FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

MODELL ROMANTiK

DFG

Jena University Language & Information Engineering Lab

Modeling Word Emotion in Historical Language:

Quantity Beats Supposed Stability in Seed Word Selection

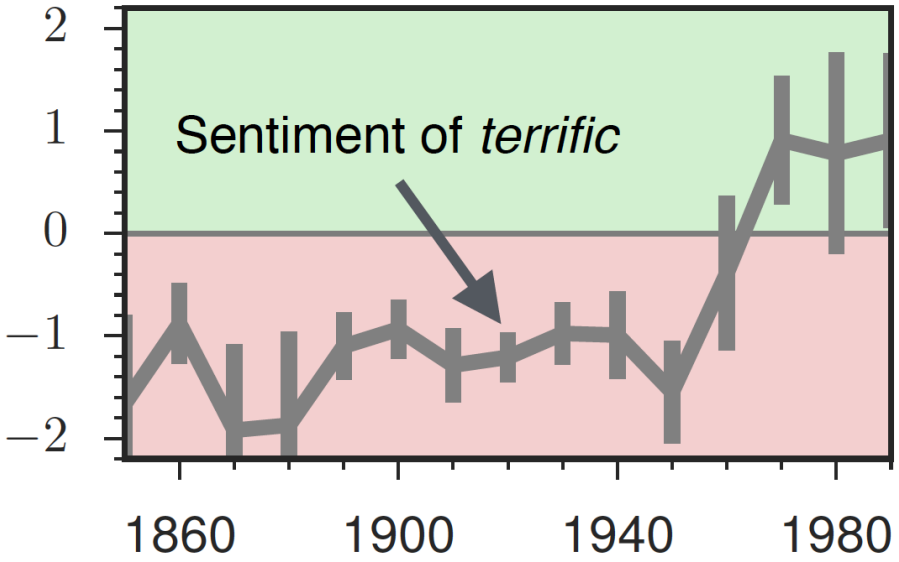
Johannes Hellrich* Sven Buechel* Udo Hahn

Jena University Language and Information Engineering (JULIE) Lab
Friedrich-Schiller-University Jena,
Jena, Germany

https://julielab.de

Introduction

Previous Work And Its Shortcomings



(Hamilton et al., EMLNP 2016)

Cook & Stevenson, LREC 2010; Jatowt & Duh, JCDL 2014; Buechel et al., LT4DH 2016

1. Reduces human emotion to polarity
2. No quantitative evaluation

Our Contribution

- First gold standard for historical word emotion (EN/DE)
  - Historical language experts instead of “native speakers”
  - Valence-Arousal-Dominance instead of polarity
- Evaluate previous approaches to historical word emotions
- Web service for visualizing emotion trajectories of words:

  JESEMEE (Hellrich et al., COLING 2018)

Building a Gold Standard for Historical Word Emotions

Emotion Lexica

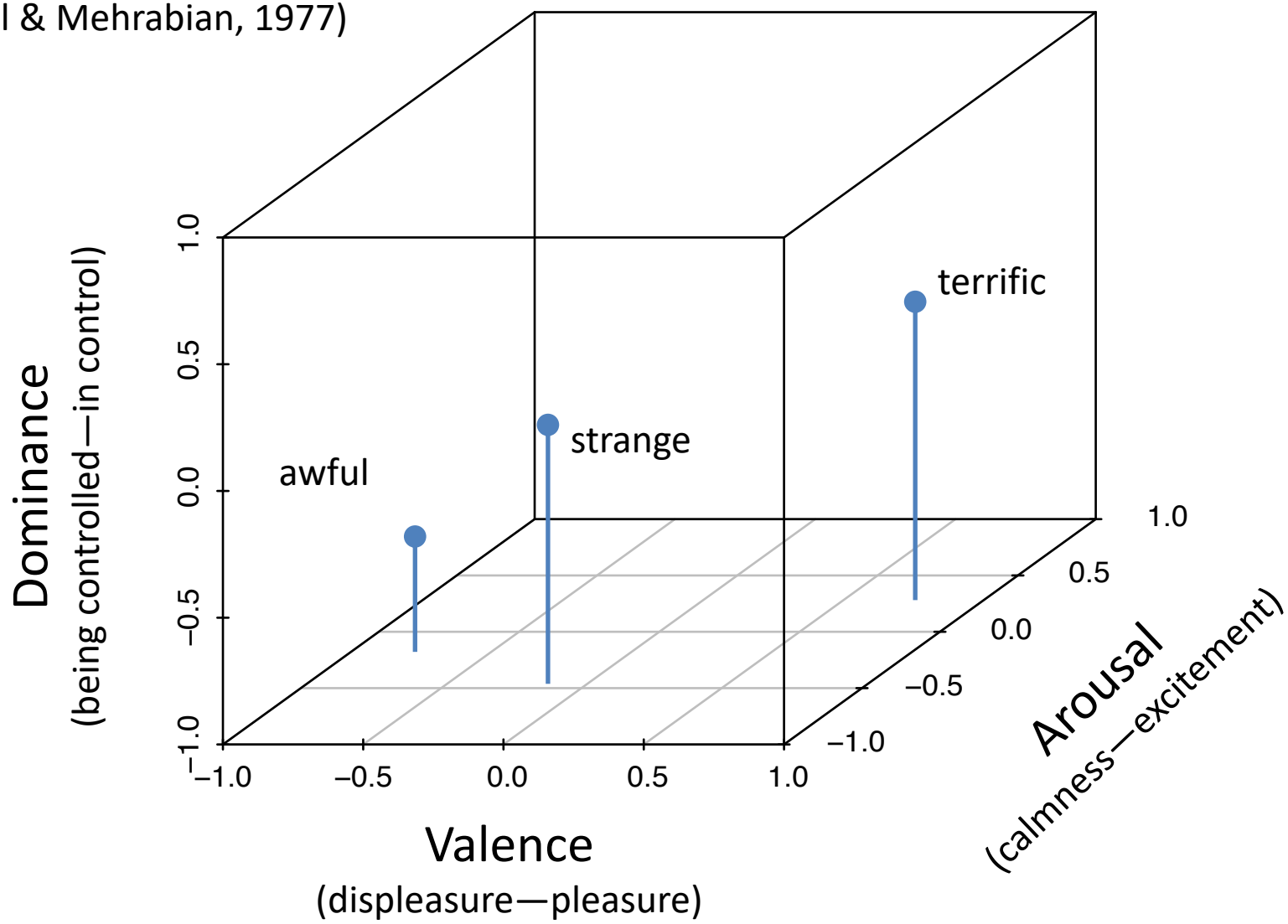
Lemma	Polarity
terrific	+
awful	–
strange	–

Emotion Lexica

Lemma	Emotion
terrific	
awful	
strange	

Valence-Arousal-Dominance

(Russell & Mehrabian, 1977)



Emotion Lexica

Lemma	Emotion		
	V	A	D
terrific	7.2	5.5	6.3
awful	2.3	4.9	3.0
strange	4.7	3.5	5.3

- Average ratings of multiple annotators
- Very popular in psychology
- Contemporary lexica are available for 13+ languages
  (Buechel & Hahn, LREC 2018)

Annotation Process

- Language stage around 1830
- Selection of raw data
  - English: COHA; German: DTA
  - selected 100 of the 1000 most frequent content words (good representations)
  - Too small for training but usable for evaluation
- Annotators
  - PhD students (EN 2, DE 3) experienced in interpreting 19th century texts
  - Asked to put themselves in position of person of that time
  - Best possible surrogate for actual native speakers
- Agreement comparable to contemporary emotion lexica

Examples from Gold Standard

	historical			modern		
	V	A	D	V	A	D
daughter	3.5	4.0	4.0	6.7	5.0	5.1
divine	7.0	7.0	2.0	7.2	3.0	6.0
strange	2.0	6.5	1.0	4.7	3.5	5.3

Methods for Modeling Historical Word Emotions

Overview of Considered Methods

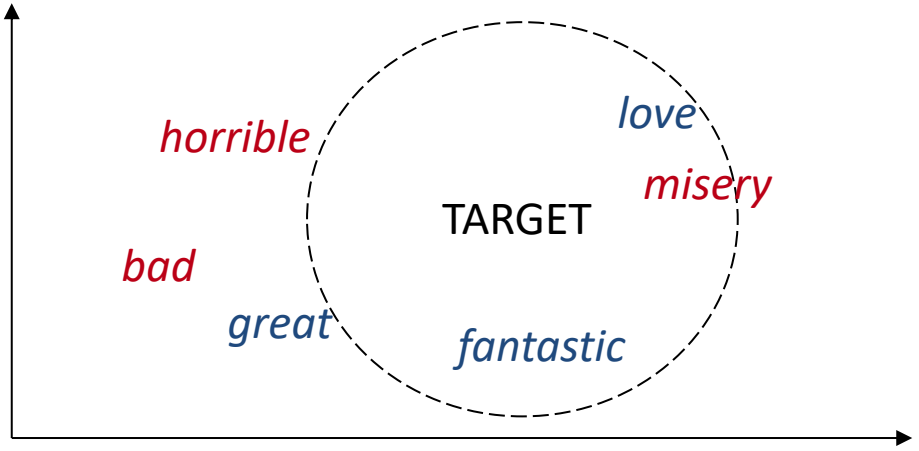
kNN

RandomWalk

ParaSim

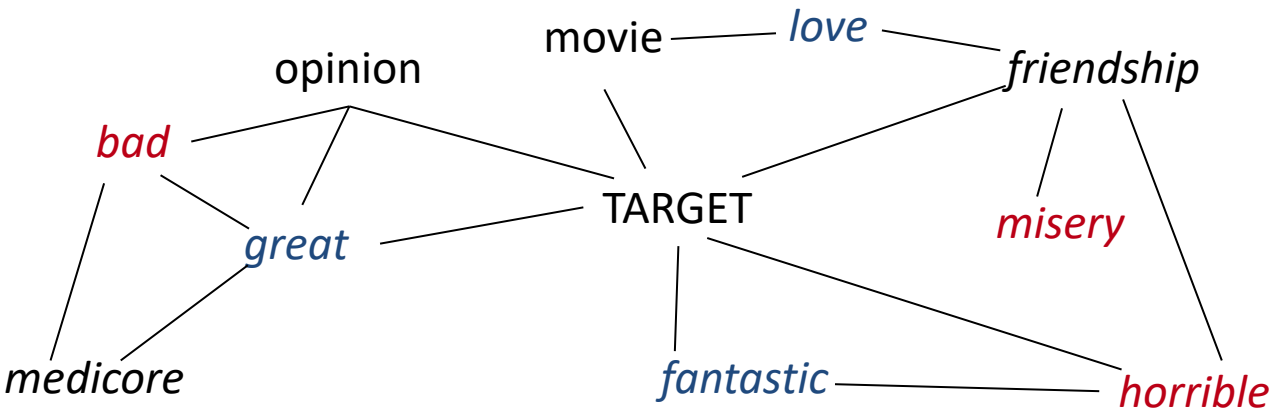
- Previously used in historical applications
- Predictions based on word embedding similarity

K Nearest Neighbor Regression (kNN)



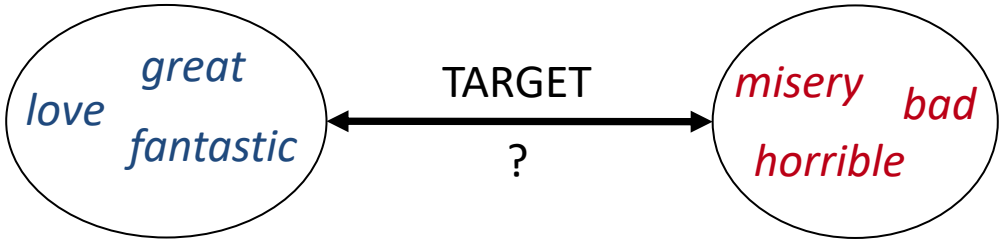
- Historical application: Buechel et al. (LT4DH, 2016)

Graph-Based Polarity Propagation (RandomWalk)



- Algorithm by Zhou et al. (NIPS 2004)
- Historical application: Hamilton et al. (EMNLP 2016)

Similarity to Paradigm Words (ParaSim)



- Turney & Littman (ACM TOIS 2003)
- Historical application: Cook & Stevenson (LREC 2010)
- Embedding similarity instead of word association
  (Buechel & Hahn, NAACL 2018)

Seed Word Selection Strategies

- Methods need seeds / training data
- Not enough historical ratings available
- Fallback to present-language emotion lexica
- Which part of the lexica do you use?
  1. **Full**: Use everything
  2. **Limited**: only semantically stable words (Hamilton et al., EMNLP 2016)

Experiments on Modeling Historical Word Emotions

Outline of Experiments

- Synchronic (background measure)
  - Full seed set: ANEW (1000 words; Bradley & Lang, 1999)
  - Limited seed set: Selection by Hamilton et al. (19 words; EMNLP 2016)
  - Test set E-ANEW (14K words; Warriner et al., 2013)
- Diachronic (actual experimental conditions)
  - Seeds as in synchronic experiment
  - Test set EN / DE historical gold standard

Outline of Experiments

- Synchronic (background measure)
  - Full seed set: ANEW (1000 words; Bradley & Lang, 1999)
  - Limited seed set: Selection by Hamilton et al. (19 words; EMNLP 2016)
  - Test set E-ANEW (14K words; Warriner et al., 2013)
- Diachronic (actual experimental conditions)
  - Seeds as in synchronic experiment
  - Test set EN / DE historical gold standard
- Reliability problem of embedding neighborhoods

  (Hellrich & Hahn, COLING 2016; Hellrich et al., RepEval 2019)
  - SGNS : stochastic optimization
  - SVD_{PPMI} : deterministic mathematical procedure

Outline of Experiments

- Synchronic (background measure)
  - Full seed set: ANEW (1000 words; Bradley & Lang, 1999)
  - Limited seed set: Selection by Hamilton et al. (19 words; EMNLP 2016)
  - Test set E-ANEW (14K words; Warriner et al., 2013)
- Diachronic (actual experimental conditions)
  - Seeds as in synchronic experiment
  - Test set EN / DE historical gold standard
- Reliability problem of embedding neighborhoods

  (Hellrich & Hahn, COLING 2016; Hellrich et al., RepEval 2019)
  - SGNS : stochastic optimization
  - SVD_{PPMI} : deterministic mathematical procedure
- Evaluation in Pearson correlation *r*

Synchronic Evaluation

Algorithm	Seed Set	SVD_{PPMI}	SGNS
kNN	full	**.55**	.49
ParaSim	full	**.56**	.49
RandomWalk	full	**.54**	.43
kNN	limited	.18	.17
ParaSim	limited	.25	.19
RandomWalk	limited	.33	.18

- Full seed set > set of stable words
- SVD_{PPMI} > SGNS

Diachronic Evaluation

Algorithm	Seed Set	SVD_{PPMI}	SGNS
kNN	full	**.31**	**.37**
ParaSim	full	**.35**	**.36**
RandomWalk	full	**.35**	**.36**
kNN	limited	**.27**	**.15**
ParaSim	limited	**.30**	**.23**
RandomWalk	limited	**.31**	.04

- Full seed set > set of stable words
- RandomWalk is quite jumpy
- SVD_{PPMI} competitive for English, superior for German (not shown; but otherwise consistent)

Main Findings

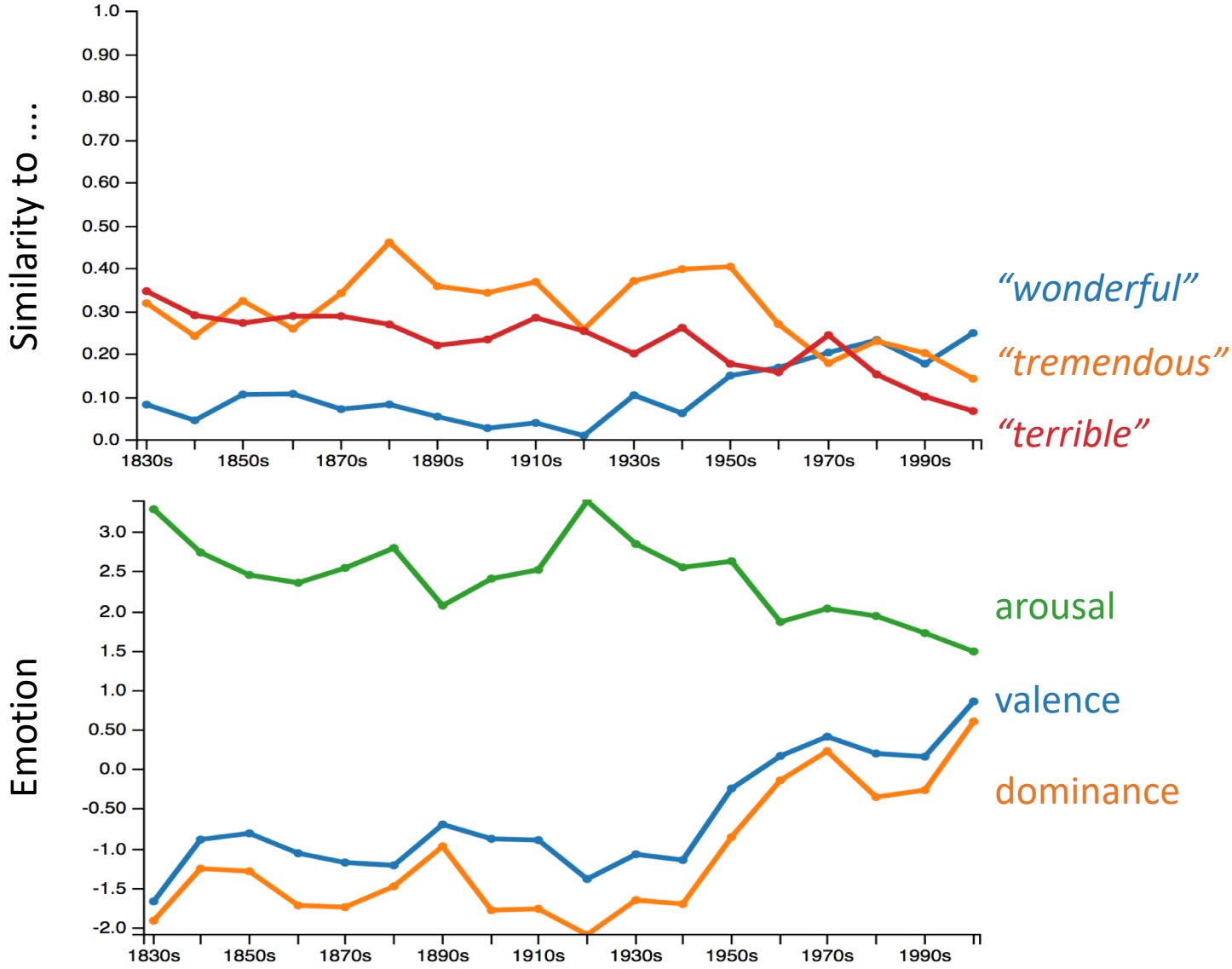
- **SVD_{PPMI}** about as good SGNS but stable
- **ParaSim** competitive + no hyperparameters
- **Full seed** set always outperforms limited one

JeSemE: Word Embedding Exploration for DH

http://jeseme.org/

(Hellrich, Buechel & Hahn, COLING 2018)

Meaning and Emotion of *Terrific* over Time



Conclusion

Conclusion

- Evaluation problem for historical language: no native speakers!
- First gold standard for 19th century word emotion by historical language experts: `https://github.com/JULIELab/histEmo`
- Evaluation of previous methodological approaches
  - Quantity beats stability regarding seed word selection
  - Insights incorporated into the JeSemE web tool: `http://jeseme.org`

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

MODELL ROMANTiK

DFG

Modeling Word Emotion in Historical Language:

Quantity Beats Supposed Stability in Seed Word Selection

Johannes Hellrich* Sven Buechel* Udo Hahn

Jena University Language and Information Engineering (JULIE) Lab
Friedrich-Schiller-University Jena,
Jena, Germany

https://julielab.de