Poincaré Embeddings for Document Summarisation

supervised by Diego Antognini and Aleksei Triastcyn

Keywords: *document summarisation, hyperbolic geometry, embeddings*

Description

Following the rapid development of social media and accessibility of digital information there is a growing need to filter and summarise this information. The motivation for this ranges from reducing the time spent on browsing and searching for relevant information to filtering out false information.

Recent approaches to document summarisation make use of word embeddings and underlying graph structure of sentence relations [1]. At the same time, most real-world datasets, including texts, bare inherent hierarchical structure that may help in assessing semantic similarities between objects. The goal of this project is to improve the quality of such methods by incorporating hierarchical structure in the embedding space using hyperbolic geometry.

We will make use of recent work on hyperbolic embeddings based on Poincaré ball model [2] and integrate such embeddings in a document summarisation system. To further improve the quality of summarisation, we will experiment with various ways of getting graph structure from data and updating embeddings online.

Prerequisites

- Good knowledge of machine learning
- Strong mathematical background

References

[1] Yasunaga, Michihiro, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev. "Graph-based Neural Multi-Document Summarization." *arXiv preprint arXiv:1706.06681* (2017).

[2] Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." In *Advances in Neural Information Processing Systems*, pp. 6341-6350. 2017.