

Graduiertenkolloquium Angewandte Informatik

Learning Latent Features using Stochastic Neural Networks on Graph Structured Data

M.Sc. Tobias Weller

Graph structured data are ubiquitous data structures, used to model relationships between entities. Graphs have become an important foundation to represent interactions between users in social networks, items in recommender systems, and interactions between drugs in bioinformatics. The main research problems in these areas include node clustering, node classification and link prediction. Especially the link prediction task is in bioinformatics of special interest toward the identification and development of new uses of existing or abandoned drugs since drug development is currently very time consuming and expensive.

In recent years, network embeddings are often trained for encoding the nodes of a network into a low-dimensional space whilst preserving the graph structure. Based on the trained embeddings, machine learning techniques are applied to address the main machine learning tasks, such as link prediction. In most of the published methods, like e.g. DeepWalk, node2vec and LINE, random walks procedures are used to efficiently explore diverse neighbourhoods and compute embeddings based on them. However, these methods assume a connected graph, otherwise the random walks are not sufficient to gather enough information of nodes in the neighbourhoods.

In this work we address three types of problems: Link prediction on bipartite networks, link prediction on knowledge graphs and a semantic grouping of nodes and links in graphs.

We use a stochastic factorization model to learn a target distribution over the graph structured data, allowing to predict unknown links and embed the nodes into a low-dimensional space whilst preserving the distribution of interactions within the graph.

Compared to the existing methods that use random walks, our approach is much more robust in terms of the connectivity of the networks.

Results show that the proposed method outperforms current state-of-the-art models in several studied complex networks and sets a new baseline in link prediction on disconnected graph structured data.

For this presentation we focus on the first two research problems, namely link prediction on bipartite networks and link prediction on knowledge graphs. We show how a stochastic factorization model is used to learn a target distribution on the graph structured data, based on the learned target distribution predict missing links and the benefit of our method in terms of the connectivity of the graph structured data.

Termin: Mittwoch, 27.Mai 2020, 15.45 Uhr

Ort: voraussichtlich Onlineveranstaltung

Join Zoom Meeting

<https://kit-lecture.zoom.us/j/92868749876?pwd=ZmF0NmJjRmdYeTByWGlrYVYyaHoxQT09>

Meeting ID: 928 6874 9876

Password: hFischer#1

Veranstalter: Institut AIFB, Forschungsgruppe Web Science

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

A. Oberweis, H. Sack, A. Sunyaev, Y. Sure-Vetter (Org.), M. Volkamer, J. M. Zöllner