

Distributional Semantics and Ground Truths for Neo-Latin

Digital Humanities and Neo-Latin Studies, 15 April 2021
Maria Chiara Parisi & Yvette Oortwijn

Overview

- Introduction
- Challenges
- Initial tasks
 - Vicipaedia definitional dataset evaluation
 - Neo-latin dataset evaluation
- Results on initial tasks and discussion
- Grounded evaluation
- Grounding evaluation: example
- Conclusion

Introduction

Aim: Constructing and evaluating distributional semantics models for Neo-Latin

Test case: Christian Wolff (1679-1745)'s mathematical method and its spread in 18th cent. Germany

Corpus: tiny Neo-Latin dataset, from bilingual German and Neo-Latin corpus

Challenges

- Latin word embeddings: highly inflected language, Latin vs. Neo-Latin
- Tiny data
- Domain specific data
- Evaluation

The Neo-Latin Corpusculum

- 30 target terms, for which we have 3 sentences each.
- The target term is replaced in the snippets with “___”.

Target	Snippet 1	Snippet 2	Snippet 3
Mathematica	methodum meditationis, et inveniendi, et probandi, potest huc commode referri methodus ___.	Unde reliquis omnibus praeferenda methodus ___.	Ubi methodus___ adhibetur, ibi & principia sunt indubitata, & modus concludendi legitimus.
Axiomata	inveniuntur ___ duobus fontibus adhibitis vel per inductionem vel per definitiones illa probatur omnium exemplorum inductiones [...].	quae enim immediate ex definitiones fluunt propositiones si theroreticae sunt ___ si practicae postulata vocantur.	secundò has ipsas definitiones in se considerabo et hinc deductas proprietates appellabo ___.
Definitio	ad cognitionem tria requiruntur ___ divisio demonstratio.	est autem ___ nihil aliud quam rei naturae ex qua ipsa componitur commoda explicatio.	et ___ quidem est propositio rem ita determinans ut ab aliis omnibus semper distingui possit.

Tasks

Two initial evaluations of Nonce2Vec on Latin data

- (1) **Vicipaedia** definitional dataset evaluation

Compare vectors from N2V one shot learning of terms based on definitional sentences from Vicipaedia to W2V over entire Vicipaedia

- (2) **Neo-Latin mathematical method** dataset evaluation

Consistency metric to check the stability of vector spaces generated by N2V on tiny Neo-Latin dataset

Results

Results of two initial evaluations of Nonce2Vec on Latin data

(1) **Vicipaedia** definitional dataset evaluation

N2V can learn Latin word embeddings from a single definitional sentence, though slightly less well than for English

(2) **Neo-Latin mathematical method** dataset evaluation

Consistent Neo-Latin word embeddings can be learned by N2V

Discussion of results

What do the results of these two evaluations mean?

- Is the quality good enough for **downstream application**?
- Do results on Vicipaedia definitional dataset **generalise** to other Neo-Latin data?
- Are the embeddings **consistently good or consistently bad**?

→ **Further evaluation is needed**

Grounding evaluation

Why grounding evaluation:

- Evaluation of each specific **task-*cum*-dataset**
- Distinguishing the **consistently good** from consistently bad
- Using a **predefined standard**

Grounding evaluation: example

Computational task: for a term, retrieve passages that are relevant to the meaning of this term

Evaluation method:

- **Ideally:** annotate all paragraphs of the corpus on their relevance to the research question/term
- **Practically:** six-step procedure (Betti et al., 2020) combining qualitative, quantitative and computational based on domain expert knowledge

Grounding evaluation: example

Step 1: initial research question & conceptual model

Are *axioms* grounded in *definitions*?

Step 2: term based research question, and term lists

Axioma: axiom*, princip*

Definitio: definiti*, essenti*

Deductio: deduc*, flu*, proba*, deriv*, demonstrat*

Step 3: check term list expansions

flu* → flumen ✗, fluō ✓

Grounding evaluation: example

Step 4: paragraph retrieval

Find all paragraphs with, e.g., a term from one of the lists in it, or a term from each.

Step 5: paragraph annotation

Quae enim immediate ex definitiones fluunt, propositiones, si theroreticae sunt, axiomata , si practicae, postulata, vocantur.	1
Alta definitio vel pleniter televisio alta definitio visa est unus ex modis televisionem signis mittendis, et genus televisionis digitalis, in quo nunc qualitas imaginum optime videtur.	-1

Grounding evaluation: example

Step 6: corpus-dependent concept-modeling ground truth

Compare performance of methods by using precision and recall of top 10/100/500 results.

Notes:

- Count all paragraphs that are not annotated as a no-hit
- This does not give an absolute performance score

Discussion

- For evaluation, a **combination of intrinsic and extrinsic evaluation** methods is necessary
- Evaluation **other types of tasks**
- When are results **“good enough”**?
- **Our use case:** beyond single-language models
→ matching Neo-Latin and German terms by word meaning, i.e. they relate to the same concept

Conclusion

- Ensuring a faithful interaction with data and its representation for humanities *can* and *should* depend on expert-constructed ground truths
- We have shown an example of one such **domain expert ground truth**
- **Testing** against expert-constructed ground truths will show applicability of DS methods to Neo-Latin data and whether **further development** is still needed

References & Related work

- Betti, A., & van den Berg, H. (2014). Modelling the History of Ideas. *British Journal for the History of Philosophy*, 22(4), 812–835. <https://1drv.ms/b/s!AgPq3zEkkYuOgYNIZz-93F2nMbXLCw>
- Herbelot, A., & Baroni, M. (2017). High-risk learning: Acquiring new word vectors from tiny data. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 304–309. <https://www.aclweb.org/anthology/D17-1030>
- Bloem, J., Fokkens, A., & Herbelot, A. (2019). Evaluating the Consistency of Word Embeddings from Small Data. *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2019*, 132–141. <http://lml.bas.bg/ranlp2019/proceedings-ranlp-2019.pdf>
- Betti, A., Reynaert, M., Ossenkoppele, T., Oortwijn, Y., Salway, A., & Bloem, J. (2020). Expert Concept-Modeling Ground Truth Construction for Word Embeddings Evaluation in Concept-Focused Domains. *Proceedings of the 28th International Conference on Computational Linguistics*, 6690–6702. <https://doi.org/10.18653/v1/2020.coling-main.586>
- Bloem, J., Parisi, M. C., Reynaert, M., Oortwijn, Y., & Betti, A. (2020). Distributional Semantics for Neo-Latin. *Proceedings of LTHALA 2020 - 1st Workshop on Language Technologies for Historical and Ancient Languages*, 84–93. <https://www.aclweb.org/anthology/2020.lt4hala-1.13>
- Oortwijn, Y., Van den Berg, H., & Betti, A. (2020). Ground Truths in the Humanities. *5th Workshop on Visualization for the Digital Humanities (VIS4DH)*. 2020 Workshop on Visualization for the Digital Humanities, Salt Lake City. <https://arxiv.org/pdf/2103.12841.pdf>
- Raji, I. D., Bender, E. M., Denton, E., Hanna, A., & Paullada, A. (2020). AI and the Everything in the Whole Wide World Benchmark. *Proceedings of the NeurIPS 2020 Workshop: ML Retrospectives, Surveys & Meta-Analyses (ML-RSA)*. NeurIPS 2020, Virtual.
- Oortwijn, Y., Bloem, J., Sommerauer, P., Meyer, F., Zhou, W., & Fokkens, A. (2021, May 24). Challenging Distributional Models with a Conceptual Network of Philosophical Terms. *Proceedings of the 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. NAACL 2021, Online.

Acknowledgements

We thank the organisers of this conference for inviting us.

This talk was based on joint work with Jelke Bloem, Martin Reynaert and Arianna Betti.

This research was supported by grants *e-Ideas* (VICI, 277-20-007) and *CatVis* (314-99-117), funded by the Dutch Research Council (NWO), and by the Human(e)AI grant *Small data, big challenges* funded by the University of Amsterdam.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).