

WARUM WIRD WAS WIE KLASSIFIZIERT?

Scalable Reading + eXplainable AI am Beispiel historischer Lebensverläufe

Patrick D. Brookshire, M.A.

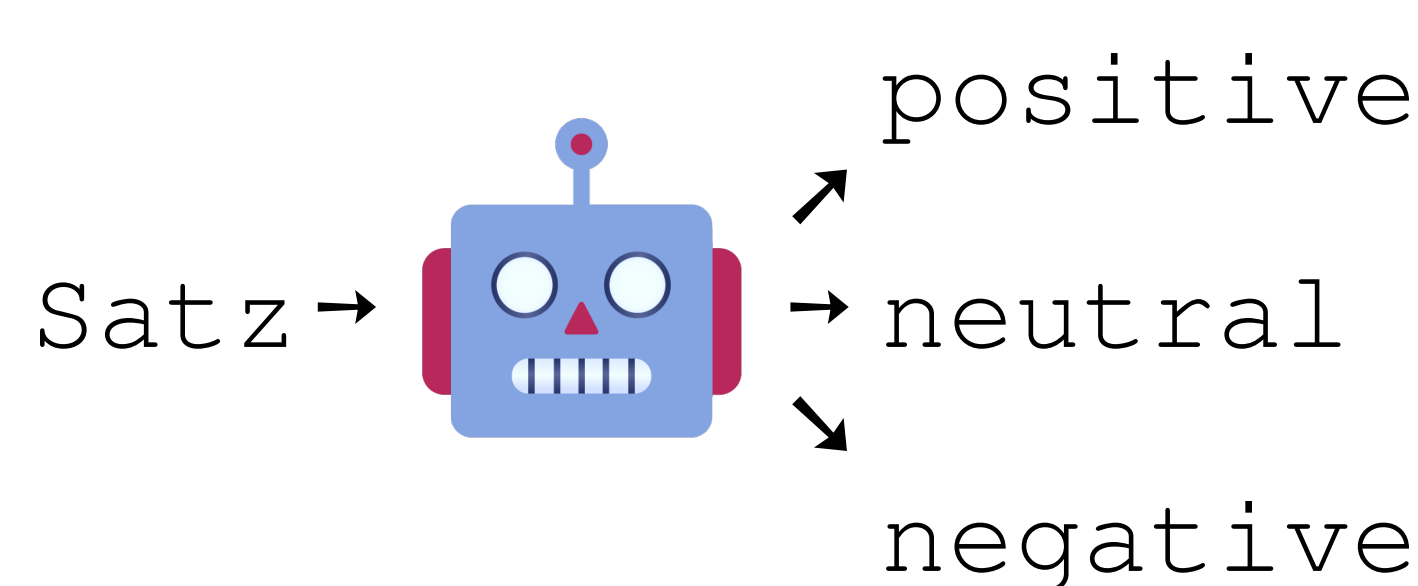
Hintergrund

Klassifikationen sind in den Digital Humanities häufig „under construction“. Bei kontextabhängigen Tasks sind **Deep-Learning-Ansätze** **performanter** als regelbasierte, aber **weniger interpretierbar** und ebenfalls nie zu 100% treffsicher.

→ Wie kann ein **modulares** und **iteratives Verfahren** zu (noch) reflektierteren und valideren Analysen führen?

1) Klassifikation

Bsp. satzweise Sentimentanalyse:



2) Transformation

nominal → **ordinal**

positive	→	+1
neutral	→	0
negative	→	-1

3) Segmentierung

Einzelklassifikationen (hier Sätze) werden mit unterschiedlichsten Strategien (z. B. Mittelwertbildung) zu größeren **Segmenten** (z. B. Textabschnitten gleicher Länge) zusammengefasst.

5) Erklärung

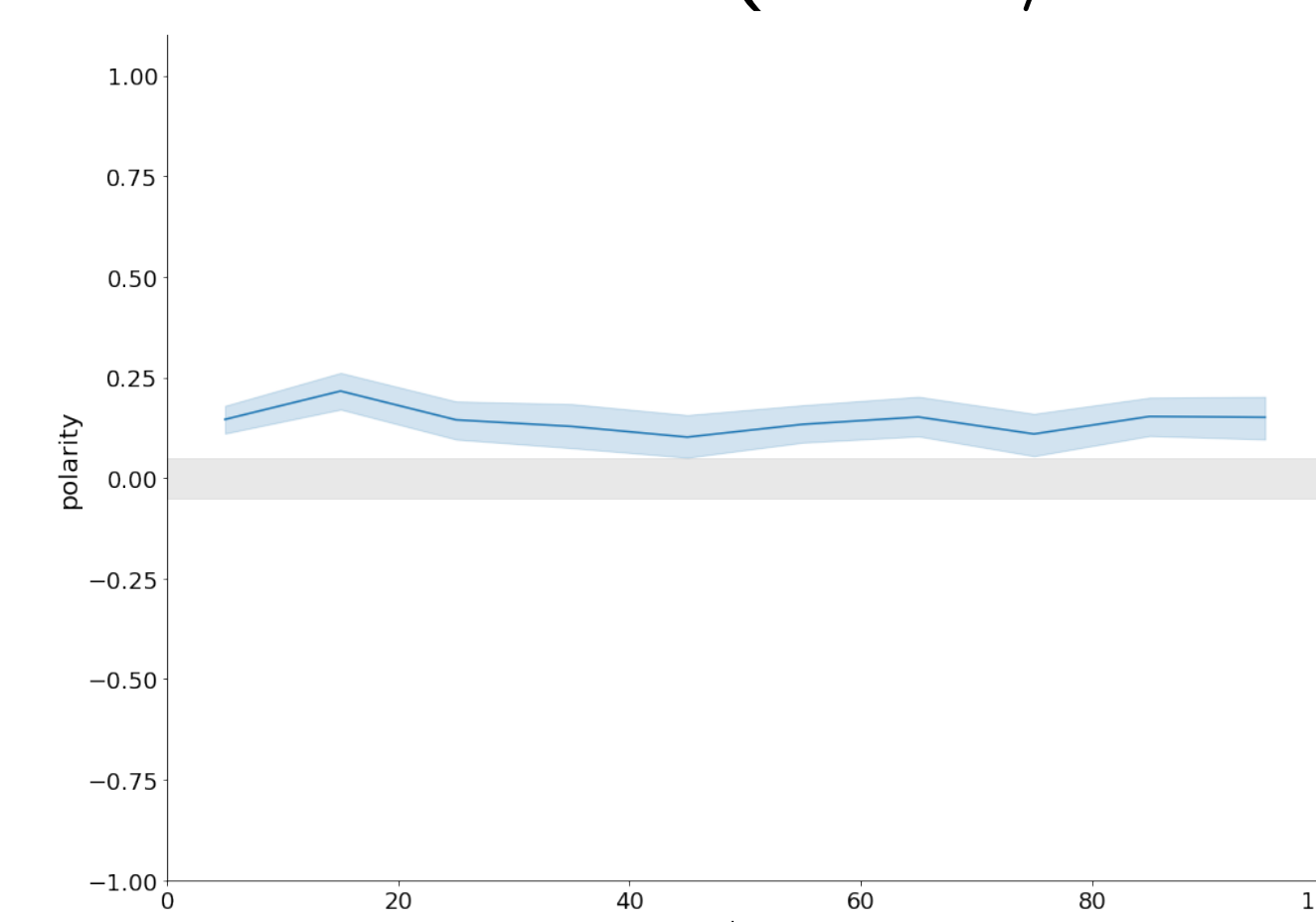
XAI-Modelle (z. B. BERT_explainability, Chefer et al. 2021) lassen sich **als weitere (Zoom-)Skala** einbinden, um relevante/interessante Klassifikationen nachzuvollziehen.

source	n	text	label	negative	neutral	positive
sfz80512	54	Während die gemeine Erkenntniß an den endlichen Gegensätzen des mannichfaltigen Daseins festhält und lediglich ihre Beziehungen und Verhältnisse zu ermitteln sucht, ist die höhere Erkenntniß das Wesen selbst zu seinem Inhalt, wie es seine endlichen Bestimmungen in sich erzeugt und diese Differenzierungen wieder in sich vereinigt.	neutral	0.66%	88.92%	10.42%
sfz80512	55	Nur in unserer endlichen Existenz sondert sich das Wesen von seinen Bestimmungen, ist unser Denken ist darin besteht, daß es von dem Allgemeinen zum Besonderen übergeht, und nur durch diesen Uebergang jedes der Entgegengesetzten als das auftritt, was es ist .	negative	66.20%	20.06%	13.74%
sfz80512	56	An und für sich ist die Erkenntniß Einheit des Allgemeinen und Besonderen, und also auch Einheit der Form und des Stoffes sein.	neutral	2.88%	86.32%	10.80%
sfz80512	57	Das aber ist die Grundbedeutung der Idee .	neutral	37.95%	61.11%	0.94%
sfz80512	58	Und weil nun die Idee als vollkommene Einheit der Stoffe mit der Form erkannt wird, so kann und ist sie in ihrer Richtung auf die Existenz auf zweifache Weise gefaßt werden: einmal nämlich als dasjenige, was die Einheit mit sich selbst in unser Bewußtsein, das andere Mal als das, was Einheit in die Gegensätze bringt, in welchen die äußeren Gegenstände unserer Erkenntniß miteinander stehen.	negative	94.80%	3.08%	2.12%
sfz80512	59	Die Ideen der ersten Art beziehen sich auf den Willen, die der zweiten auf die Welt der von unserem Bewußtsein unabhängigen Gegenstände oder die Natur.	neutral	1.14%	97.05%	1.81%
sfz80512	60	Was die Offenbarung des göttlichen Bewußtseins in uns wirkt, das ist die Kraft unseres eigenen Bewußtseins, indem es in die Gegensätze und Vermittlungen seiner eigenen Existenz verflochten ist, und die Erschaffung unseres eigenen wahren Wesens, welches in Wahrheit kein anderes als das göttliche selber ist.	negative	92.78%	2.36%	4.85%

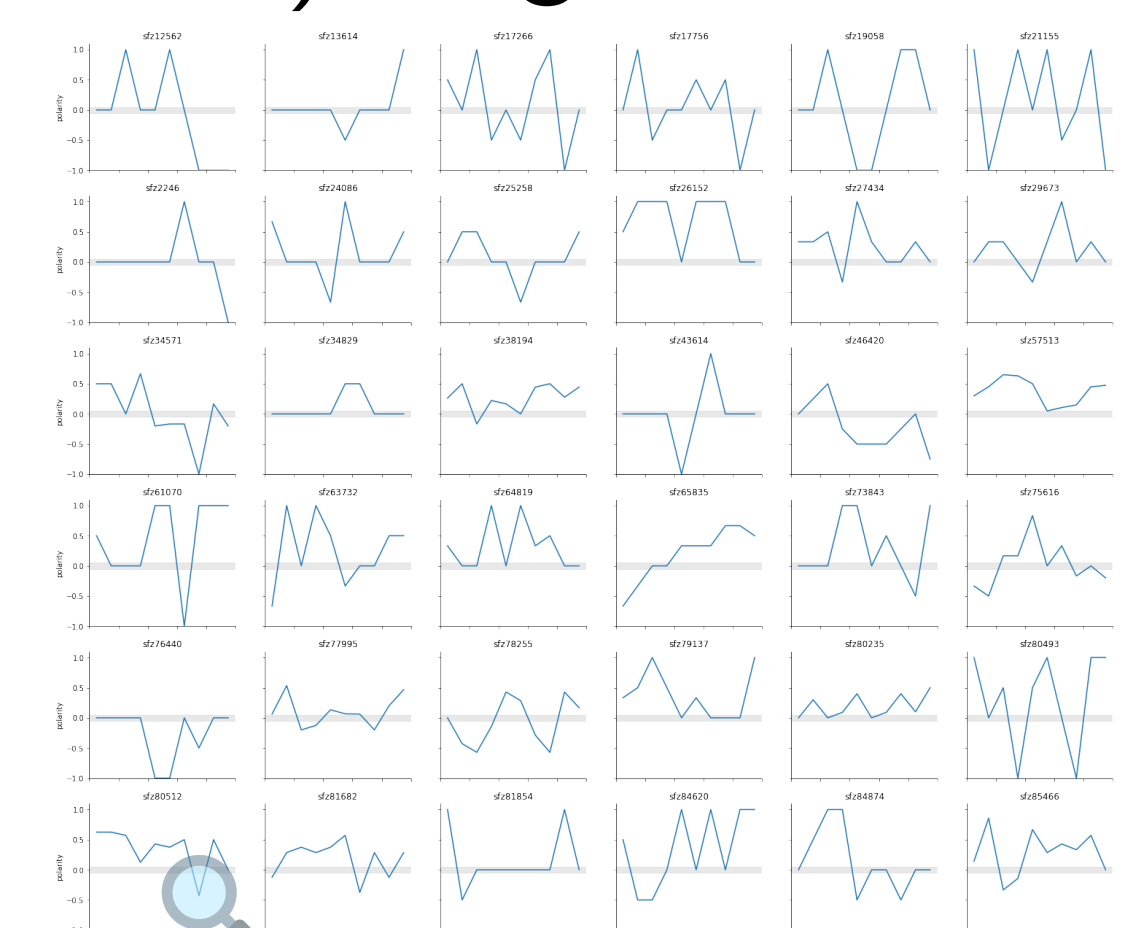
Auswirkung einzelner Subwordtokens auf Klassifikation

4) Aggregation

Einzeltextsegmente werden zu **Samples** (z.B. 360 ADB-Biographien aus dem 18./19. Jhd.) aggregiert, wobei ein **Skalenwechsel** (Rein-/Rauszoomen) möglich bleibt.



Polaritätsverlauf (Gesamtkorpus)



Polaritätsverläufe (Biographien)

Fazit

Wegen Segmentierungs- und Aggregationsschritten beeinflussen Fehler nachgelagerte Analyseschritte **unterschiedlich stark**. Das vorgeschlagene Verfahren unterstützt die **Identifikation relevanter Fehlklassifikationen** und sieht im Sinne eines „**Human-in-the-Loop**“ immer auch Korrekturen vor.

→ Die Methodologie ist auf beliebige Klassifikationsaufgaben übertragbar.

Literatur

- Brookshire, P. D. & Reiter, N. (2024). Modeling Moravian Memoirs: Ternary Sentiment Analysis in a Low Resource Setting. *LaTeCH-CiFL*, 91–100.
- Chefer, H., Gur S. & Wolf L. (2021). Transformer Interpretability Beyond Attention. *IEEE/CVF*, 782–791.
- Hutto, C. J. und Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Rule-Based Model for Sentiment Analysis of Social Media Text. *ICWSM*, 216–222.
- Jockers, M. L. (2015). Revealing Sentiment and Plot Arcs with the Syuzhet Package.
- Qi, P. et al. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *ACL-Demos*, 101–108.
- Reinert, M., Schrott, M. und Ebner, B. (2015). From Biographies to Data Curation – the Making of www.deutsche-biographie.de. *BD*, 13–19.
- Singh, L. G. und Singh, S. R. (2021). Empirical Study of Sentiment Analysis Tools and Techniques on Societal Topics. *Journal of Intelligent Information Systems* 56 (2): 379–407.
- Weitin, T. (2017). Scalable Reading. *Zeitschrift für Literaturwissenschaft und Linguistik* 47 (1): 1–6.

Code/Data

