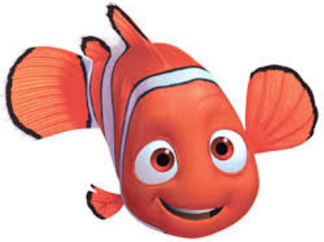


The Search for Emotions in Language

Saif M. Mohammad

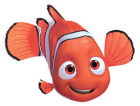
Senior Research Scientist
National Research Council Canada

Emotions

- Determine human experience and behavior
- Condition our actions
- Central in organizing meaning
  - No cognition without emotion

Outline

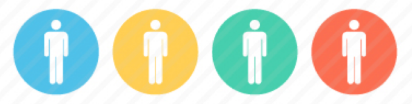
- Introduction
  - emotion and language

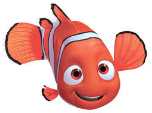


Outline

- Introduction
  - emotion and language

- The Search for Emotions (humans)
  - annotating words, sentences, tweets,…

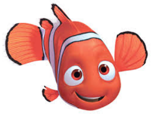




Outline

- Introduction
  - emotion and language

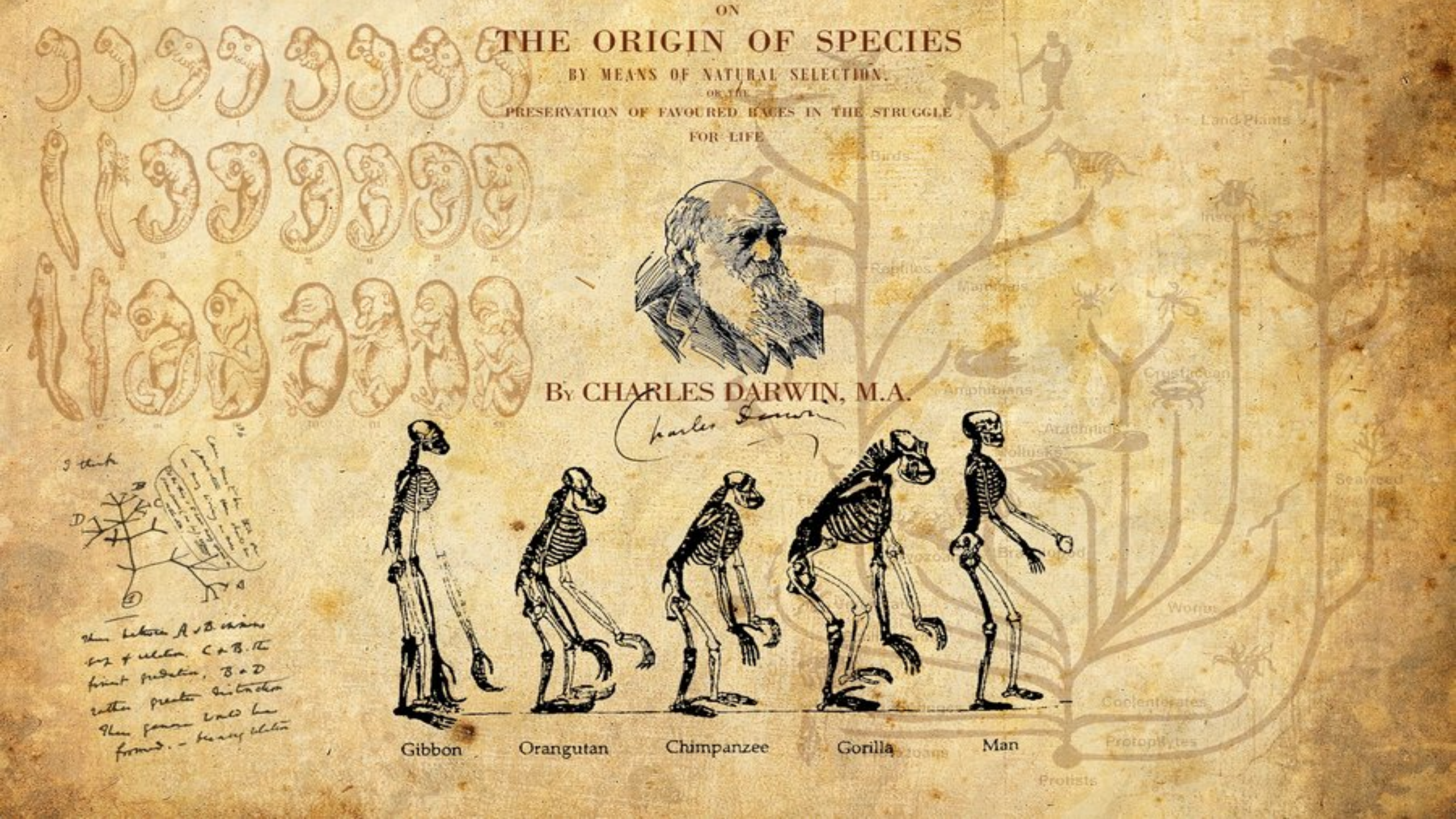
- The Search for Emotions (humans)
  - annotating words, sentences, tweets,…



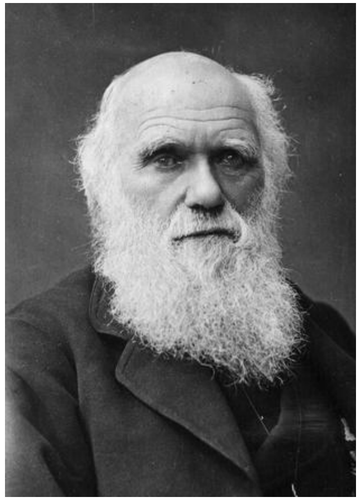
- The Search for Emotions (machines)
  - automatic systems for emotion, sentiment, stance, personality, music generation, argumentation,…

Introduction

Psychological Models of 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.
Charles Darwin
I think
Gibbon
Orangutan
Chimpanzee
Gorilla
Man
Land Plants
Birds
Reptiles
Insects
Amphibians
Crustacean
Arachnids
Mollusks
Seaweed
Worms
Coelenterates
Protophytes
Protists

Charles Darwin



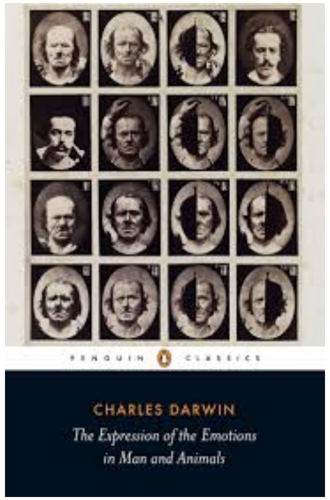
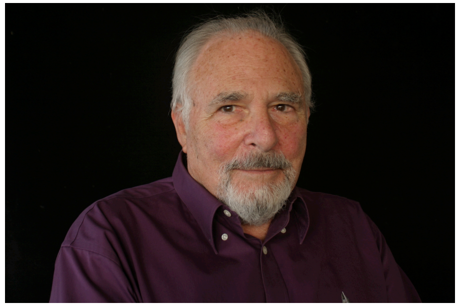


FIG. 20.—Terror,
from a photograph by Dr. Duchenne.

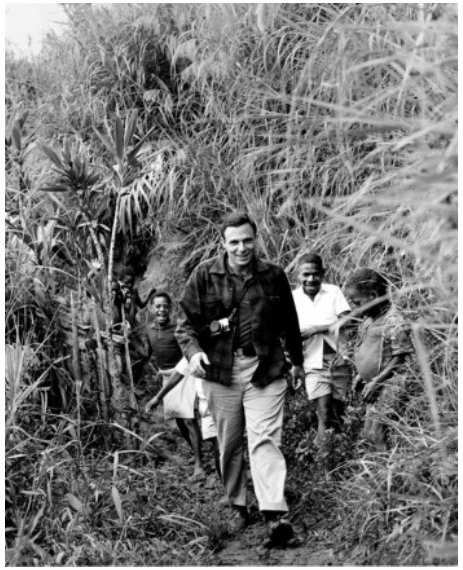
- published *The Expression of the Emotions in Man and Animals* in 1872
- seeks to trace the animal origins of human characteristics
  - pursing of the lips in concentration
  - tightening of the muscles around the eyes in anger
- claimed that certain facial expressions are universal
  - these facial expressions are associated with emotions

Debate: Universality of Perception of Emotions

Margaret Mead
Cultural anthropologist



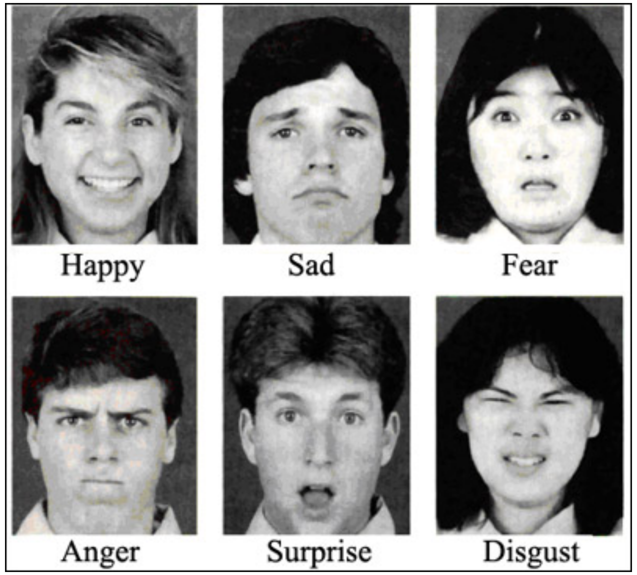
Paul Ekman
Psychologist and discoverer of micro expressions.



- Circa 1950's, Margaret Mead and others believed facial expressions and their meanings were culturally determined
  - behavioural learning processes
- Paul Ekman provided the strongest evidence to date that Darwin, not Margaret Mead, was correct in claiming facial expressions are universal
- Found universality of six emotions

Paul Ekman, 1971: Six Basic Emotions

- Anger
- Disgust
- Fear
- Joy
- Sadness
- Surprise



Plutchik, 1980: Eight Basic Emotions

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust

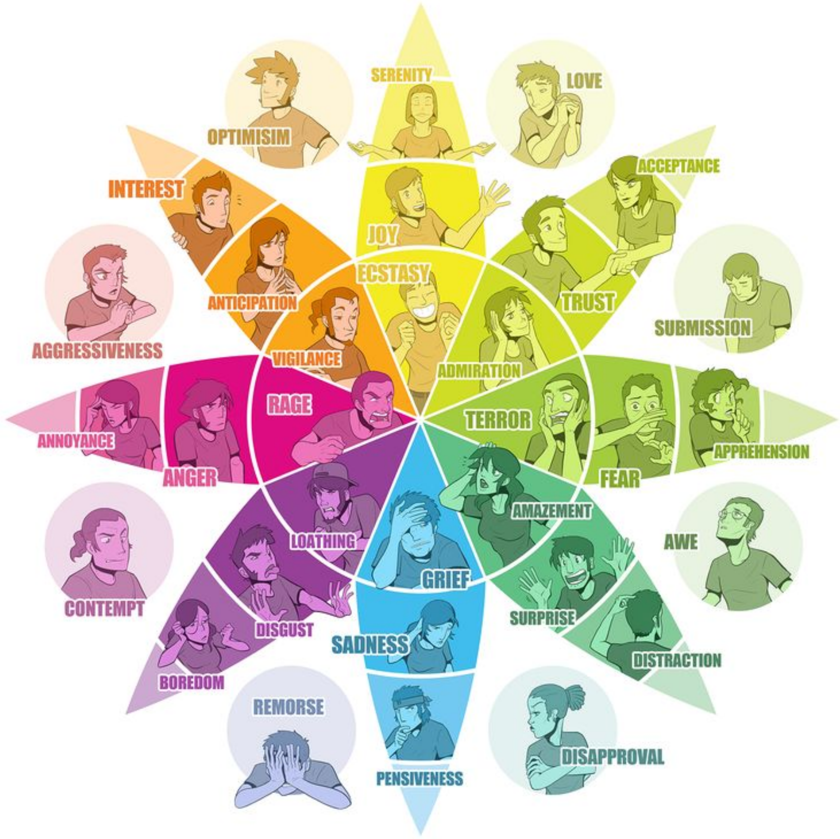


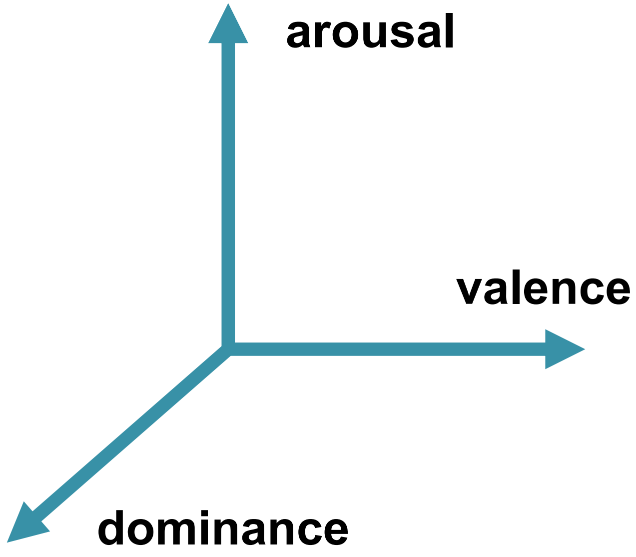
Image credit: Julia Belyanevych

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

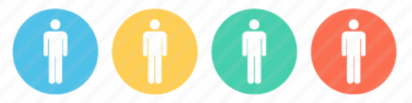


Psychological Models of Emotions

We annotate data for both:

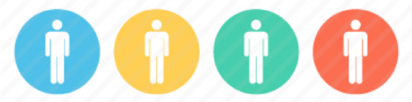
- the valence, arousal, and dominance model
- the basic emotions model

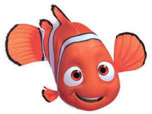
Motivation

Human annotations of words and tweets for emotions 

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

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





Finding Emotions (humans)

- annotating words, phrases, sentences, tweets
- crowdsourcing
- obtaining reliable fine-grained annotations

Word-Emotion Associations

Words have associations with emotions:

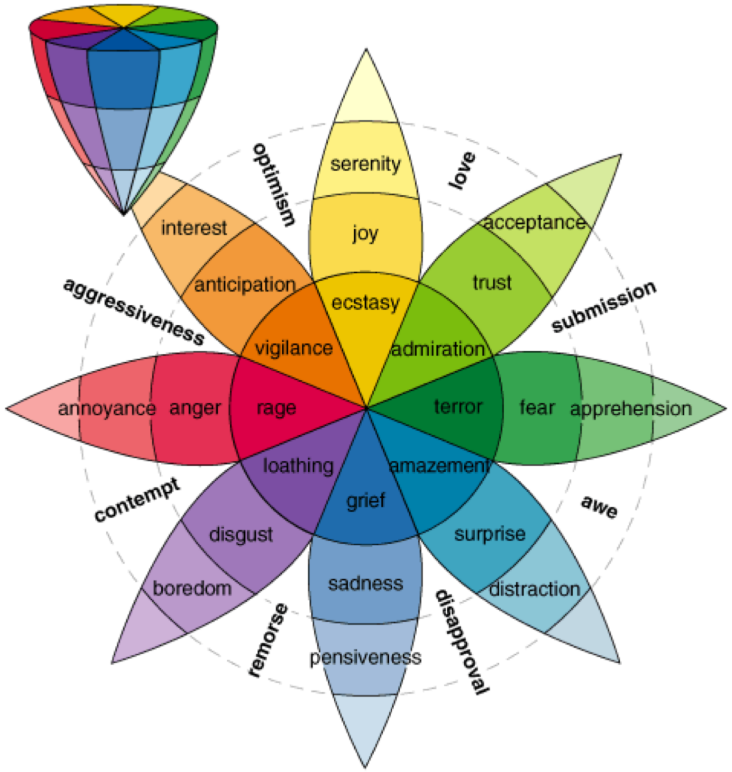
- attack and public speaking typically associated with fear
- yummy and vacation typically associated with joy
- loss and crying typically associated with sadness
- result and wait typically associated anticipation

Goal: Capture word-emotion associations.

Which Emotions?

Plutchik, 1980: Eight Basic Emotions

- Anger
- Anticipation
- Disgust
- Fear
- Joy
- Sadness
- Surprise
- Trust



Goal: We chose to capture word-emotion associations for the 8 Plutchik emotions.

Annotations by Crowdsourcing

- Benefits
  - Inexpensive
  - Scales well to large-scale annotations
- Challenges
  - Quality control
    - Malicious/random annotations
  - Words used in different senses are associated with different emotions.

Word-Choice Question

Peter Turney

Q1. Which word is closest in meaning to *cry*?

- *car*
- *tree*
- *tears*
- *olive*

- Generated automatically
  - Near-synonym taken from thesaurus
  - Distractors are randomly chosen
- Guides Turkers to desired sense
- Aids quality control
  - If Q1 is answered incorrectly:
    - Responses to the remaining questions for the word are discarded

Association Questions

Q2. How much is *cry* associated with the emotion sadness?

(for example, *death* and *gloomy* are strongly associated with sadness)

- *cry* is not associated with sadness
- *cry* is weakly associated with sadness
- *cry* is moderately associated with sadness
- *cry* is strongly associated with sadness

- Eight such questions for the eight basic emotions.
- Two such questions for positive or negative sentiment.

Better agreement when asked ‘associated with’ rather than ‘evoke’.

Emotion Lexicon

- NRC Emotion Lexicon
  - sense-level lexicon
    - word sense pairs: 24,200
  - word-level lexicon
    - union of emotions associated with different senses
    - word types: 14,200

Available at: www.saifmohammad.com

Paper:

Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence*, 29 (3), pages 436-465, 2013.

Use of The NRC Emotion Lexicon

- For research by the scientific community
  - Computational linguistics, psychology, digital humanities, robotics, public health research, etc.

- To analyze text
  - Brexit tweets, Radiohead songs, Trump tweets, election debates,…
  - **Wishing Wall**, uses the NRC Emotion lexicon to visualize wishes. Displayed in:
    - **Barbican Centre**, London, England, 2014
    - **Tekniska Museet**, Stockholm, Sweden, 2014
    - **Onassis Cultural Centre**, Athens, Greece, 2015
    - **Zorlu Centre**, Istanbul, Turkey, 2016

- In commercial applications

How to capture fine-grained affect intensity associations reliably?

Humans are not good at giving real-valued scores:

- hard to be consistent across multiple annotations
- difficult to maintain consistency across annotators
- 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

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)

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

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 positive sentiment)
  to 1 (most association with positive sentiment)

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

Comparative Annotations

Best–Worst Scaling (Louviere & Woodworth, 1990):

- 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)

Best-Worst Scaling Lexicons

Svetlana Kiritchenko
NRC

Lexicon	Language	Domain
1. Affect Intensity Lexicon	English	General
2. SemEval-2015 English Twitter Sentiment Lexicon	English	Twitter
3. SemEval-2016 Arabic Twitter Sentiment Lexicon	Arabic	Twitter
4. Sentiment Composition Lexicon for Negators, Modals, and Adverbs (SCL-NMA)	English	General
5. Sentiment Composition Lexicon for Opposing Polarity Phrases (SCL-OPP)	English	General

Lexicons and papers available at:
http://saifmohammad.com/WebPages/lexicons.html

Affect Intensity Lexicon: Example entries

Highest anger intensity:

outraged	0.964
brutality	0.959
hatred	0.953

Lowest anger intensity:

sisterhood	0.015
musical	0.011
tree	0.000

Highest fear intensity:

torture	0.984
terrorist	0.972
horrific	0.969

Lowest fear intensity:

volunteer	0.031
lines	0.031
romance	0.031

Scores are in the range 0 (lowest intensity) to 1 (highest intensity).

English Twitter Lexicon:
Examples sentiment scores obtained using BWS

Term	Sentiment Score -1 (most negative) to 1 (most positive)
awesomeness	0.827
#happygirl	0.625
cant waitttt	0.601
don't worry	0.152
not true	-0.226
cold	-0.450
#getagrip	-0.587
#sickening	-0.722

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 General Inquirer
- Terms from the Warriner et al. (2013) VAD lexicon
- Terms common in tweets

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**

number of pairs of best—worst annotations

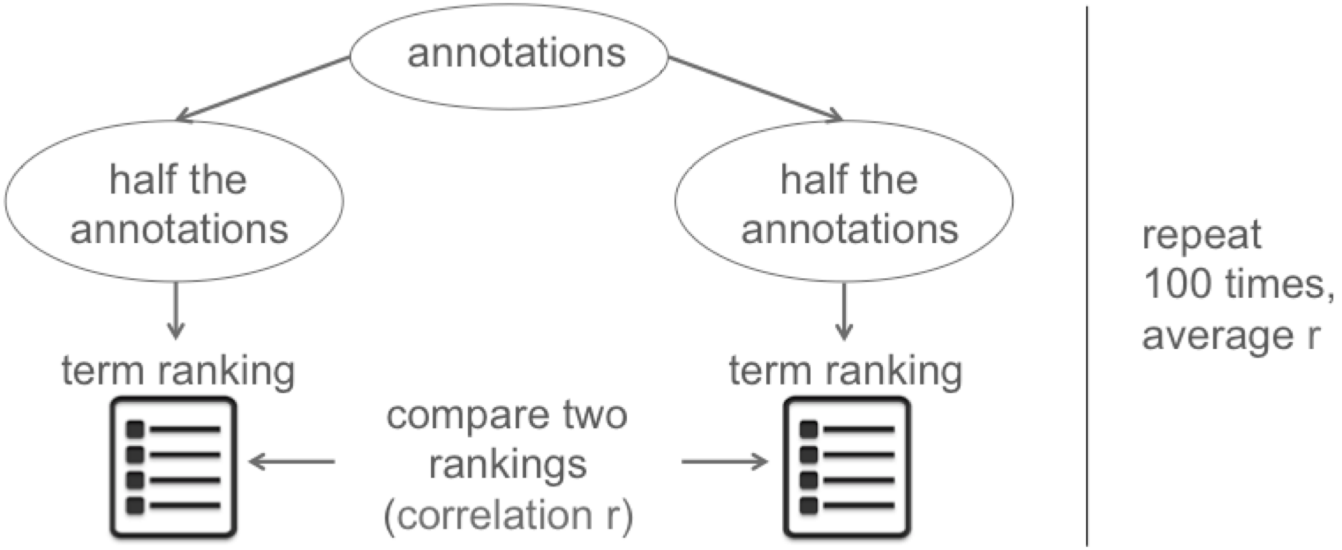
Example Entries in the VAD Lexicon

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).

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 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

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.

Papers:

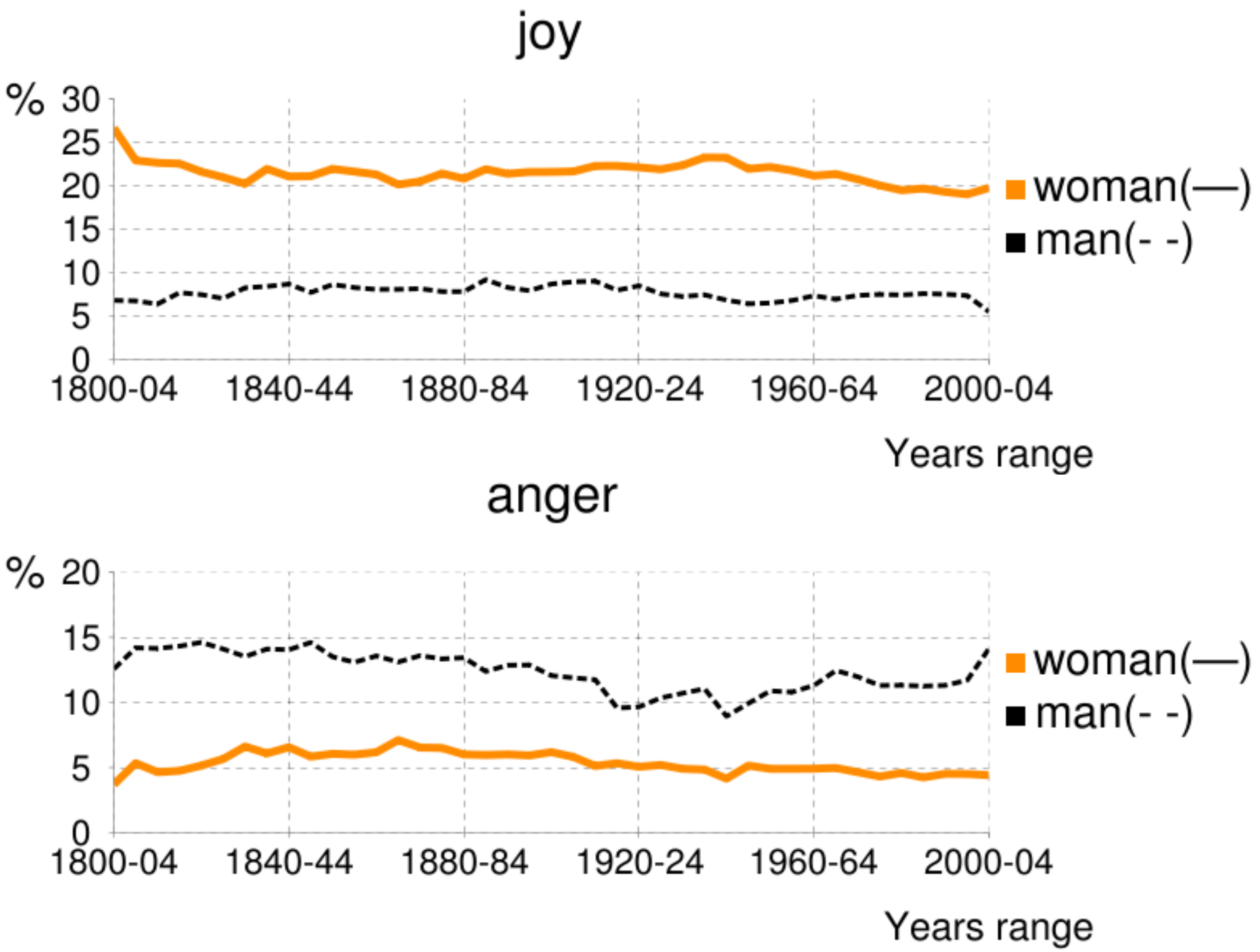
- Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best-Worst Scaling. Svetlana Kiritchenko and Saif M. Mohammad. In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. June 2016. San Diego, CA.
- Word Affect Intensities. Saif M. Mohammad. In Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), May 2018, Miyazaki, Japan.
- Sentiment Composition of Words with Opposing Polarities. Svetlana Kiritchenko and Saif M. Mohammad. In Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. June 2016. San Diego, CA.
- The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition. Svetlana Kiritchenko and Saif M. Mohammad, In Proceedings of the NAACL 2016 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA), June 2014, San Diego, California.
- Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In Proceedings of the International Workshop on Semantic Evaluation (SemEval '16). June 2016. San Diego, California.

Finding Emotions (machines)

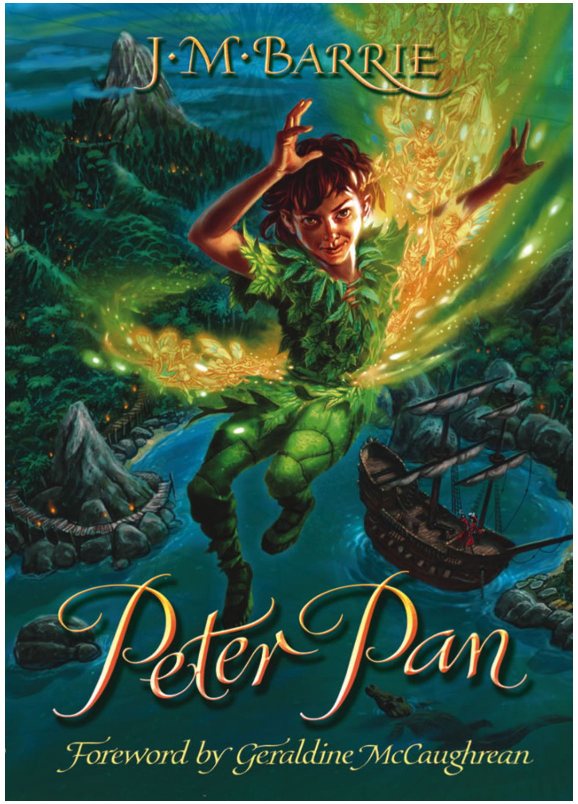
- automatic systems for emotion, sentiment, personality, literary analysis, music generation,…

Tony Yang, Simon Fraser University

Visualizing Emotions in Text



Percentage of joy and anger words in close proximity to occurrences of man and woman in books.

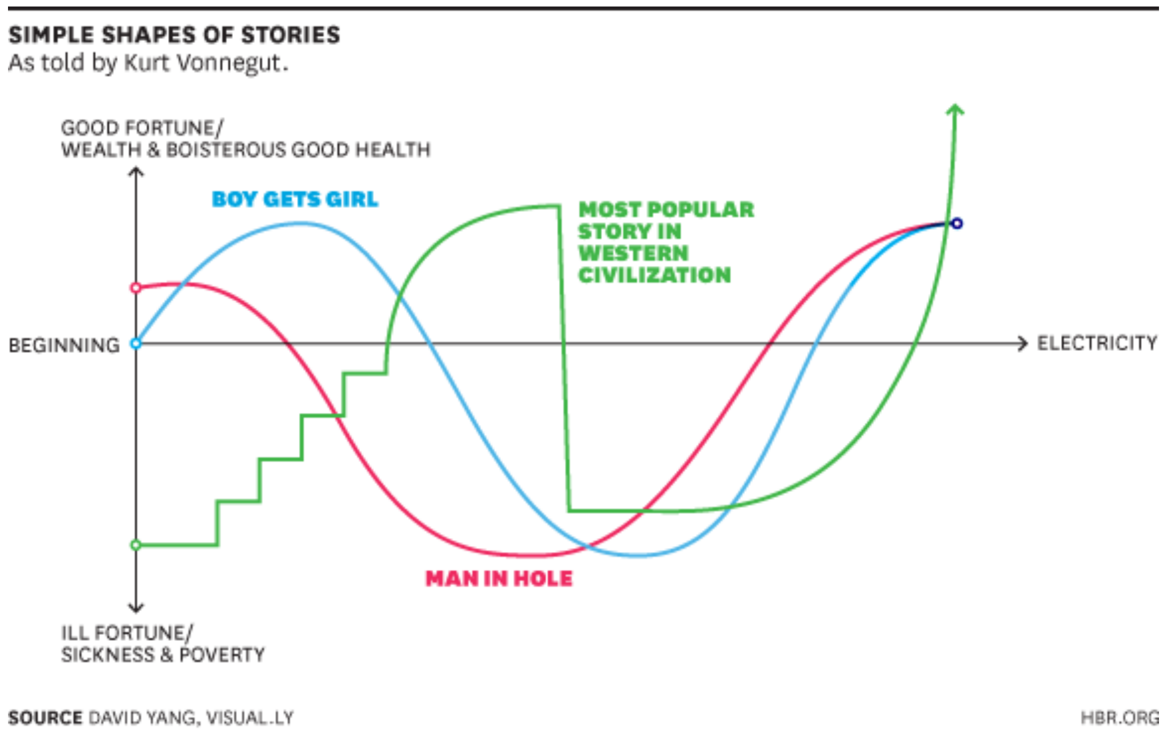


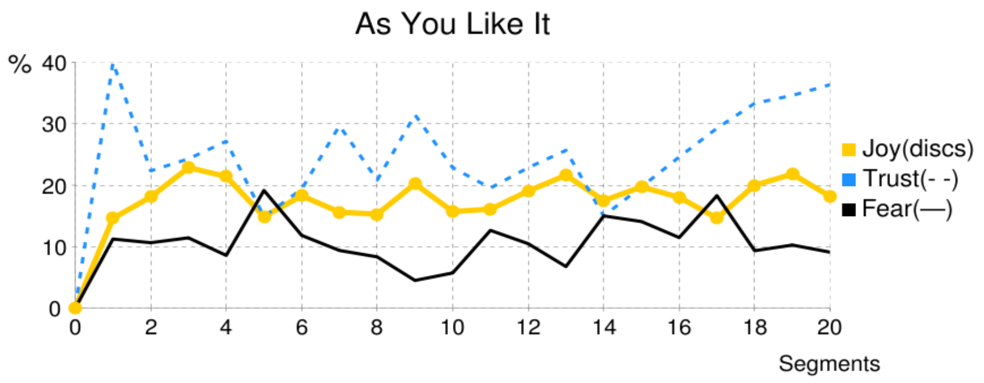
Stories

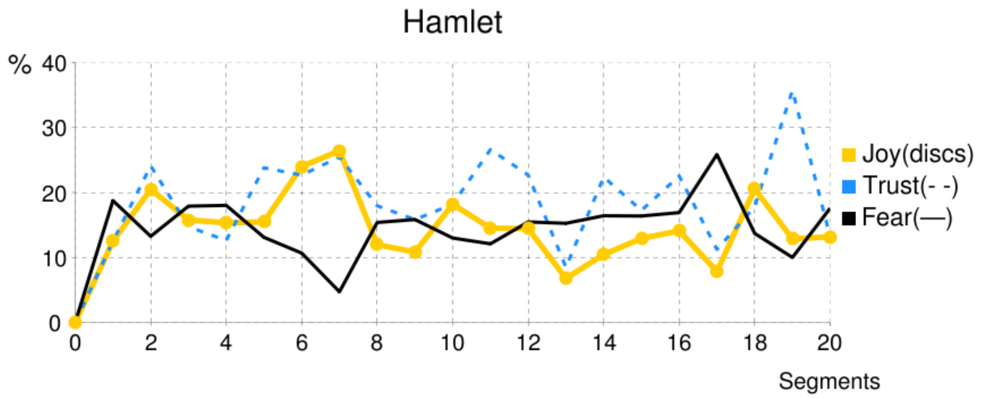
STORIES

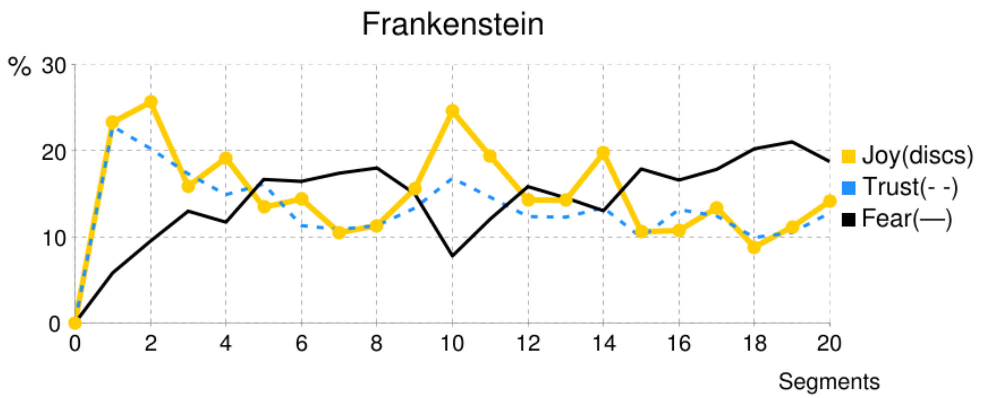
Tracking Emotions in Stories

- Can we automatically track the emotions of characters?
- Are there some canonical shapes common to most stories?
- Can we track the change in distribution of emotion words?




As You Like It
%
0
10
20
30
40
0
2
4
6
8
10
12
14
16
18
20
Segments
Joy(discs)
Trust(- -)
Fear(—)


Hamlet
%
0
10
20
30
40
0
2
4
6
8
10
12
14
16
18
20
Segments
Joy(discs)
Trust(- -)
Fear(—)


Frankenstein
%
0
10
20
30
0
2
4
6
8
10
12
14
16
18
20
Segments
Joy(discs)
Trust(- -)
Fear(—)

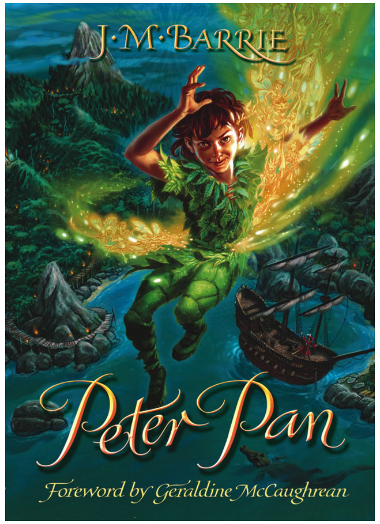
Work on shapes of stories

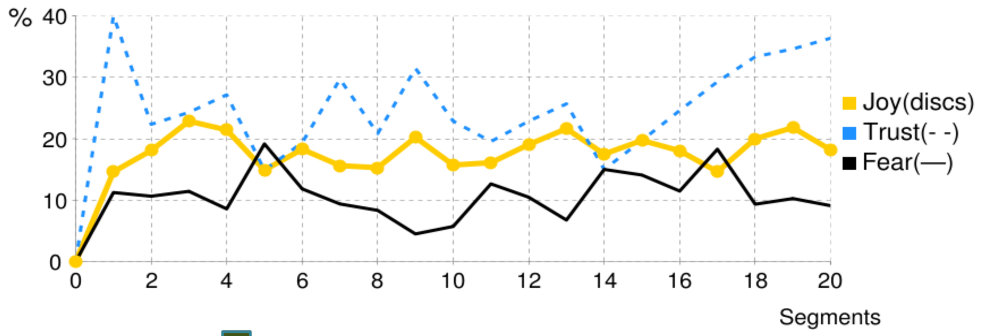
- From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.
- Character-based kernels for novelistic plot structure. Elsner, M., 2012, April. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 634-644). Association for Computational Linguistics.
- A novel method for detecting plot. M. Jockers http://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/, June 2014.
- The emotional arcs of stories are dominated by six basic shapes. Reagan, A.J., Mitchell, L., Kiley, D., Danforth, C.M. and Dodds, P.S., 2016. EPJ Data Science, 5(1), p.31.

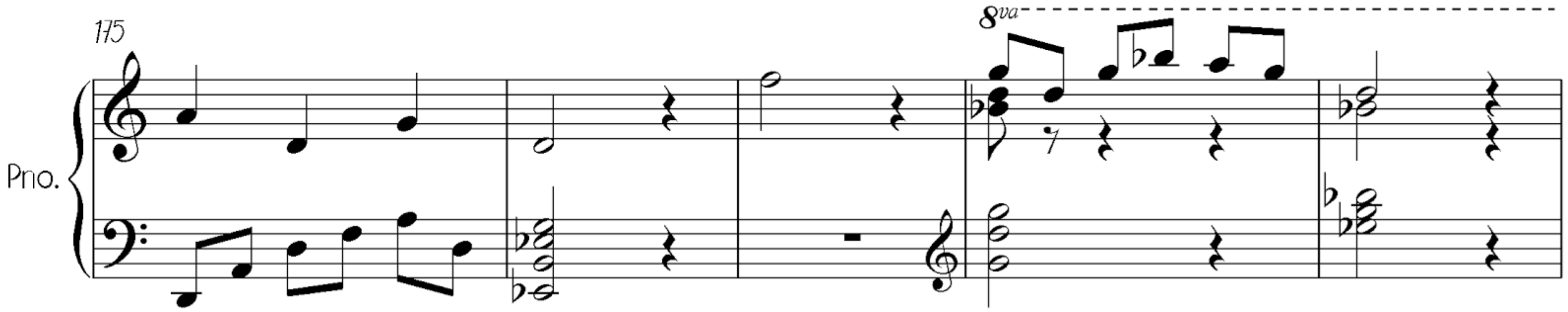
Generating music from text

Paper:

- Generating Music from Literature. Hannah Davis and Saif M. Mohammad, In Proceedings of the EACL Workshop on Computational Linguistics for Literature, April 2014, Gothenburg, Sweden.







A method to generate music from literature.

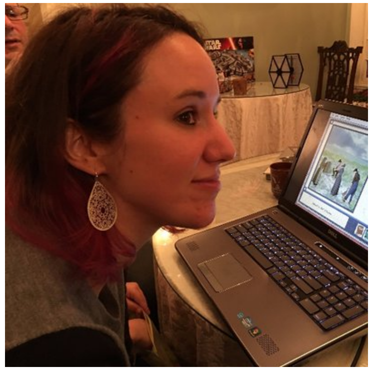
- music that captures the change in the distribution of emotion words.

Challenges

- Not changing existing music -- generating novel pieces
- Paralysis of choice
- Has to sound good
- No one way is the right way -- evaluation is tricky

Music-Emotion Associations

- Major and Minor Keys
  - major keys: happiness
  - minor keys: sadness
- Tempo
  - fast tempo: happiness or excitement
- Melody
  - a sequence of consonant notes: joy and calm
  - a sequence of dissonant notes: excitement, anger, or unpleasantness



Hannah Davis
Artist/Programmer

Hunter et al., 2010, Hunter et al., 2008, Ali and Peynirciolu, 2010,
Gabrielsson and Lindstrom, 2001, Webster and Weir, 2005

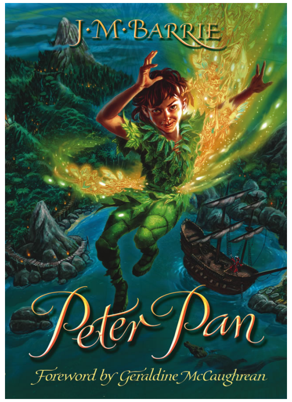
TransProse

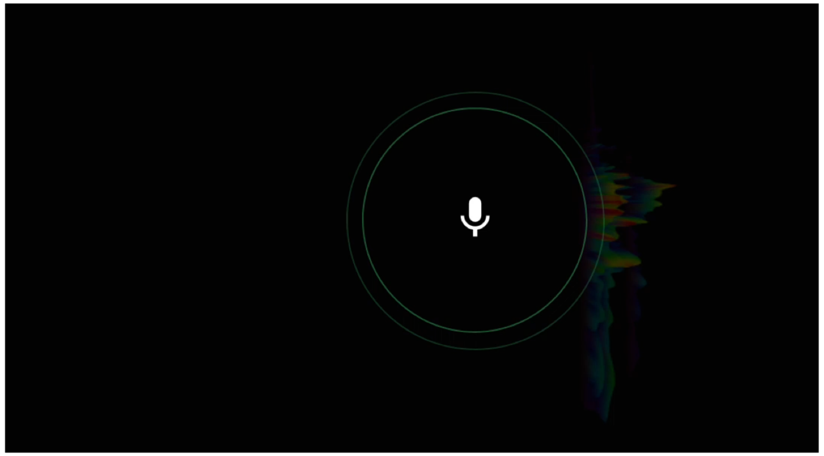
Automatically generates three simultaneous piano melodies pertaining to the dominant emotions in the text, using the NRC Emotion Lexicon.

TransProse

Automatically generates three simultaneous piano melodies pertaining to the dominant emotions in the text, using the NRC Emotion Lexicon.

Examples

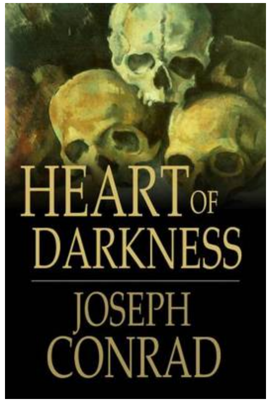




TransProse

Automatically generates three simultaneous piano melodies pertaining to the dominant emotions in the text, using the NRC Emotion Lexicon.

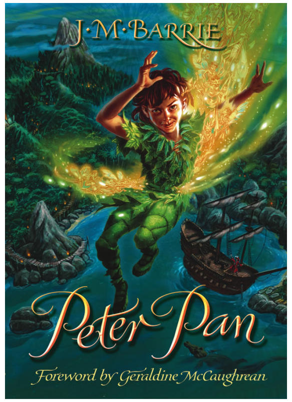
Examples

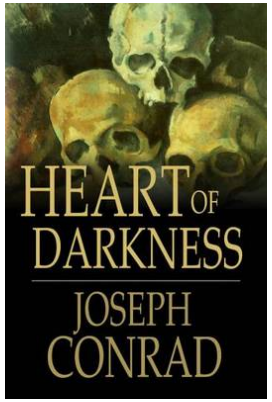


TransProse

Automatically generates three simultaneous piano melodies pertaining to the dominant emotions in the text, using the NRC Emotion Lexicon.

Examples





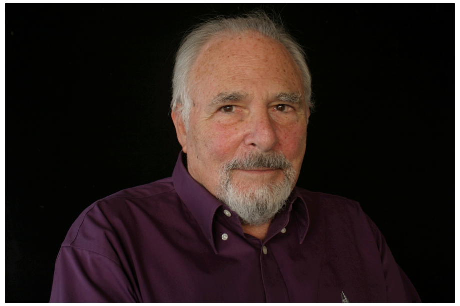
TransProse: www.musicfromtext.com
Music played 300,000 times since website launched in April 2014.

TransProse Music Played by an Orchestra, at the Louvre Museum, Paris

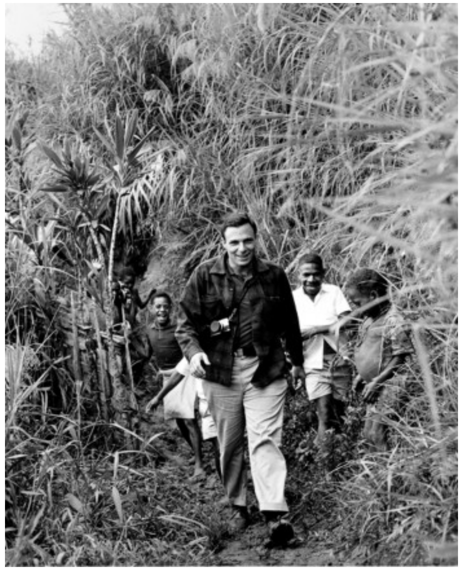
A symphony orchestra performs under the glass of the Louvre museum in Paris on Sept. 20. Accenture Strategy has created a symphonic experience enabled by human insight and artificial intelligence technology. (Michel Euler/AP)

Debate: Universality of Perception of Emotions

Margaret Mead
Cultural anthropologist



Paul Ekman
Psychologist and discoverer of micro expressions.



Lisa Barrett
University Distinguished Professor of Psychology, Northeastern University

- Grad school experiment on people’s ability to distinguish photos of depression from anxiety
  - one is based on sadness, and the other on fear
  - found agreement to be poor

Some Emotions more basic than others?

may be not…

Hashtagged Tweets

- Hashtagged words are good labels of sentiments and emotions

  Some jerk just stole my photo on #tumblr #grrr **#anger**

- Hashtags are not always good labels:
  - hashtag used sarcastically

    The reviewers want me to re-annotate the data. **#joy**

Paper:

#Emotional Tweets, Saif Mohammad, In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem), June 2012, Montreal, Canada.

Data to Model Hundreds of Emotions

Papers:

- Using Nuances of Emotion to Identify Personality. Saif M. Mohammad and Svetlana Kiritchenko, In *Proceedings of the ICWSM Workshop on Computational Personality Recognition*, July 2013, Boston, USA.
- Using Hashtags to Capture Fine Emotion Categories from Tweets. Saif M. Mohammad, Svetlana Kiritchenko, Computational Intelligence, Volume 31, Issue 2, Pages 301-326, May 2015.

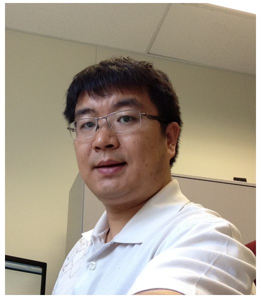
Sentiment Lexicons

Created a sentiment lexicon using a Turney (2003) inspired method that uses PMI of a word with co-occurring positive and negative seed hashtags.

Positive
spectacular 0.91
okay 0.3

Negative
lousy -0.74
murder -0.95

Svetlana Kiritchenko
NRC



Xiaodan Zhu
NRC

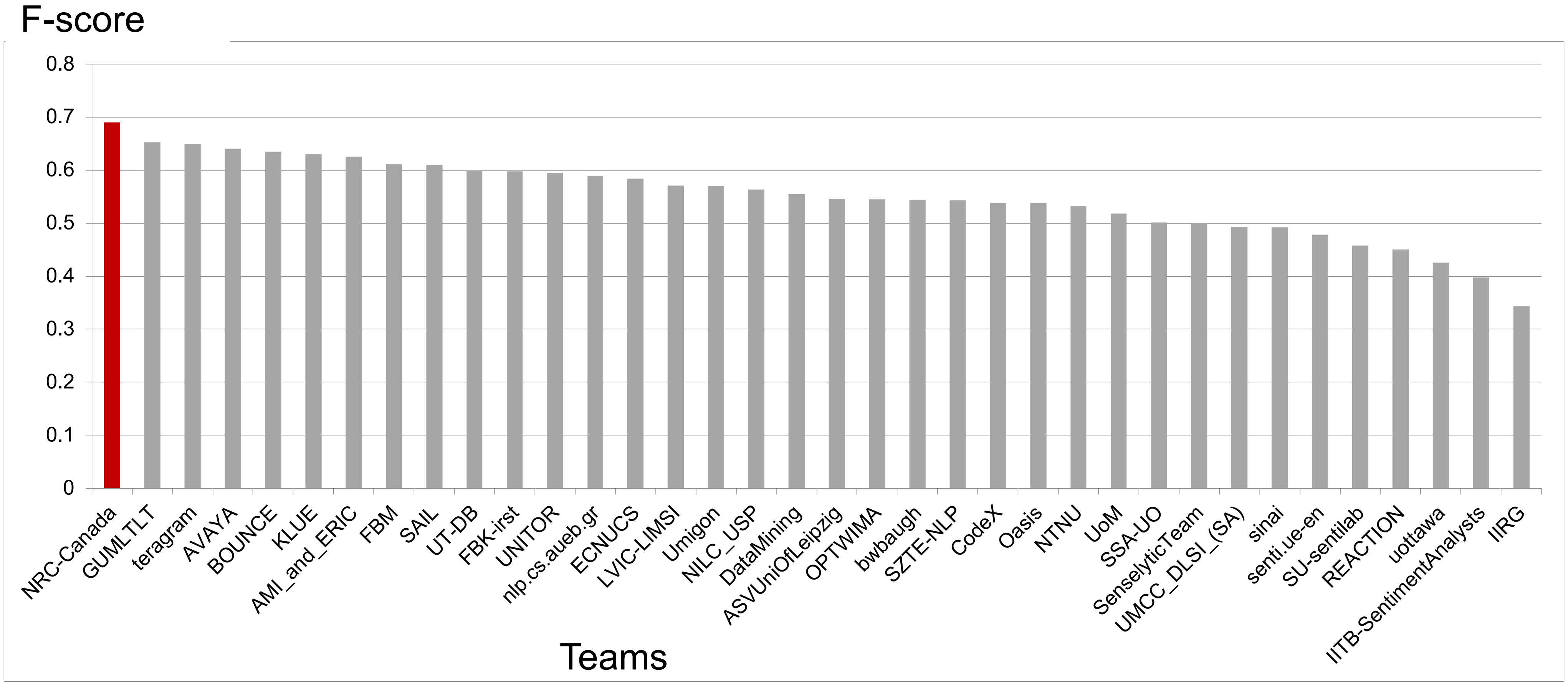
SemEval Shared task on the Sentiment Analysis of Tweets

Papers:

- Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. *Journal of Artificial Intelligence Research*, 50, August 2014.
- NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, June 2013, Atlanta, USA.

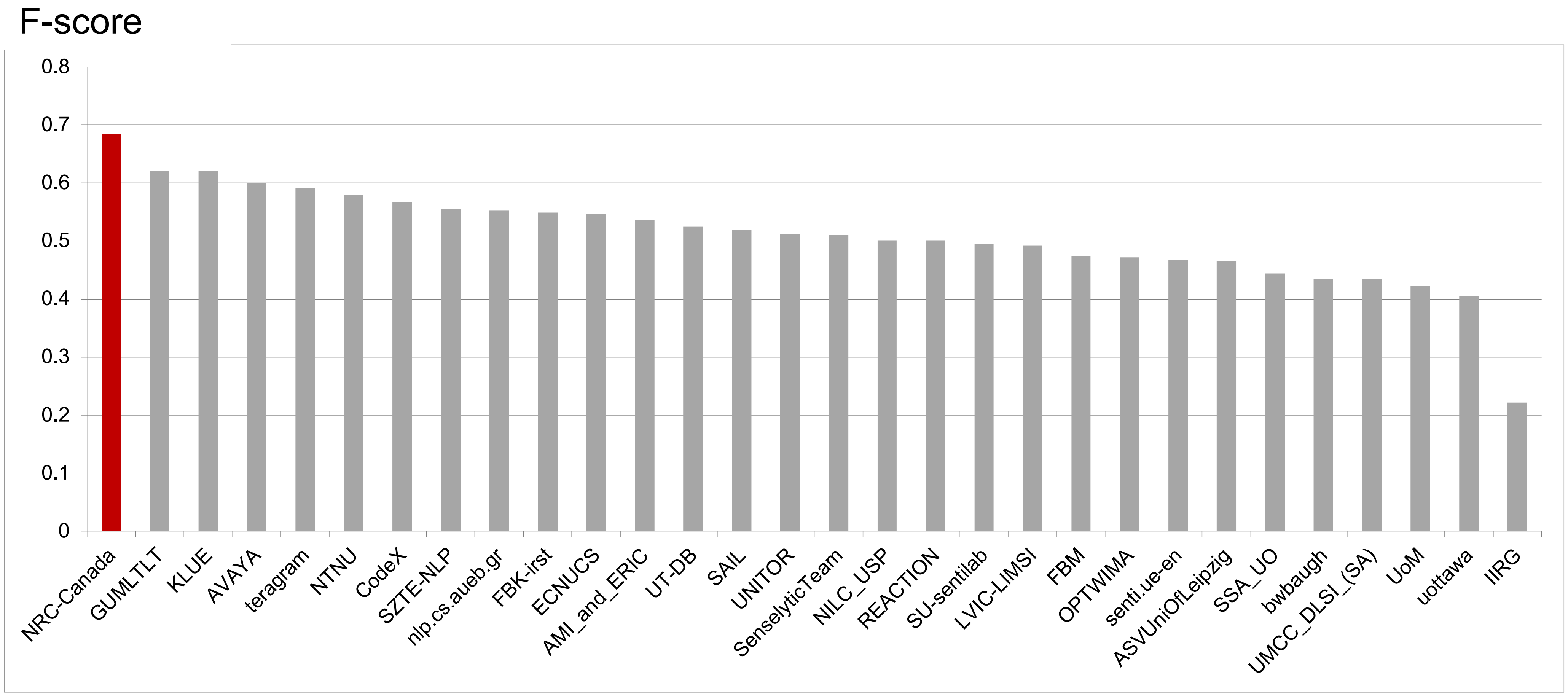
Sentiment Analysis Competition

SemEval-2013: Classify Tweets, 44 teams

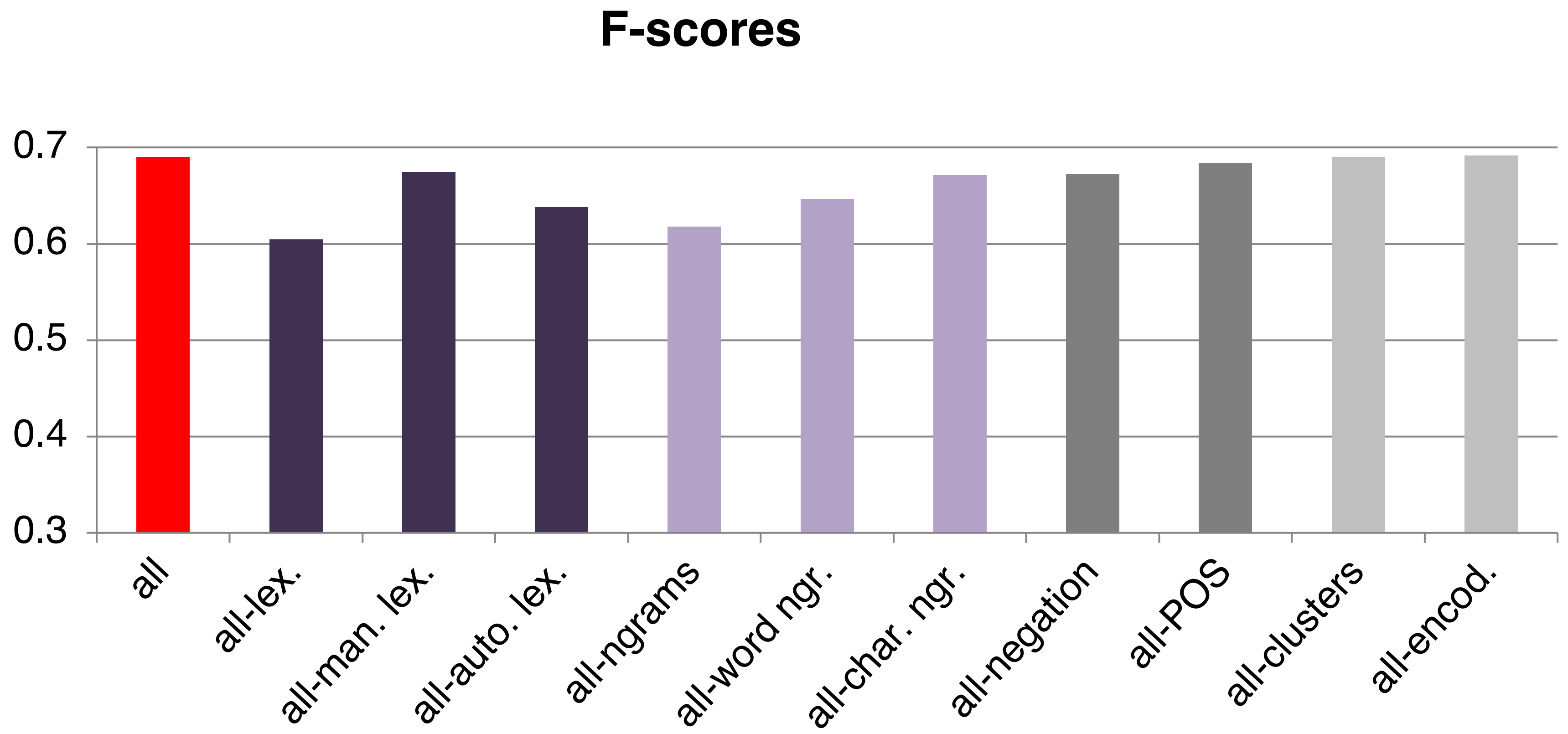


Sentiment Analysis Competition

SemEval-2013: Classify SMS messages, 30 teams



Feature Contributions (on Tweets)

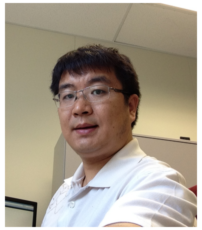


Detecting Stance in Tweets

favor against neither


Parinaz Sobhani


Svetlana Kiritchenko


Xiaodan Zhu


Colin Cherry

Given a tweet text and a target determine whether:

- the tweeter is in favor of the given target
- the tweeter is against the given target
- neither inference is likely

Example 1:

Target: Donald Trump
Tweet: Jeb Bush is the only sane candidate in this republican lineup.

Systems have to deduce that the tweeter is likely against the target.

Example 2:

Target: pro-life movement
Tweet: The pregnant are more than walking incubators, and have rights!

Systems have to deduce that the tweeter is likely against the target.

SemEval-2018 Task 1: Affect in Tweets

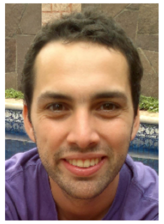
https://competitions.codalab.org/competitions/17751

Tasks: Inferring likely affectual state of the tweeter

- emotion intensity regression (EI-reg)
- emotion intensity ordinal classification (EI-oc)
- sentiment intensity regression (V-reg)
- sentiment analysis, ordinal classification (V-oc)
- multi-label emotion classification task (E-c)

English, Arabic, and Spanish Tweets

75 Team (~200 participants)



Felipe José Bravo Márquez

Mohammad Salameh

Svetlana Kiritchenko

Semeval-2018 Task 1: Affect in tweets. Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. In Proceedings of International Workshop on Semantic Evaluation (SemEval-2018), New Orleans, LA, USA, June 2018.

Participating Systems: ML algorithms

ML algorithm	#Teams				
	EI-reg	EI-oc	V-reg	V-oc	E-c
AdaBoost	1	1	3	1	0
Bi-LSTM	10	8	10	6	6
CNN	10	8	7	6	3
Gradient Boosting	8	3	5	4	1
Linear Regression	11	2	7	2	1
Logistic Regression	9	7	8	6	6
LSTM	13	9	10	5	4
Random Forest	8	7	5	6	6
RNN	0	0	0	0	1
SVM or SVR	15	9	8	6	6
Other	14	16	13	12	7

Participating Systems: features

Features/Resources	El-reg	El-oc	V-reg	V-oc	E-c
	#Teams				
affect-specific word embeddings	10	8	9	9	5
affect/sentiment lexicons	24	16	16	15	12
character ngrams	6	4	3	4	2
dependency/parse features	2	3	3	3	2
distant-supervision corpora	10	8	7	5	4
manually labeled corpora (other)	6	4	4	5	3
AIT-2018 train-dev (other task)	6	5	5	5	3
sentence embeddings	10	8	7	8	6
unlabeled corpora	6	3	5	3	0
word embeddings	32	21	25	21	20
word ngrams	19	14	12	10	9
Other	5	5	5	5	5

SemEval-2018 Task 1: Affect in Tweets

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

Tasks: Inferring likely affectual state of the tweeter

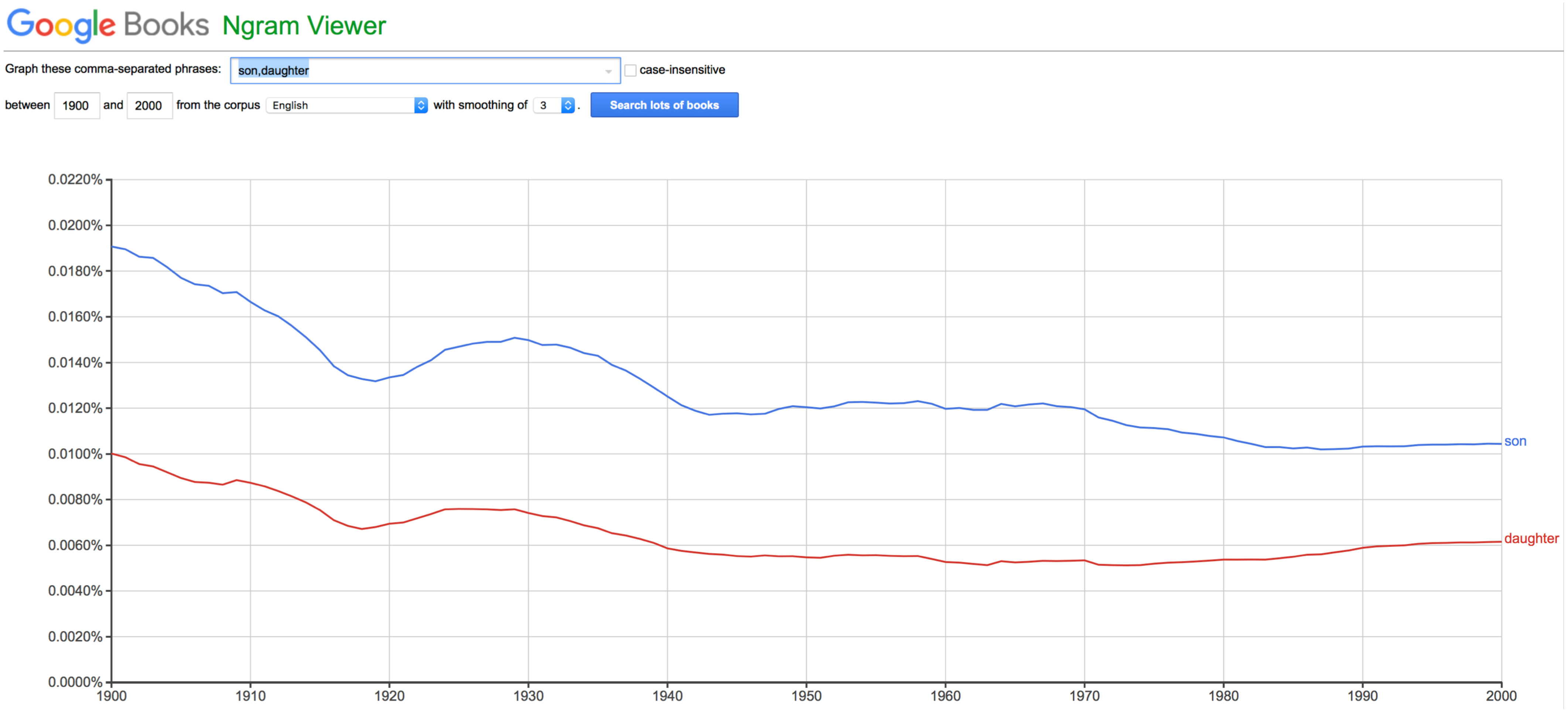
- emotion intensity regression
- emotion intensity ordinal classification
- sentiment intensity regression
- sentiment analysis, ordinal classification
- emotion classification task

English, Arabic, and Spanish Tweets

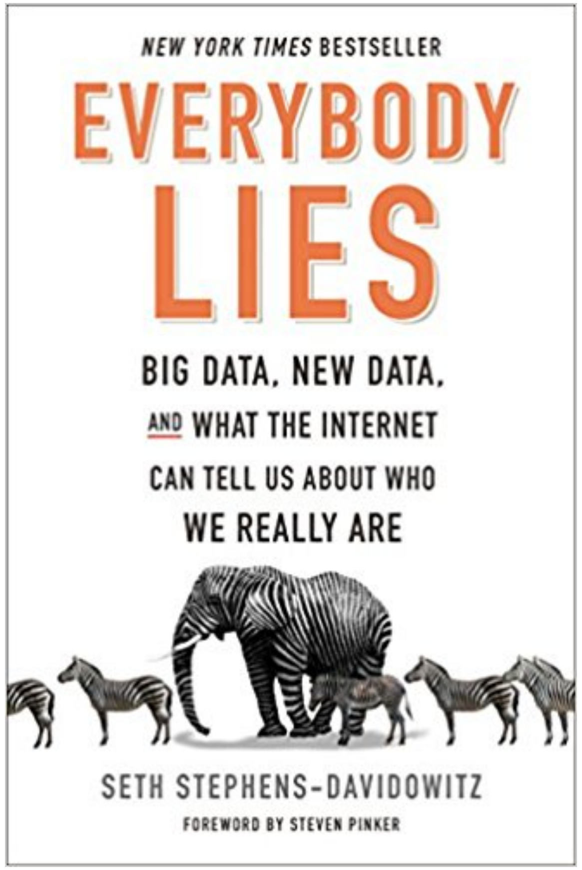
75 Team (~200 participants)

Includes a separate evaluation component for biases towards race and gender.

Occurrences of “son” and “daughter” in the Google Books Ngram corpus



Occurrences of “genius son” and “genius daughter” in the Google Books Ngram corpus



Showed that parents search disproportionately more on Google for:

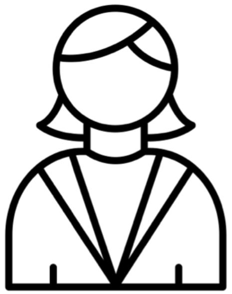
- is my son gifted? than is my daughter gifted?
- is my daughter overweight? than is my son overweight?

Do Machines Make Fair Decisions?

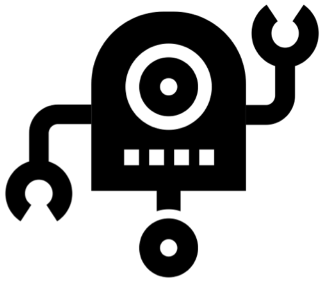
YES:

- they do not take bribes
- they can make decisions without being influenced by the user's gender, race, or sexual orientation

And NO—recent studies have demonstrated that predictive models built on historical data may inadvertently inherit inappropriate human biases



Created by Made
from Noun Project



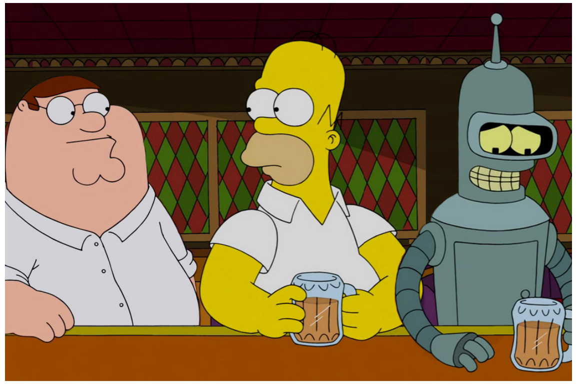
Created by Oksana Latysheva
from Noun Project

Do Machines Make Fair Decisions?

YES:

- they do not take bribes
- they can make decisions without being influenced by the user's gender, race, or sexual orientation

And NO—recent studies have demonstrated that predictive models built on historical data may inadvertently inherit inappropriate human biases



Previous Studies

- focus on one or two systems or resources
  - word embeddings (Bolukbasi et al., 2016; Caliskan et al., 2017; Speer, 2017)
- no benchmark dataset for examining inappropriate biases

Svetlana Kiritchenko

Our Work

- Equity Evaluation Corpus (EEC)—a dataset of 8,640 English sentences carefully chosen to tease out biases towards certain races and genders
- using the EEC, examine the output of 219 sentiment analysis systems that took part in the SemEval-2018 Affect in Tweets shared task

Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of *Sem*, New Orleans, LA, USA, June 2018.

Art and Emotions

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

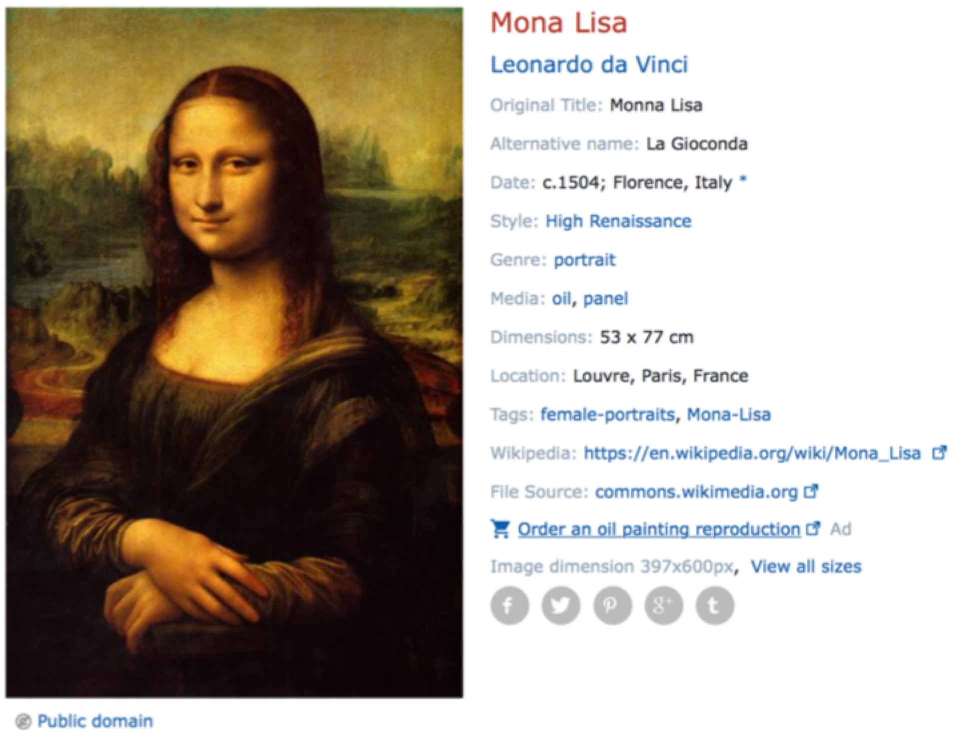


Figure 1: WikiArt.org's page for the *Mona Lisa*.

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.

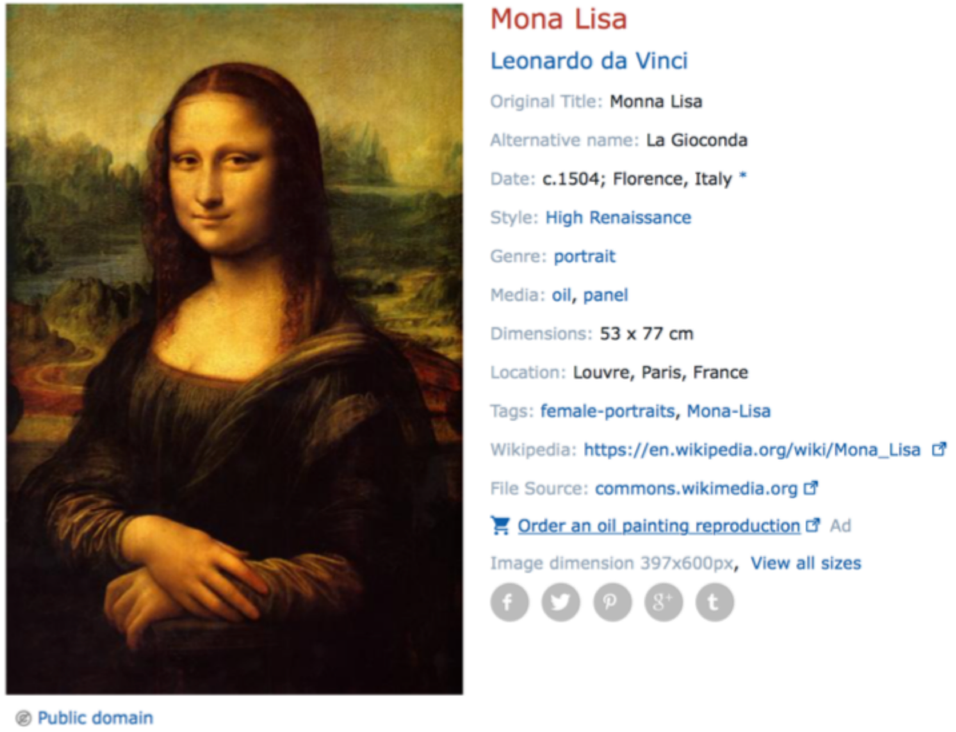
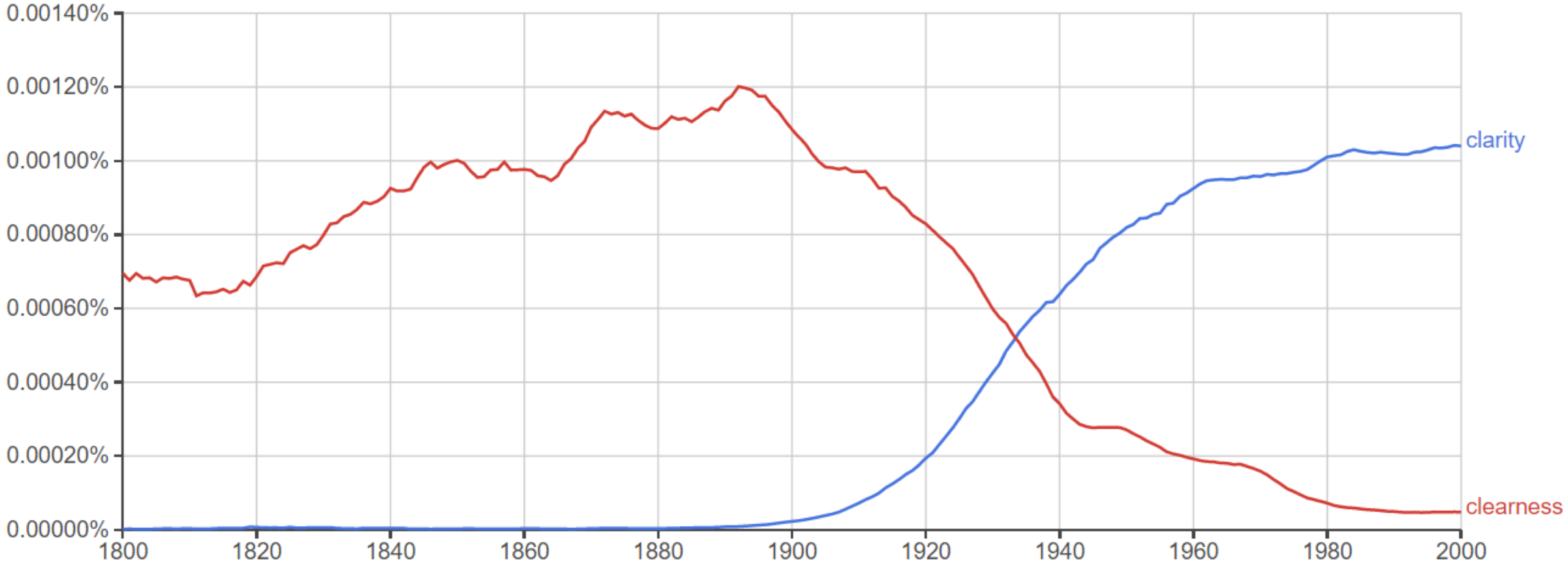


Figure 1: WikiArt.org's page for the *Mona Lisa*. In the WikiArt Emotions Dataset, the *Mona Lisa* is labeled as evoking happiness, love, and trust; its average rating is 2.1 (in the range of −3 to 3).

Clearness versus Clarity in the Google Books Ngrams Corpus

Why did clearness fade away, replaced by clarity?



Peter Turney

The Natural Selection of Words: Finding the Features of Fitness.

Peter Turney and Saif M. Mohammad.

Resources Available at: www.saifmohammad.com

- Sentiment and emotion lexicons and corpora
- Links to shared tasks
- Interactive visualizations
- Tutorials and book chapters on sentiment and emotion analysis

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