Joint Spatiotemporal Topic-Sentiment Modeling of Social Media Data with Neural Networks

David Hanny[1]

[1]Faculty of Digital and Analytical Sciences,
University of Salzburg

Master Thesis Presentation

Supervisor: Assoc. Prof. Bernd Resch

Table of Contents

Motivation

Social media analysis in GI Science:

- Often used to model and understand real-world phenomena, e.g. natural disasters.
- Techniques:
  - Text processing: topic modeling, semantic analysis, sentiment analysis
  - Spatial (point-patterns, hotspot) and temporal analyses
  - (Image processing)
- Semantics, sentiment, space and time are usually considered separately.

$\rightarrow$ **Integrated multimodal topic model?**

Table of Contents

Methodological Contributions

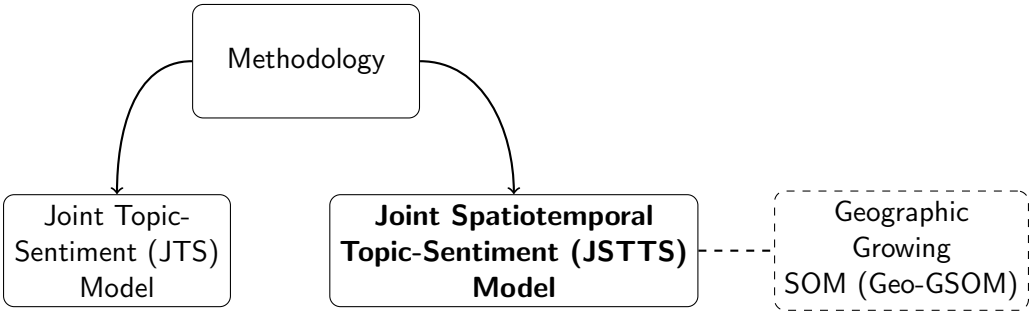

Figure: Core contributions of the thesis.

Model Architecture

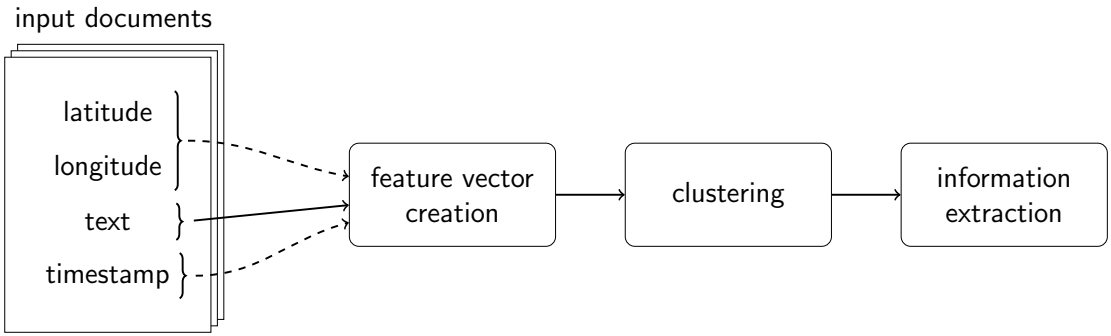

Figure: Overall architecture of the presented models.

Joint Spatiotemporal Topic-Sentiment Model

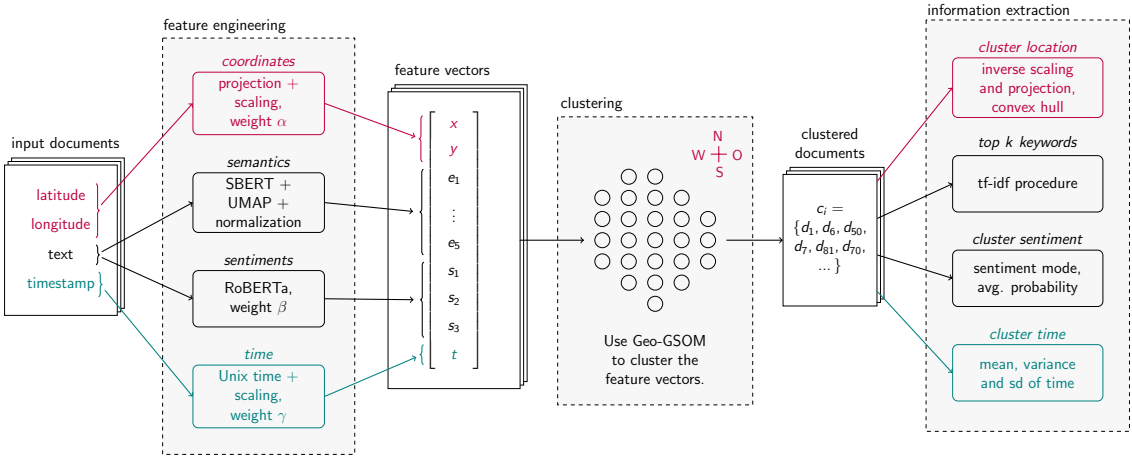

Figure: Methodological workflow of the JSTTS model.

Geographic Growing Self-Organising Map (Geo-GSOM)

Four-stage algorithm based on the GSOM[1] and Geo-SOM[2]: (1) Initialization, (2) Growing, (3) Smoothing, (4) Geographic Smoothing.

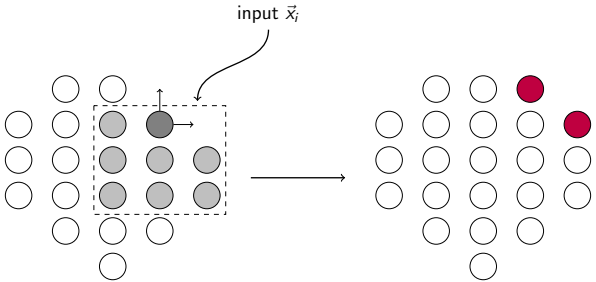

Figure: Visualisation of the growing process of new nodes.

[1] Alahakoon, Halgamuge, and Srinivasan 2000.
[2] Bação, Lobo, and Painho 2004.

Table of Contents

Quantitative Evaluation

- Comparison of JTS model against previous topic-sentiment models and a sequential approach using the TweetEval data set[3]:
  - Semantic topic quality (coherence $\cdot$ diversity) of 0.17 vs 0.12 over comparable approaches.
  - Sentiment classification accuracy of 0.72.
  - Sentiment uniformity of 0.9 within clusters.
- Evaluation of GSOM vs. Geo-GSOM on artificial test data ($I = 0.86$).
  - GSOM $\rightarrow \hat{I} = 0.47$
  - Geo-GSOM $\rightarrow \hat{I} = 0.77$
- Sensitivity analysis for JSTTS model.

[3] Barbieri et al. 2020.

Use Case Evaluation

Case study regarding the Ahr Valley flooding in July 2021 using 11 177 geo-referenced tweets.

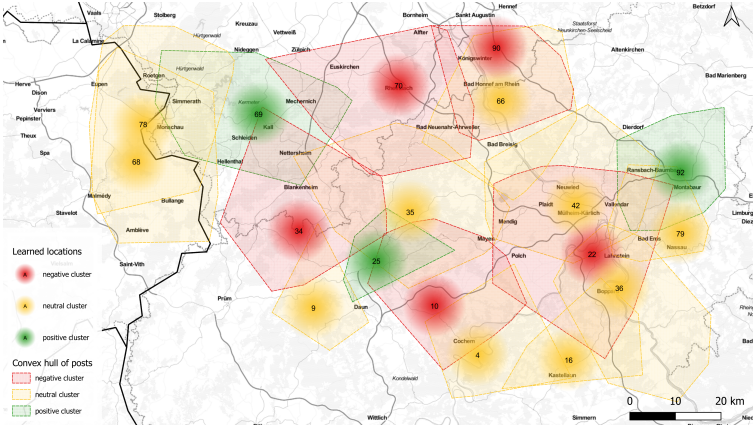

Figure: Cluster map of Ahr Valley tweets.

Example Output

cluster	size	top 8 keywords	sentiment	SI
4	66	cochem, mosel, pfalz, rheinland, notfallseelsorge, burg, einsatz, eltz	neutral	0.737
22	225	helfen, menschen, unbegreiflich, deutschland, einfach, katastrophengebiet, leid, loki	negative	0.790
35	110	nürburgring, mayen, einsatz, gerade, hochwasser, gepostet, foto, ahrweiler	neutral	0.708
36	108	koblenz, notfallseelsorge, einsatz, gerade, boppard, pfalz, gepostet, rheinland	neutral	0.780
66	174	shuttle, ahrweiler, foto, gerade, gepostet, germany, rheinland, info	neutral	0.777
70	352	leider, warum, hochwasser, einfach, ahrweiler, sehen, müssen, başkan	negative	0.787

Table: Selected model outputs for the Ahr Valley data.

Table of Contents

Limitations and Outlook

Some of the topics that are subject to further research:

- Better semantic topic representations, e.g. with generative language models.
- Additional input dimensions, e.g. images, movement patterns or person densities based on mobile phone data.
- Transferability to other use cases.
- Comparison with other clustering and embedding algorithms.