University of Stuttgart
Institute for
Natural Language Processing

Emotion Analysis

Intensity Prediction

Dec 24, 2022

Roman Klinger
(includes material
by Laura Oberländer)

Outline

Recap and where we are

- Emotion theories
  Fundamental emotions, Valence-Arousal Dominance, Appraisal, Components, Regulation
- Corpus creation (incl. Assignment 1)
  Annotation, quality assessment, crowdsourcing, existing corpora
- Dictionaries
  Classification, applications, creation, existing lexicons
- Evaluation-based approaches
  OCC model, rules, appraisal
- Classification (incl. Assignment 2)
  Features, deep learning, weak labeling, transfer/multitask learning
- Assignment 3: Context
- Role labeling, Structured Prediction (incl. Assignment 4)
- **Intensity prediction**

Examples for Emotion Role Labeling

- [Djokovic]EXPERIENCER [happy]CUE [to carry on cruising]STIMULUS
- [#Republicans]TARGET are a joke . [Clint Eastwood]STIMULUS is their mascot ! America is in trouble if [these idiots]CUE win ! #RNC
- [Trump]EXPERIENCER [upbeat]CUE [on potential for US-Japan trade deal.]STIMULUS
- [Obama Voter]TARGET [Says Vote for Obama]STIMULUS [YES WE CAN AGAIN !]CUE

Examples from Oberländer et al. (2020): Experiencers, Stimuli, or Targets: Which Semantic Roles Enable Machine Learning to Infer the Emotions? COLING.

Task Definition: Relations, spans, or clauses?

- Relation detection:

cue · stimulus · target · experiencer

A couple | infuriated | officials | by | landing their helicopter in the middle of a nature reserve.

- Sequence labeling:

A couple | infuriated | officials | by | landing their helicopter in the middle of a nature reserve.

target · cue · experiencer · stimulus

John | is | happy | that | Peter meets him, but | Peter | is | annoyed.

- experiencer cue · stimulus · experiencer · cue

→ trade-off between task complexity and accurateness

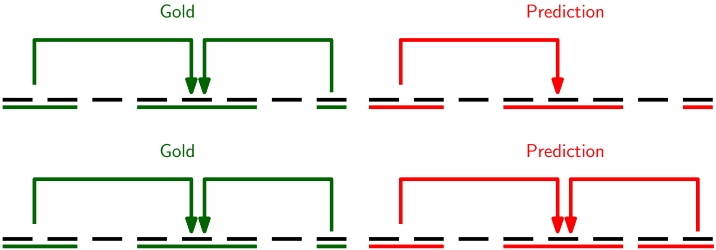
- Clause classification:

A couple infuriated officials | by landing their helicopter in the middle of a nature reserve.

emotion clause · cause/stimulus clause

Evaluation of Graphs

How many TP for spans? How many for relations?



⇒ Error propagation during evaluation.

Take Away

- The task of emotion intensity prediction
- How to annotate for intensities: Best-worst scaling
- Resources that contain emotion intensity annotations
- Computational models to predict intensities

Plan and Remarks

- Next session: Assignment 4 discussion
  - Submission is next Sunday.
  - Teams who did not present yet are asked to present.
  - Any questions regarding the assignment?
- Exam
  - Exam takes place on February 7, 2023
  - 45 minutes exam
  - Room: PWR 7, V7.03 – please come to the room at 5:30 (campus says official start is at 6pm, but we'll start earlier, this time assignment has technical reasons)
  - Any questions regarding the exam?
- Evaluation
  - Evaluation results are online at `https://romanklinger.de/teaching/eaws2223-evaluation.pdf` (English version hopefully available soon, but not yet)

Outline

1. Recap
2. The Emotion Intensity Prediction Task
3. Intensity Annotation of Text
4. Resources
5. Shared Tasks Systems

Recap: Emotion Classification

- Evgeny is happy about the offer.

  Joy

- Sorry, chocolate ice cream is not available today.

  Sadness

- My dog just died.

  Sadness

- This spider might jump into your bed.

  Fear

Emotion Intensity Prediction

Example	Joy	Sadness	Anger	Fear
Evgeny is happy.	0.7	0.0	0.0	0.0
Chocolate not available.	0.0	0.5	0.2	0.1
My dog died.	0.0	0.9	0.2	0.0
Spider jumps into bed.	0.0	0.2	0.2	0.8

(the set of emotions is a parameter)

Emotion Intensity Prediction

Definition

- Given a tweet and an emotion X, the goal is to determine the intensity or degree of emotion X felt by the speaker—a real-valued score between 0 and 1.
- Annotate instances for degree of affect.

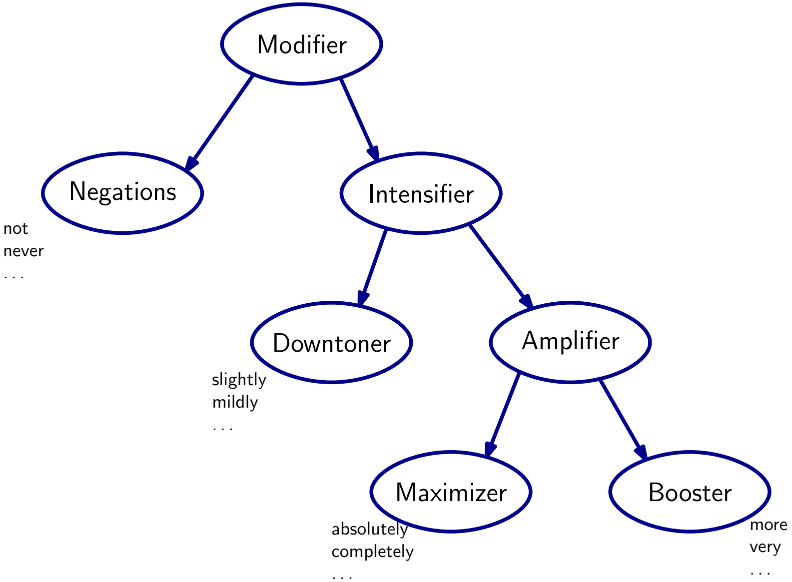
Mohammad, Bravo-Marquez (2017):
WASSA-2017 Shared Task on Emotion Intensity.

How are different intensities of an emotion expressed?

How to determine the intensity of an emotion?

- Paul is happy. — joy
- Paul is excited. — amplified joy
- Paul is very happy. — amplified joy
- Paul is a bit happy. — downtoned joy
- Paul is not happy. — neutral? different emotion?

Modifiers


Modifier
Negations
Intensifier
not
never
...
Downtoner
Amplifier
slightly
mildly
...
Maximizer
Booster
absolutely
completely
...
more
very
...

Studies on Negations, Intensifiers, and Modality

- I like Hilton's hotels. I do not like Hilton's hotels.
  ⇒ negation can flip polarity
- Breakfast is really good. Breakfast is hardly good.
  ⇒ intensifier can flip polarity
- I should eat healthy. I would love you anyways.
  ⇒ modals do not always have the same effect

Carrillo-de-Albornoz, Plaza (2013): An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification

Studies on Negations, Intensifiers, and Modality

Carrillo-de-Albornoz, Plaza (2013): An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification

The authors define three effect types:

1. Might Must Must
   low degree of uncertainty ⇒ no effect
2. Would May May
   desires or high degree of uncertainty ⇒ decrease of emotion
3. Could Could Might Would Should Should
   needs, obligations, disagreements ⇒ reversal of emotion, decrease of strength

Studies on Negations, Intensifiers, and Modality

1. Might Must Must

low degree of uncertainy ⇒ no effect

2. Would May May

desires or high degree of uncertainy ⇒ decrease of emotion

3. Could Could Might Would Should Should

needs, obligations, disagreements ⇒ reversal of emotion, decrease of strength

- Effect Type 1: low change of uncertainty
  - Pictures online must be a different hotel.
- Effect Type 2: desire
  - It would be good to improve food.
- Effect Type 3: disagreement
  - Coffee making facilities in the rooms would have been good.

Carrillo-de-Albornoz, Plaza (2013): An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification

Studies on Negations, Intensifiers, and Modality

TABLE 1. List of negation signals.

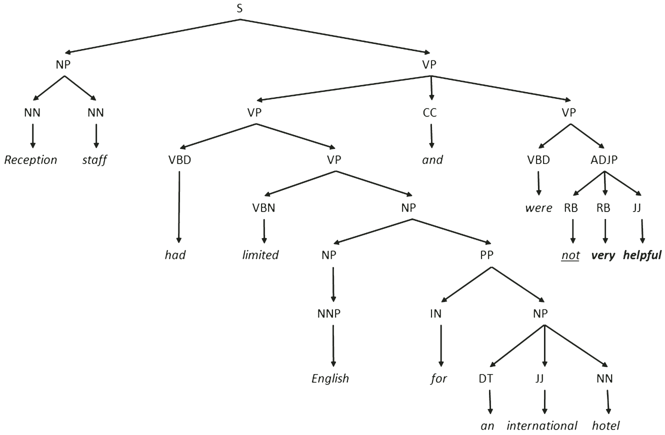
No	None	Non	Nor	Nothing	Neither	Nobody	Doesnt
Never	Nowhere	Not	N't	Don't	Dont	Doesn't	Cant
Won't	Wont	Didn't	Didnt	Haven't	Havent	Can't	Shan't
Cannot	Couldn't	Couldnt	Needn't	Neednt	Wasn't	Wasnt	Isn't
Shant	Weren't	Werent	Daren't	Darent	Hadn't	Hadnt	Hasn't
Isnt	Aren't	Arent	Oughtn't	Oughtnt	Wouldn't	Wouldnt	
Hasnt	Mightn't	Mightnt	Mustn't	Mustnt	Shouldn't	Shouldnt	

TABLE 2. List of intensifiers and their weights.

Term	%	Term	%	Term	%	Term	%	Term	%
Very	75	Small	−50	So	75	Only	−25	Little	−75
Great	60	Really	80	More	35	Fewest	−85	Less	−55
Much	35	Ridiculously	−65	Most	75	Lot	75	Total	75
Some	40	Extraordinarily	75	Barely	−75	Difficult	−65	Big	65
Hardly	−85	Almost	−20	Slightly	−75	Lowest	−85	Huge	75
Relatively	−50	Somewhat	−60	Fairly	65	Few	−75	Fully	85
Pretty	75	Thoroughly	65	Quite	35	Fewer	−65	Complete	75
Perfectly	75	Obviously	75	Certainly	75	Minor	−75	Bigger	35
Completely	85	Definitely	95	Absolutely	95	Low	−75	Absolute	75
Highly	75	Tremendously	85	Especially	70	Lower	−65	Incredible	75
Particularly	45	Significantly	45	Awfully	75	Higher	45	Utter	75
Totally	85	Tremendous	85	Entirely	75	Highest	65	Biggest	75
Strongly	55	Extremely	95	Incredibly	85	Real	20	Super	75
Terribly	75	Immensely	75	Such	55	Extra	20	Rather	75
Exceptionally	75	Exceedingly	85	Vastly	75	Major	35	High	55
Bit	−35								

Carrillo-de-Albornoz, Plaza (2013): An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification

Studies on Negations, Intensifiers, and Modality



Carrillo-de-Albornoz, Plaza (2013): An Emotion-Based Model of Negation, Intensifiers, and Modality for Polarity and Intensity Classification

Data-driven learning of modifier weights

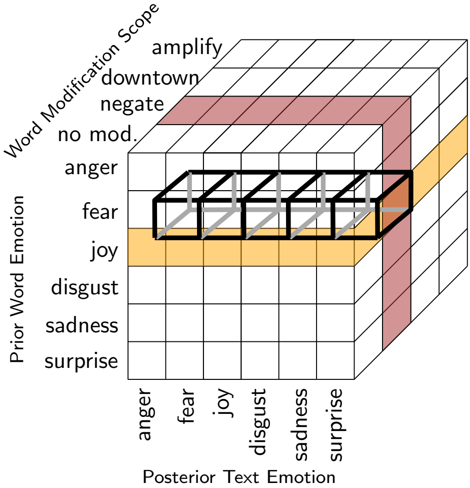
2018 International Conference on Data Science and Advanced Analytics

An Empirical Analysis of the Role of Amplifiers, Downtoners, and Negations in Emotion Classification in Microblogs

Florian Strohm and Roman Klinger
Institut für Maschinelle Sprachverarbeitung
University of Stuttgart
70569 Stuttgart, Germany
Email: {roman.klinger,florian.strohm}@ims.uni-stuttgart.de

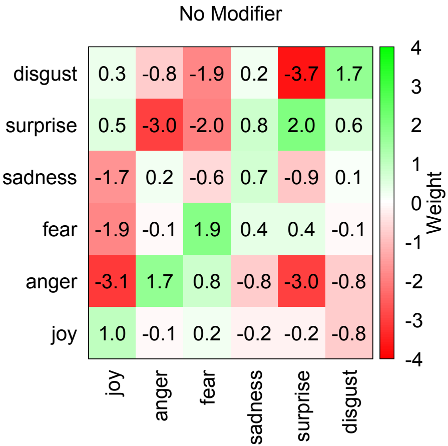
- Study on modifier weights in a dictionary-based classification setting

Dictionary-based Classification: Setting



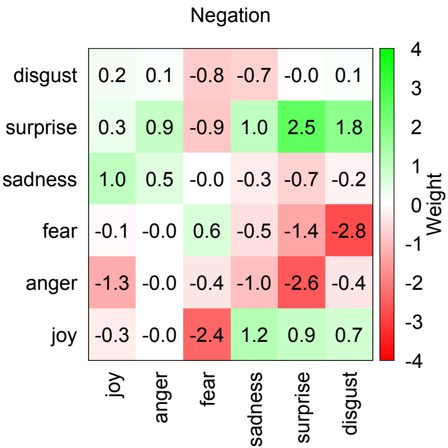
- Training:
  Hill climbing for F_1 on balanced training set
- Inference:
  Maximum a postiori
- Example: "not happy"

Results: No Modifier



- Diagonal has highest values (green)
- Some emotion words do not change other emotions (white)
- Being angry doesn't go well with joy or surprise, surprise not with anger (red)

Results: Negations



- Diagonal has low absolute values (except for surprise)
- Neg. joy → sadness
- Neg. surprise → surprise
- Neg. sadness → joy
- Mostly lower positive weights, some strong negative weights
- Some negations mean "nothing": anger, disgust

Negation Examples

Joy ⇒ Sadness

“Not sure how this happened but in two days I’ve somehow gained 5 lbs...so not happy about this. #ugly #fatty #depressed #sad”

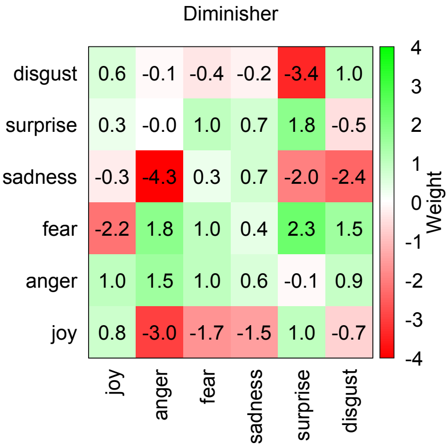
Sadness ⇒ Joy

“Yes! I’m about to eat this piece of cheesecake and I don’t feel guilty about it. #indulgingalittle #cheesecake #happy”

Fear ⇒ Fear

“Don’t worry, let God take control. #worry”
“”No fear is stronger than you are.” - Mark David Gerson #fear #quote #spirituality”

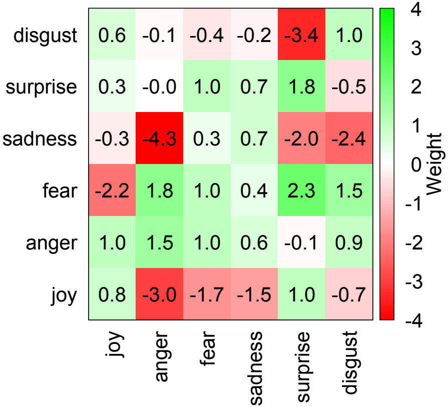
Results: Downtoner



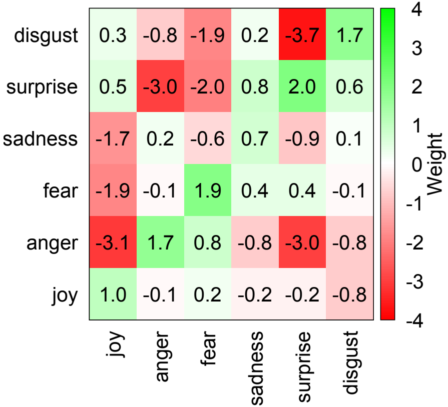
- Nothing surprising, similar to no modifier, mostly lower weights
- Some exceptions, e.g. “a bit sad” → no anger at all

Results: Downtoner

Diminisher



No Modifier



Downtoner Example

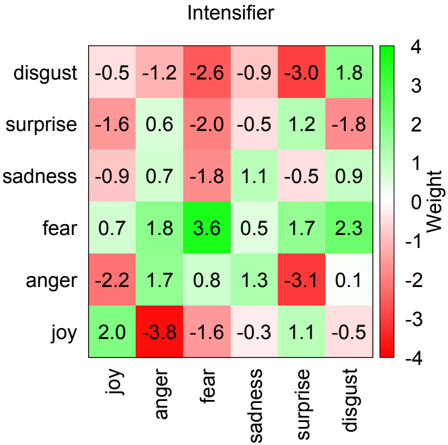
Sadness $\Rightarrow$ Sadness, Joy

"pray more and worry less #pray #faith #love #peace #happiness..."

Joy...

"Just a bit happy to be back in Ibiza..."

Results: Amplifier



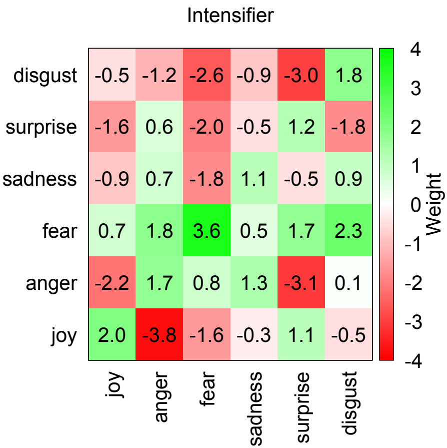
- "Stronger" weights
- Especially clearer separation from (some) other emotions

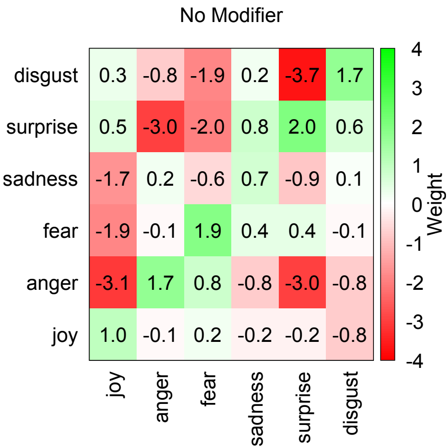
Amplifier Example

Joy $\Rightarrow$ 2$\cdot$ Joy

"Wishing you a very happy day! #happiness #positivity'

Results: Amplifier





Discussion: Relation to Other Variables

- Probability of emotion classification prediction?
  Confidence of annotators?
  - Might be correlated, intense emotions are easier to recognize
  - Troiano et al. 2021: Emotion Ratings: How intensity, annotation confidence and agreements are entangled.
- Valence–Arousal
  - Valence: Degree of positivity, not the same (but correlated?).
  - Arousal: Degree of activiation, not the same (but correlated?).
  - $\Rightarrow$ No clear answers.

Outline

1. Recap
2. The Emotion Intensity Prediction Task
3. Intensity Annotation of Text
4. Resources
5. Shared Tasks Systems

Intensity Annotation

Which score [0;I] should be assigned to these instances?

- He is super happy!!!
  - Perhaps something close to one?
- He is more happy than he has even been before.
  - Perhaps this one is even higher?
- Ideas:
  - Do not assign scores in isolation.
  - Consider multiple instances at the same time to increase the availability of context.

- Best worst scaling:
  Finn, A. and Louviere, J. J. 1992. Determining the appropriate response to evidence of public concern: The case of food safety. Journal of Public Policy and Marketing., 11: 19–25
- First use for emotion analysis: Mohammad, Bravo-Marquez (2017): WASSA-2017 Shared Task on Emotion Intensity.

Intensity Annotation, Desiderata

- Consistency
  - Annotations by different people should be comparable
  - Annotations by the same person should be comparable on same/comparable instances
- Granularity
  - We would like to have an interpretable scale.
  - Meanings should be ‘evenly distributed’ (no bias towards one side)

https://xkcd.com/1098/

Best-Worst Scaling

- Organize n items to be rated in m 4-tuples
- Annotators are presented with one 4-tuple at a time and answer two questions:
  1. Which item is associated with the highest intensity of anger?
  2. Which item is associated with the lowest intensity of anger?

he's angry
the face shows rage
john frowns
I am happy

↓

the face shows rage
he's angry ; john frowns
I am happy

How much do we get from such Quadruple?

Annotation result:

A B ; C D

- $A > B$
- $A > C$
- $A > D$
- $B > D$
- $C > D$
- Don't know B–C.

- Could we just do pairwise annotations?
  - If we show 5 pairs, annotators read A and D three times, C and B twice. More work, less context.
- Could we increase efficiency by showing k-tuples with larger k?
  - The amount of elements in the middle which don't receive judgements increases.
  - For quintuple (A best, E worst), we get 7 pairs: $A > B$, $A > C$, $A > D$, $A > E$, $B > E$, $C > E$, $D > E$
  - Don't know B-C and B-D and C-D

How to get scores from BWS?

- Each tuple with the BWS questions shown to annotators
- Obtain real valued scores for all the terms using the formula:

$$\text{score}(\text{item}) = \frac{\#\text{best}(\text{item}) - \#\text{worst}(\text{item})}{\#\text{annotations}(\text{item})}$$

- $\text{score}(\cdot) \in [-1, 1]$
- Normalize and scale as needed

How to put the quadruples together?

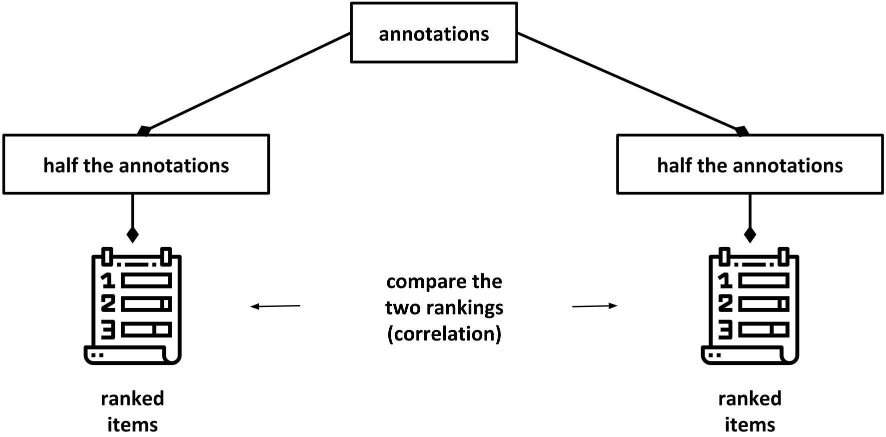
Rules to balance the tuples:

- No two samples have the same four items (in any order)
- No two items within a sample are identical
- Each item occurs in j different samples
- Each pair of items appears in the same number of samples

How many do we need?

- Empirical observation:
  Reliable results can be obtained with $m = 2n$ quadruples.
- With $j = 8$ and three annotators Mohammad (2017) gets 24 rating per item

Quality Assessment for BWS: Split-half Reliability



Hands on BWS

Use BWS to annotate joy in following 5 Tweets, use 10 4-tuples.

```
I feel so blessed to work with the family
Today I reached 1000 subscribers on YT!!
Quinn's short hair makes me sad.
Be happy not because everything is good
try asking for a cheeseburger with only onion
```

- What are the scores for each instance?
- Optional: What is the reliability calculated with Pearson's r?

You can use the tuples at `https://www.emotionanalysis.de/lecture/s08bws.pdf`, if you wish, created via

```
for (( i=0 ; i<10 ; i++ )) ; do shuf 5instances | head -n 4 > $i ; done
```

Lecture on Emotion Analysis in Text

In-Class Exercise on Best-Worse-Scaling for Emotion Intensity Annotation

Roman Klinger

Mark those instances with the highest intensity of joy and with the lowest intensity of joy

```
Quinn's short hair makes me sad.
Be happy not because everything is good
I feel so blessed to work with the family
Today I reached 1000 subscribers on YT!!

Quinn's short hair makes me sad.
try asking for a cheeseburger with only onion
Today I reached 1000 subscribers on YT!!
Be happy not because everything is good

Today I reached 1000 subscribers on YT!!
Quinn's short hair makes me sad.
I feel so blessed to work with the family
Be happy not because everything is good

I feel so blessed to work with the family
try asking for a cheeseburger with only onion
Be happy not because everything is good
Quinn's short hair makes me sad.

Be happy not because everything is good
I feel so blessed to work with the family
Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!

Be happy not because everything is good
Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family

Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!
Be happy not because everything is good
I feel so blessed to work with the family

Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family
Be happy not because everything is good
Quinn's short hair makes me sad.

Today I reached 1000 subscribers on YT!!
Be happy not because everything is good
I feel so blessed to work with the family
try asking for a cheeseburger with only onion

try asking for a cheeseburger with only onion
Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family
Be happy not because everything is good
```

Lecture on Emotion Analysis in Text

In-Class Exercise on Best-Worse-Scaling for Emotion Intensity Annotation

Roman Klinger

**Mark those instances with the highest intensity of joy
and with the lowest intensity of joy**

Quinn's short hair makes me sad.
Be happy not because everything is good
I feel so blessed to work with the family
Today I reached 1000 subscribers on YT!!

Quinn's short hair makes me sad.
try asking for a cheeseburger with only onion
Today I reached 1000 subscribers on YT!!
Be happy not because everything is good

Today I reached 1000 subscribers on YT!!
Quinn's short hair makes me sad.
I feel so blessed to work with the family
Be happy not because everything is good

I feel so blessed to work with the family
try asking for a cheeseburger with only onion
Be happy not because everything is good
Quinn's short hair makes me sad.

Be happy not because everything is good
I feel so blessed to work with the family
Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!

Be happy not because everything is good
Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family

Quinn's short hair makes me sad.
Today I reached 1000 subscribers on YT!!
Be happy not because everything is good
I feel so blessed to work with the family

Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family
Be happy not because everything is good
Quinn's short hair makes me sad.

Today I reached 1000 subscribers on YT!!
Be happy not because everything is good
I feel so blessed to work with the family
try asking for a cheeseburger with only onion

try asking for a cheeseburger with only onion
Today I reached 1000 subscribers on YT!!
I feel so blessed to work with the family
Be happy not because everything is good

Hands on BWS – Solution

- I feel so blessed to work with the family

  0.938 in EmoInt Data
- Today I reached 1000 subscribers on YT!!

  0.845 in EmoInt Data
- Quinn’s short hair makes me sad.

  0.083 in EmoInt Data
- Be happy not because everything is good

  0.627 in EmoInt Data
- try asking for a cheeseburger with only onion

  0.567 in EmoInt Data

Hands on BWS – Solution (based on annotation example)

- I feel so blessed to work with the family
  - $9 \times$ best; $0 \times$ worst
  - $(9-0)/9 = 1$ 0.938 in EmoInt Data
- Today I reached 1000 subscribers on YT!!
  - $1 \times$ best; $0 \times$ worst
  - $(1-0)/1 = 1$ 0.845 in EmoInt Data
- Quinn's short hair makes me sad.
  - $0 \times$ best; $8 \times$ worst
  - $(0-8)/8 = -1$ 0.083 in EmoInt Data
- Be happy not because everything is good
  - $0 \times$ best; $2 \times$ worst
  - $(0-2)/2 = -1$ 0.627 in EmoInt Data
- try asking for a cheeseburger with only onion
  - $0 \times$ best; $0 \times$ worst
  - ? 0.567 in EmoInt Data

Outline

1. Recap
2. The Emotion Intensity Prediction Task
3. Intensity Annotation of Text
4. Resources
5. Shared Tasks Systems

Dictionaries with Intensity Scores

- G. P. Strauss, D. N. Allen (2008): Emotional intensity and categorisation ratings for emotional and nonemotional words. Cognition and Emotion 22 (1):114-133.
  - Manual annotation of 463 words with 200 students
  - (can't say more, paper is behind paywall, 43 USD)
- Saif M. Mohammad (2018): Word Affect Intensities
  - Create lexicon with BWS with intensity scores for 6000 words
  - `http://www.saifmohammad.com/WebPages/AffectIntensity.htm`
- Koeper, Kim, Klinger (2018): IMS at EmoInt-2017: Emotion Intensity Prediction with Affective Norms, Automatically Extended Resources and Deep Learning
  - Use neural network which takes word embeddings as input and outputs intensity score, trained on manually labeled data.
  - `http://www.ims.uni-stuttgart.de/data/ims_emoint`

WASSA-2017/SemEval-2018 Twitter Corpora

- Emotion Intensity Shared Task at WASSA-2017:
  First initiative for emotion intensity prediction in shared task
- 7102 Tweets with annotations freely available
  (separated in train, dev, test)
- Share task description paper: Mohammad/Bravo-Marques (2017): WASSA-2017 Shared Task on Emotion Intensity.
  https://www.aclweb.org/anthology/W17-5205.pdf
- Details on data set: Mohammad/Bravo-Marques (2017): Emotion Intensities in Tweets https://www.aclweb.org/anthology/S17-1007.pdf
- Data: https://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html
- Extended for SemEval-2018 Shared Task with more data and more languages
  - English + Arabic and Spanish
  - Shared task website: https://competitions.codalab.org/competitions/17751
  - Shared task paper: https://www.aclweb.org/anthology/S18-1001/

Details on the dataset creation

Query Twitter

- Synonyms from Roget's Thesaurus
- anger: angry, mad, frustrated, annoyed, peeved, irritated, miffed, fury, antagonism...
- sadness: sad, devastated, sullen, down, crying, dejected, heartbroken, grief, weeping...
- At most 50 tweets per query term
- At most 1 tweet for every tweeter-query-term combination
- Include variants without emotion hashtags to the data (to study effect of hashtags)

Crowdsourcing Annotation

Details:

- 4 tweets at a time (4-tuple)
- 1 tweet appeared in 8 different 4-tuples
- 3 independent annotators

Quality assurance:

- 5% annotated by authors (gold questions)
- Accuracy of annotations on gold questions below 70%, annotator removed

Example of annotation for the degree of Anger

Which one is ‘most’ and which one is ‘least’ angry?

1. `Someone stole my photo on tmblr #grrr`
2. `I didn't find out about this till today due to my bff telling me. I am so disgusted and offended by this.`
3. `why are people so angry toward veggie burgers?`
4. `That grudge you're holding keeps making an appearance because #God wants you to deal with it.`

Examples of annotated tweets

- IreneEstry can't wait to see you Hun #cuddles #gossip
  joy, 0.77
- *Sigh* #depression #saddness #afterellen #shitsucks
  sadness, 0.91
- ima kitchen sink
  sadness, 0.33
- like srsly somebody help me deal
  with this social anxiety
  fear, 0.97
- When you just want all the attention #cantsleep
  fear, 0.50
- DJ_JeanFranko growl!!!
  anger, 0.50

Split-Half Reliability in Created Dataset

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

Other Corpora (I)

- Bostan, Kim, Klinger: GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception (2020)
  - Corpus focused on development of spans of feelers, targets, stimuli, and associated emotions
  - Discrete emotions also labeled with intensities: Low, Medium, High, for multiple annotators in crowdsourcing
- Strapparava, Mihalcea: SemEval-2007 Task 14: Affective Text (2007)
  - "The interval for the emotion annotations was set to [0, 100], where 0 means the emotion is missing from the given headline, and 100 represents maximum emotional load"
  - Main annotation task modelled as scoring, not as categorization

Other Corpora (II)

- Aman, Szpakowicz:
  Identifying Expressions of Emotion in Text (2007)
  - "The second kind of annotations involved assigning emotion intensity (high, medium, or low) to all emotion sentences in the corpus, irrespective the emotion category assigned to them. No intensity label was assigned to the no emotion sentences."
  - Intensity in addition to categorization
- Alm, Roth, Sproat: Emotions from text: machine learning for text-based emotion prediction (2005)
  - Main annotation task is categorization
  - Contains intensity annotations from 1–3 (no details on intensity annotation process given).

⇒ WASSA2017/SemEval2018 corpora are the first resources annotated with a focus on intensity

Outline

Baseline System and its features

AffectiveTweets System with features:

- Features:
  - Sparse Features (Word N-grams and Character N-grams)
  - Affect Lexicon
  - Word Embeddings
- System:
  - Implemented in Weka
  - Pretty easy to use, even without programming skills

	Pearson correlation r				
	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
Over the subset of test set where intensity ≥ 0.5					
WN + WE + L	0.51	0.51	0.40	0.49	0.47

WASSA-2017 Official Results of the Shared Task

- 22 teams participated
  (48 on English data in SemEval 2018,
  76 across all languages)
- Only 7 teams above baseline (0.66)

Team Name	r avg. (rank)	r fear (rank)	r joy (rank)	r sadness (rank)	r anger (rank)
1. Prayas	0.747 (1)	0.732 (1)	0.762 (1)	0.732 (1)	0.765 (2)
2. IMS	0.722 (2)	0.705 (2)	0.726 (2)	0.690 (4)	0.767 (1)
3. SeerNet	0.708 (3)	0.676 (4)	0.698 (6)	0.715 (2)	0.745 (3)
4. UWaterloo	0.685 (4)	0.643 (8)	0.699 (5)	0.693 (3)	0.703 (7)
5. IITP	0.682 (5)	0.649 (7)	0.713 (4)	0.657 (7)	0.709 (5)
6. YZU NLP	0.677 (6)	0.666 (5)	0.677 (8)	0.658 (6)	0.709 (5)
7. YNU-HPCC	0.671 (7)	0.661 (6)	0.697 (7)	0.599 (9)	0.729 (4)

Pearson correlations (r) and ranks (in brackets) obtained by the first seven systems on the full test sets.

WASSA-2017 Shared Task on Emotion Intensity

Commonly Used Setups:

- Features:
  Word embeddings, Sentence embeddings, Affective Lexicons
- Regression Methods: Neural Models, SVM or SVR
- Toolkits, libraries: Keras & Sci-kit learn

Prayas – Winning System at WASSA 2017

- Ensemble of several approaches
- Approach 1: Feed forward neural network
  - Word2vec word embeddings, many lexicons
- Approach 2: Multi-task learning NN
  - Share network properties across different emotions
- Approach 3: Sequence learning with CNNs/LSTMs

Pranav Goel, Devang Kulshreshtha, Prayas Jain and K.K. Shukla (2017): Prayas at EmoInt 2017: An Ensemble of Deep Neural Architectures for Emotion Intensity Prediction in Tweets. https://www.aclweb.org/anthology/W17-5207.pdf

IMS – Second Position at WASSA 2017

- Main model architecture: random forest regressor
- Features:
  - Manually created dictionaries
  - Automatically extended dictionaries (with additional Twitter data)
  - CNN-LSTM regressor

Maximilian Köper, Evgeny Kim, Roman Klinger (2017): IMS at EmoInt-2017: Emotion Intensity Prediction with Affective Norms, Automatically Extended Resources and Deep Learning. https://www.aclweb.org/anthology/W17-5206/

Take Away

- The task of emotion intensity prediction
- How to annotate for intensities: Best-worst scaling
- Resources that contain emotion intensity annotations
- Computational models to predict intensities

Plan and Remarks

- Next session: Assignment 4 discussion
    - Submission is next Sunday.
    - Teams who did not present yet are asked to present.
    - Any questions regarding the assignment?
- Exam
    - Exam takes place on February 7, 2023
    - 45 minutes exam
    - Room: PWR 7, V7.03 – please come to the room at 5:30 (campus says official start is at 6pm, but we'll start earlier, this time assignment has technical reasons)
    - Any questions regarding the exam?
- Evaluation
    - Evaluation results are online at `https://romanklinger.de/teaching/eaws2223-evaluation.pdf` (English version hopefully available soon, but not yet)