Benjamin Matthias Ruppik

Heinrich Heine University Düsseldorf
Universitätsstr. 1
40225 Düsseldorf
NRW, Germany

benjamin.ruppik@hhu.de
https://www.ruppik.net/

1 Research interests

My research area is **topological deep learning**, where I apply techniques from **topological data analysis** to **word embedding spaces**. My goal is to use these mathematical methods to understand and improve dialogue systems.

1.1 Word embeddings

Word embeddings associate each word or token in a text corpus with a dense vector in an ambient space, such that words with similar meanings are close together. Figure 1 shows a section of the latent space produced by a pre-trained contextual language model. These language vectors form the backbone of many dialogue system components, especially for natural language understanding, dialogue state tracking and natural language generation. Natural language is thus modelled via **point clouds** in higher dimensional spaces, whose dimensions are usually in the hundreds, if not thousands. From the **shape** of word embeddings, a multitude of features can be extracted, to form the basis for various downstream tasks in dialogue system applications.

We expect representations of real-world datasets in higher-dimensional space to lie in proximity to lower-dimensional sub-manifolds. Typically, one suspects that data manifolds can be described locally by a handful of coordinates, modelled on a low-dimensional Euclidean space. In stark contrast to this **manifold hypothesis**, in previous work, our research group detected singularities in a *static* word embedding resulting from polysemous words (Jakubowski et al., 2020). At words which have multiple meanings, different pieces of the *static* word space appear to converge and coincide. Since modern language models utilize *contextual embeddings*, I have recently focused on applying similar topological tools to *contextual* latent spaces. This comes with new challenges, such as increased computational cost and difficulties in interpretability, but also offers very interesting new perspectives.

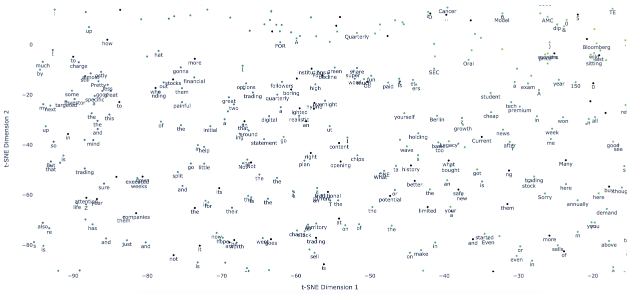

Figure 1: t-SNE projection of a subsection of the contextual embedding space produced by a `roberta-base` model. Topological data analysis is a collection of tools to study point clouds like this both locally and globally.

1.2 Topological Data Analysis

For us three-dimensional humans, high dimensional data spaces are hard to understand and visualize. **Topological data analysis (TDA)** is a mathematical toolkit which enables glimpses into high-dimensional point clouds by measuring their geometry at various scales. **Topology** is a branch of mathematics which studies the properties of geometric spaces that are invariant under continuous deformations. In data analysis, this involves studying the connected components, non-trivial holes and cavities of spaces derived from the data vectors. The major advantage of **topological features** is their invariance under small deformations and rotations, as opposed to using the embedding vectors directly. This leads to characteristics that are generalizable and not dependent on the exact data set used for training. To that end, I am interested in methods for defining and computing various flavours of **persistent homology** – topological features which can be detected at different size scales.

TDA has already shown its strengths in natural language processing, for instance, in probing model architectures. Kushnareva et al. (2021) build an increasing union of graphs from the attention scores of an input sentence in a pre-trained language model. The persistence features of the resulting objects are utilized for artificial text detection, a supervised classification task. Tulchinskii et al. (2023) topologically estimate the dimension of an embedding text paragraph to decide whether it was generated by an AI.

An important aspect of **topological machine learning**,

The 20th Annual Meeting of the Young Researchers' Roundtable on Spoken Dialogue Systems, pages 43–45
September 16–17, 2024. ©2024 Association for Computational Linguistics

the fusion of topological tools with deep learning, are vectorization methods. These transform the topological persistence diagrams into a format suitable for training downstream machine learning models. Lately, we made some progress towards including TDA feature extractors in the ConvLab-3 dialogue system code base. We have successfully applied topological features originating from *static* word embeddings to the **term extraction task** for dialogue datasets (Vukovic et al., 2022). Upcoming work extends this by showing that topological features from *contextual* word embedding spaces are even better suited for this task (Ruppik et al., 2024). This is demonstrated by showing that term extractors trained on the topological features of the MultiWOZ dialogue dataset can successfully transfer to another dialogue corpus, which contains different domains than the source datasets.

While these improvements themselves are promising, one should keep in mind that this method also leads to higher computational requirements. Our long-term research goal is to apply topological features for **unsupervised** applications, in particular the possibility of extracting meaningful **concepts** from unlabelled dialogue data to facilitate **ontology learning**.

2 Spoken dialogue system (SDS) research

Since November 2022, with the release of ChatGPT, I have been interacting with dialogue systems almost every day in one way or another: At work, they help me write emails, check grammar (including in this document), translate, answer coding questions, debug, and entertain me at home. These interactions have been mostly text-based, but as the voice mode improves and becomes more authentic, I can see myself using it more for speech in the very near future.

It is only a matter of time until such interactive dialogue agents are integrated into many aspects of our lives, such as customer support, appointment scheduling and booking, education, and all kinds of entertainment (games, AI companions, etc.). The most useful systems will be those which can be easily adapted to different domains. As a research community, we can contribute by investigating domain-agnostic architectures and representations.

Open academic research is more important than ever. We should not leave it to big tech companies to operate a walled garden of closed models, where we depend on them to get restricted access and always need to ask for permission to use state-of-the-art foundation models and build on top of them. This is especially important, since we should not need to trust secretive for-profit companies to be responsible for machine learning model use and deployment, and we should not leave all safety research to them.

One large problem that the academic community faces, including our lab at a public university in Germany, is restricted access to compute resources. This limitation prevents us from training foundation models and competing with large tech companies and their scale of operation. Nevertheless, I believe that our open basic research in academia is invaluable. Even with limited resources, we can come up with new and interesting ideas that a profit-oriented company might never have the motivation to explore. Academia, in particular, provides opportunities for cross-subject collaboration between different departments and people with vastly different backgrounds.

This interdisciplinary collaboration is especially relevant to me as I transitioned from pure mathematics to data science. I would appreciate seeing more results grounded in theoretical mathematics research find their way into practice. Effective collaboration between different subject areas depends on good communication. Mathematicians need to be able to identify possible applications of their methods, and consider implementations of algorithms and their efficiency. Dialogue systems researchers, on the other hand, need to point out places in their systems and pipelines where general methods might apply. They might need to step back from their specific implementations and identify parts that potentially work more generally than the specific setup or given dataset. The future state of dialogue systems research seems harder to predict than ever, but I believe that interdisciplinary collaborations will play a crucial role in it.

3 Suggested topics for discussion

1. Alternative evaluations of large language models: What other ways, apart from benchmarks – that might leak or suffer from overfitting – are there to evaluate language models? For example, can we intrinsically measure the quality of the embedding space produced by a model?

2. Prompting versus introspection: Should we focus on trying to apply the language generation capabilities of autoregressive language models to solve tasks, or should we take their representations and build on top of these? Can we learn everything about a model by observing how it answers the questions we present, or do we need to look into the internals of the models?

3. Reproducibility of research based on black-box language models, and their inaccessibility: How should we deal with the fact that we only have "prompting access" to the most powerful models, often hidden behind an API, and cannot observe and investigate their architecture in full?

References

Alexander Jakubowski, Milica Gašić, and Marcus Zibrowius. 2020. Topology of Word Embeddings: Singularities Reflect Polysemy. In *Proceedings of the Ninth Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, Barcelona, Spain (Online), pages 103–113. https://aclanthology.org/2020.starsem-1.11.

Laida Kushnareva, Daniil Cherniavskii, Vladislav Mikhailov, Ekaterina Artemova, Serguei Barannikov, Alexander Bernstein, Irina Piontkovskaya, Dmitri Piontkovski, and Evgeny Burnaev. 2021. Artificial Text Detection via Examining the Topology of Attention Maps. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, pages 635–649. https://doi.org/10.18653/v1/2021.emnlp-main.50.

Benjamin Matthias Ruppik, Michael Heck, Carel van Niekerk, Renato Vukovic, Hsien-Chin Lin, Shutong Feng, Marcus Zibrowius, and Milica Gašić. 2024. Local Topology Measures of Contextual Language Model Latent Spaces With Applications to Dialogue Term Extraction. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue (to appear)*. https://www.arxiv.org/abs/2408.03706.

Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey I. Nikolenko, Evgeny Burnaev, Serguei Barannikov, and Irina Piontkovskaya. 2023. Intrinsic Dimension Estimation for Robust Detection of AI-Generated Texts. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*. https://dl.acm.org/doi/10.5555/3666122.3667828.

Renato Vukovic, Michael Heck, Benjamin Ruppik, Carel van Niekerk, Marcus Zibrowius, and Milica Gašić. 2022. Dialogue Term Extraction using Transfer Learning and Topological Data Analysis. In Oliver Lemon, Dilek Hakkani-Tur, Junyi Jessy Li, Arash Ashrafzadeh, Daniel Hernández Garcia, Malihe Alikhani, David Vandyke, and Ondřej Dušek, editors, *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Edinburgh, UK, pages 564–581. https://doi.org/10.18653/v1/2022.sigdial-1.53.

Biographical sketch

Ben is a postdoctoral researcher in the Dialog Systems and Machine Learning research group led by Prof. Milica Gašić at the Heinrich-Heine-Universität Düsseldorf, which he joined in 2022. In collaboration with the Topology and Geometry group in the Mathematics Department, under the supervision of Prof. Marcus Zibrowius, Ben is developing applications of Topological Data Analysis in Natural Language Processing, focusing on dialogue systems. Before transitioning to machine learning research, Ben was a pure mathematician at the Max-Planck-Institute for Mathematics in Bonn, where he specialized in knotted surfaces in 4-dimensional manifolds. He graduated from the University of Bonn in 2022. Ben is supported by funds from the European Research Council (ERC) provided under the Horizon 2020 research and innovation programme (Grant agreement No. STG2018 804636) as part of the DYMO project.

Photo by Shutong Feng, 2023.