					✓									
NRC-Hash-Emo	✓	✓	✓	✓	✓			✓	✓													
NRC-Hash-Sent		✓	✓	✓	✓			✓														
NRC-Hashtag-Sent.	✓		✓	✓																		
NRC10E	✓	✓	✓					✓														
Sentiment140	✓	✓	✓	✓				✓														
SentiStrength		✓	✓					✓														
SentiWordNet	✓	✓	✓	✓	✓			✓														
Vader					✓																	
Word.Affect			✓																			
In-house lexicon	✓								✓								✓					
Linguistic Features									✓													
Dependency Parser									✓													

Regression System

	Team																					
Regression	1	2	3	4	5	6	7	8	9	*	10	11	12	13	14	15	16	17	18	19	20	21
AdaBoost			✓																			
Gradient Boosting			✓	✓													✓					
Linear Regression				✓																		
Logistic Regression										✓									✓			
Neural Network	✓			✓		✓	✓	✓					✓	✓		✓				✓	✓	✓
Random Forest		✓	✓						✓													
SVM or SVR			✓	✓	✓								✓			✓	✓	✓		✓		
Ensemble	✓		✓										✓			✓				✓		

Summary: Created Affect Association Lexicons

- Created the first emotion intensity dataset for tweets
  - used best–worst scaling
  - applied to tweets (not just words) for the first time
- Showed that emotion-word hashtags often impact emotion intensity
  - often conveying a more intense emotion
- Created a benchmark regression system and conducted experiments
  - showed that affect lexicons are useful
  - especially those with fine word–emotion association scores such as the NRC Hashtag Emotion Lexicon
- Showed the extent to which emotion pairs are correlated
  - fear is strongly indicative of sadness

Ongoing Work

- SemEval-2018 Task#1: Affect in Tweets
  - Nine emotion categories
  - Valence, arousal, dominance
  - English, Arabic, Spanish
- Analyzing relationship between the VAD model and the categorical model of emotions
- Analyzing interplay between emotion intensity of words and emotion intensity of sentences/tweets
  - NRC Affect Intensity Lexicon: provides real-valued affect intensity scores for words

    Word Affect Intensities. Saif M. Mohammad. arXiv preprint arXiv:1704.08798, April 2017. http://saifmohammad.com/WebPages/AffectIntensity.htm
- Developing stronger emotion intensity models
- Multimodal emotion analysis

Resources available at shared task website:
http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html

- data
- annotation questionnaires
- evaluation scripts
- interactive visualizations of the data

AffectiveTweets Package
https://github.com/felipebravom/AffectiveTweets

Various affect lexicon available here:
http://saifmohammad.com/WebPages/AffectIntensity.htm

- NRC Hashtag Emotion lexicon
- NRC Affect Intensity Lexicon
- and others

Best-Worst Scaling resources available here:
http://saifmohammad.com/WebPages/BestWorst.html

- scripts and various BWS datasets

Contact:

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

Felipe Bravo-Marquez
fbravoma@waikato.ac.nz
www.cs.waikato.ac.nz/~fbravoma/