Analysis of Poincaré word embeddings and application to taxonomy induction

Mathia Haure-Touzé – Prof. Karl Aberer
Distributed Information System Laboratory
mathia.haure-touze@epfl.ch

Introduction
This project focuses on capabilities of embed symbolic data into hyperbolic space - using the n-dimensional Poincaré ball model [1]. Words embeddings, as input features of NLP tasks, make learning faster. Assuming a latent hierarchy of symbols, this representation captures similarity and hierarchy between concepts. From NER task to taxonomy induction we seek the embeddings performance.

Theory/Experimentations
A. Poincaré Ball model
It is a model of hyperbolic metric space \((X, d)\), on the open d-dimensional unit-ball where \((u,v)\in X\):
\[
d(u,v) = \text{arccosh}\left(1 + 2 \frac{|u-v|^2}{(1-|u|^2)(1-|v|^2)}\right)
\]
The distance varies smoothly with the location of \(u\) and \(v\), introducing a structural bias to represent hierarchies.

B. Learning embeddings
The learning step [1] assumes a latent hierarchy exists between English nouns such as WordNet [2] noun hierarchy. Nevertheless, no direct access is given to this nouns hierarchy, only a set of pair of words (i.e. “is-a” relations) are considered as input. From this set, a cross-entropy loss-function make closer related words (i.e. closer vectors \(u\) and \(v\)) and farther unrelated words using negative sampling (i.e. \(u\) and \(v'\in\text{Neg}(u)\))
\[
L(\theta) = \sum_{u,v\in D} \log \frac{e^{-d(u,v)}}{\sum_{v'\in\text{Neg}(u)} e^{-d(u,v')}}
\]

C. Poincaré embeddings analysis
First, we evaluate performance of the representation on NLP tasks. Embedding specialization [3] was used for NER task (e.g. find named entity as “[Brian, person] is in the [kitchen, location]”). Recently, benchmarks [4] have been proposed to evaluate embeddings on a wide range of tasks, methods, and dataset: categorization task, analogy task (e.g. “Germany” : “Berlin” :: “France” : “Paris”) and similarity task (e.g. “car” is similar to “bike” and associated to “road”).

D. Taxonomy induction
Noisy set of “is-a” relations are generated from WordNet and embeddings learned for each one. Hierarchies are reconstructed from the embeddings and simple heuristics inferred to extract subsets of WebIsA dataset [5]. First embedded, hierarchies are reconstructed to analyze embeddings capabilities to do taxonomy induction.

Results
The evaluation shows GloVe embeddings gives better results than the H-PCA and Poincaré embeddings (i.e. no improvement).

Noise sensitivity analysis is based on mean average precision, mean rank and precision of top-1 generalization candidates. Results bellow show the decreasing results with the increasing ratio of noise.

Conclusion/Perspectives
The Poincaré method allows to embed hierarchy but does not remove noise by itself. Future work could combine a noise removing loss function with the learning step to create a end-to-end optimized system which removes noise from dataset of “is-a” relations.

Acknowledgments
Prof. Karl Aberer from EPFL, Dr. Amit Gupta and Dr. Remi Lebret

References