Construction and Analysis of an Emotion Proposition Store

Sebastian Ruder

Institute of Computational Linguistics, University of Heidelberg

21.07.15

Agenda

Introduction

- Sentiment analysis (positive/negative/neutral) $\rightarrow$ important for many down-stream applications
- Emotion detection: subtask of sentiment analysis
- Project
  1. Design and evaluation of frequent and clearly emotion-indicating patterns
  2. Acquisition of $>$ 1,700,000 propositions (emotion, emotion holder, cause) from Gigaword news corpus [1]
  3. Storage in an emotion proposition store
  4. Distributional analysis and evaluation for further insights (ambiguous concepts, etc.)

Motivation

- *(I think) people are happy because Chavez has fallen.* [3]; E: FALL(X) is a bad-for event; HAPPY_ABOUT(Y, E) has a positive emotion → requires insights about emotion expressions over propositions, positively/negatively perceived events
- Resource for emotion detection
- Pattern-based approach can be extended to other domains

Classification

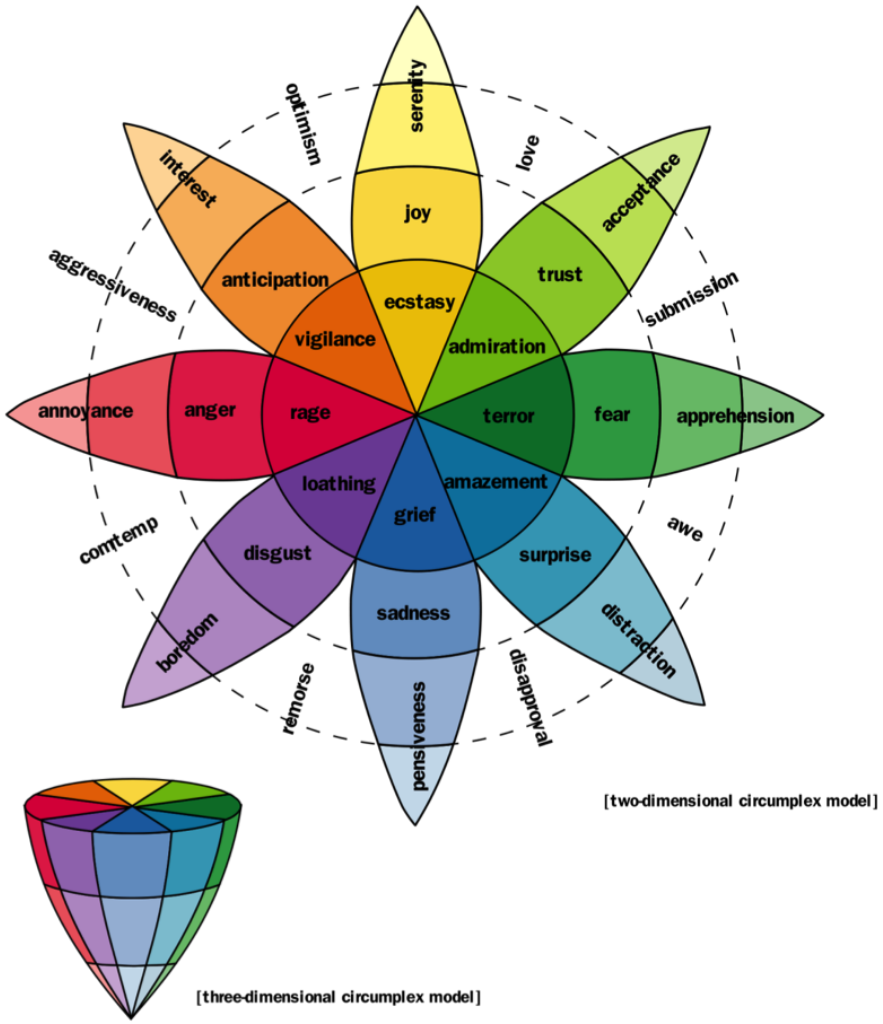

Figure : Plutchik's emotion wheel [2]

Emotions

- Diverse linguistic triggers (*disaster*, *yucky*, *betray*)
- Compositionality (*the blind* + *sees*)
- Semantic roles: Emotion holder & cause
- Ambiguity

$\rightarrow$ frequent & clearly emotion indicating predicators, e.g. *X be angry about NP / that S*, *X fears that S* for acquisition

Corpus

- Annotated Gigaword v.5 [4]

# tokens	# documents
4,032,686,000	9,876,086

Table : Number of tokens and documents for English Gigaword v.5

Patterns

fear scare/Verb NP true
joy be/Verb RB happy/JJ that/IN S false (1)

1. `fear (?! not)(?! never)scare/VB[DGPZ]/[0-9]+ NP`
2. `fear (?<= )be/VB[PDGZ]/([0-9]+)(?! not) (?! never)( [a-z]+/RB/[0-9]+)? scare/VBN/[0-9]+ that/IN/[0-9]+ S`
3. `fear (?<= )be/VB[PDGZ]/([0-9]+)(?! not) (?! never)( [a-z]+/RB/[0-9]+)? scare/VBN/[0-9]+ by/IN/[0-9]+ NP`
4. `joy (?! not)(?! never)be/VB[DGPZ]/[0-9]+ (?! not)(?! never)( [a-z]+/RB/[0-9]+)? happy/JJ/[0-9]+ that/IN/[0-9]+ S`

Pattern sources

Source	**# of patterns**
Oxford English Dictionary	24
Merriam-Webster's Dictionary	51
Roget's Thesaurus	101
Harvard General Inquirer [5]	0
NRC Emolex [6]	0
Emotion verb classes [7]	24
Adjectives [8]	126
VerbNet (*admire* + *amuse*) [9]	178
FrameNet (EMOTIONS frame) [10]	173
WordNet-Affect [11]	123

Table : Overview of the productivity of sources for pattern design

Annotator agreement on most frequent patterns

Annotated expressions	Number
Unanimous emotions	106
Unanimous emotions (including 2nd choice)	119
Majority emotions	163
Unanimous emotions + degree	39
Majority emotions + degree	131
Total	180
Fleiss' κ	0.65

Table : Number of annotated expressions for different forms of agreement

Final pattern distribution across emotions

Emotion	**# of majority patterns**
Joy	31
Trust	8
Fear	22
Surprise	16
Sadness	18
Disgust	14
Anger	29
Anticipation	25
Total	163

Table : Number of patterns that have been labeled by the majority with the same emotion

Extraction

- Constituencies << dependencies
- Collapsed Stanford dependencies (`nsubj`, `dobj`, `conj_and`, `ccomp`, `pcomp`, etc.)
- Modifiers (`nn`, `amod`, `num`)
- Prepositional objects
- Coreference resolution
- Tagging of named entities

Extraction example

- *The countries had engaged in a multimillion-dollar battle to host the tournament, with Japan relying on its economic clout and South Korea relying on its superior soccer pedigree.*
- `NYT_ENG_19960601.0010/1 trust rely on Japan/LOCATION economic clout [its/PRP$, economic/JJ, clout/NN]`
- *The Japanese, who expected to win the right to host the tournament, were dismayed.*
- `NYT_ENG_19960601.0010/9 anticipation expect Japanese win right [to/TO, win/VB, the/DT, right/NN, to/TO, host/VB, the/DT, tournament/NN]`

Emotion distribution

Emotion	Frequency	% of total extractions	# of patterns with 10+ occurrences
anticipation	966,571	54.47	22
fear	249,103	14.04	20
joy	231,967	13.07	30
trust	89,217	5.03	6
anger	64,586	3.64	28
surprise	60,221	3.39	20
disgust	59,486	3.35	15
sadness	53,269	3.00	13
Total	1,774,420	100.00	154

Table : Frequencies of emotions in extractions

NP vs. S as cause

Emotion	**# extractions with NP cause**	**# extractions with S cause**
anticipation	407,738	558,833
joy	190,484	41,483
fear	82,116	166,987
trust	72,483	16,734
surprise	59,657	564
disgust	58,942	544
anger	57,379	7,207
sadness	26,064	27,205
Total	956,392	819,557

Table : Patterns with NP vs. S cause

PMI vs. chi-square

χ^2 **bigram**	χ^2 **value**	**PMI bigram**	**PMI value**
loss of:life	9426.09	death of:NUM	3.40
loss of:innocent_life	2317.45	quake victim	3.37
tragic loss	1430.38	death of:wife	3.37
loss of:civilian_life	1416.05	loss of:man	3.37
civilian casualty	1303.09	tragic loss	3.36
death of:NUM	1208.49	death of:father	3.36
death of:NUM_people	1038.87	death of:NUM_people	3.34
NUM victim	870.43	death of:relative	3.33
unfortunate incident	677.81	loss of:friend	3.33
choice of:word	516.71	innocent victim	3.33

Table : Comparison of χ^2 and PMI values for the top 10 sadness NP cause bigrams

Most ambiguous expressions

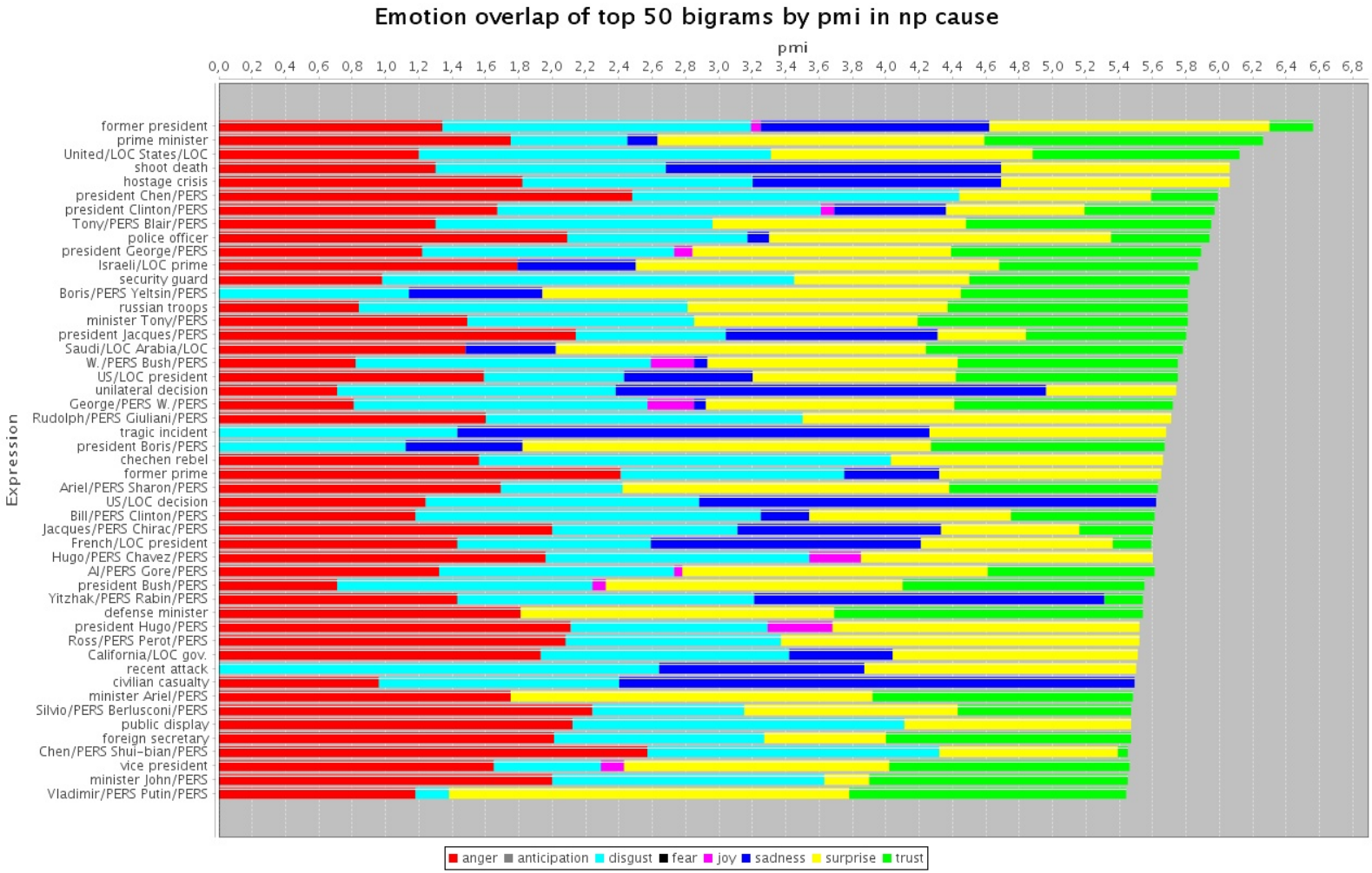

Figure : Overlap across emotions for the top 50 PMI NP cause bigrams

Agreement with NRC EmoLex

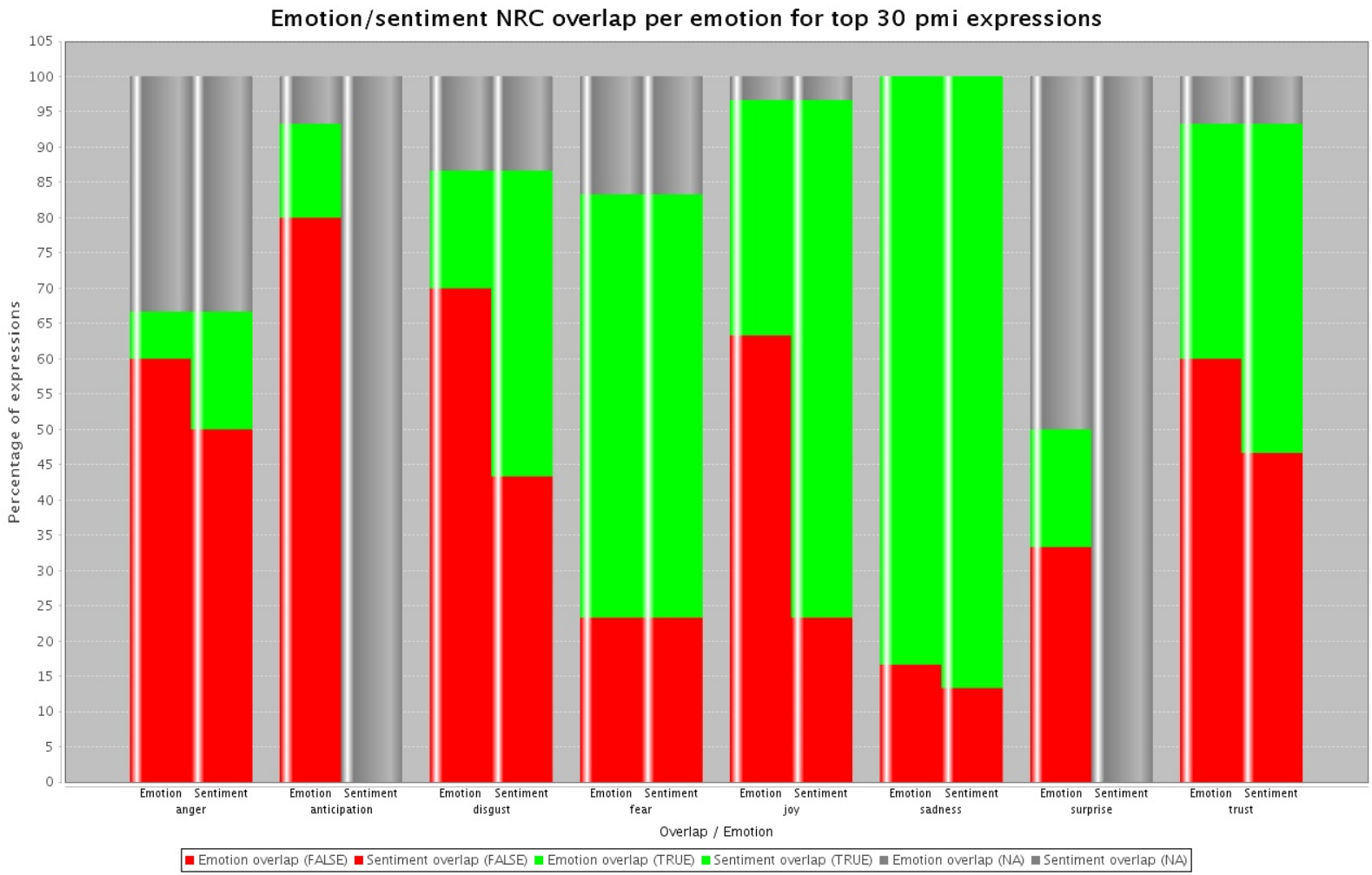

Figure : Emotion and sentiment overlap of NP cause bigrams with the NRC Emotion Lexicon

Performance against annotated gold standard

	NP cause			**S cause pred + dobj**		
Emotion/sentiment	*P*	*R*	*F1*	*P*	*R*	*F1*
anger	0.00	-	-	0.00	-	-
anticipation	0.00	-	-	0.00	-	-
disgust	0.07	1.00	0.13	0.06	1.00	0.11
fear	0.33	1.00	0.50	0.40	0.89	0.55
joy	0.90	0.95	0.92	0.83	0.60	0.70
sadness	1.00	0.67	0.80	0.36	0.45	0.40
surprise	0.00	-	-	0.00	-	-
trust	0.00	-	-	0.05	1.00	0.10
total – emotion	0.36	1.00	0.53	0.23	1.00	0.37
positive	0.56	0.96	0.71	0.40	0.57	0.47
negative	0.60	0.84	0.70	0.37	0.87	0.51
neutral	0.00	-	-	0.00	-	-
total – sentiment	0.47	1.00	0.64	0.29	1.00	0.45

Topic modeling

- LDA with different topic configurations (10, 20, 30, 50 topics) on 8 emotion-associated pseudo-documents
- Evaluation of how well topics are associated with certain emotions

Lessons learned

- Amount of available data directly impacts results (only few S causes for surprise, disgust, anger)
- Different configuration produce a plethora of different scenarios that need to evaluated
- Finding adequate resources for evaluation can be difficult; manual annotation is expensive

Bibliography I


Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda.
English gigaword fifth edition.
2011.


Robert Plutchik.
A general psychoevolutionary theory of emotion.
Theories of emotion, 1, 1980.


Lingjia Deng, Yoonjung Choi, and Janyce Wiebe.
Benefactive/malefactive event and writer attitude annotation.
In *ACL (2)*, pages 120–125, 2013.

Bibliography II

Courtney Napoles, Matthew Gormley, and Benjamin Van Durme.
Annotated gigaword.
In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, pages 95–100. Association for Computational Linguistics, 2012.

Philip J Stone and Earl B Hunt.
A computer approach to content analysis: studies using the [g]eneral [i]nquirer system.
In *Proceedings of the May 21-23, 1963, spring joint computer conference*, pages 241–256. ACM, 1963.

Bibliography III


Saif M Mohammad and Tony Wenda Yang.
Tracking sentiment in mail: how genders differ on emotional axes.
In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011*, pages 70–79, 2011.


Christiane Fellbaum and Yvette Yannick Mathieu.
A corpus-based construction of emotion verb scales.
Linguistic Approaches to Emotions in Context, 241:99, 2014.


Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archna Bhatia, Manaal Faruqui, and Chris Dyer.
Augmenting English adjective senses with supersenses.
In *Proc. of LREC*, 2014.

Bibliography IV


Karin Kipper-Schuler.
A Broad-Coverage Comprehensive Verb Lexicon.
PhD thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA, 2005.


Collin F Baker, Charles J Fillmore, and John B Lowe.
The berkeley framenet project.
In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics, 1998.


Carlo Strapparava and Alessandro Valitutti.
WordNet Affect: an affective extension of WordNet.
In *LREC*, volume 4, pages 1083–1086, 2004.