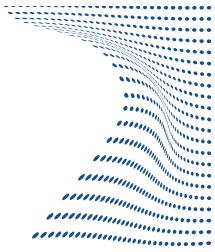
Knowledge Graph Embeddings for Recommender Systems

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR



Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (32^{nd} cycle)

Knowledge Graph Embeddings for Recommender Systems

By

Enrico Palumbo

******

Supervisors:
Prof. Elena Baralis, Politecnico di Torino
Dr. Giuseppe Rizzo, LINKS Foundation
Prof. Raphaël Troncy, EURECOM

Doctoral Examination Committee:
Prof. Paolo Cremonesi, Politecnico di Milano (referee)
Dr. Cataldo Musto, Università degli Studi di Bari "Aldo Moro" (referee)
Prof. Alejandro Bellogin, Universidad Autónoma de Madrid
Prof. Silvia Chiusano, Politecnico di Torino
Prof. Paolo Garza, Politecnico di Torino

Politecnico di Torino
2020

Abstract

Recommender systems are ubiquitous on the Web, improving user satisfaction and experience by providing personalized suggestions of items they might like. In the past years, knowledge-aware recommender systems have shown to generate high-quality recommendations, combining the best of content-based and collaborative filtering. The crucial point to leverage knowledge graphs to generate item recommendations is to be able to define effective features for the recommendation problem. Knowledge graph embeddings learn a mapping from the knowledge graph to a feature space solving an optimization problem, minimizing the time-consuming endeavor of feature engineering and leading to higher quality features. Thus, the main pillar of this thesis investigates the use of knowledge graph embeddings for recommender systems. In this thesis, we introduce entity2rec, which learns user-item relatedness for item recommendation through property-specific knowledge graph embeddings. entity2rec has been benchmarked with a set of existing knowledge graph embeddings algorithms (translational models, node2vec) that we have applied to the recommendation problem and with popular collaborative filtering algorithms on three standard datasets. entity2rec has shown to generate accurate and non-obvious recommendations, achieving high accuracy, serendipity, and novelty, and to be particularly effective when the dataset is sparse and has a low popularity bias. Furthermore, entity2rec is based on a recommendation model that encodes the semantics of the knowledge graph and can thus be interpreted and configured for a particular recommendation problem. entity2rec has also been tested in a cold start scenario with real new users through a web application called TinderBook. TinderBook is a web application that recommends books to users, given a single book that they like, leveraging an item-item relatedness measure based on entity2rec.

In addition to defining effective features, a crucial element for the quality of knowledge-aware recommender systems is the quality of the knowledge graph itself. Typically, when building a knowledge graph from a set of heterogeneous data sources,

duplicates are a major source of noise in the data. Thus, the second part of the thesis deals with the entity matching problem in the process of knowledge graph generation. In this thesis, we introduce "STEM: Stacked Threshold-based Entity Matching". STEM is a machine learning layer that can be 'stacked' upon existing threshold-based classifiers to improve their precision and recall for the entity matching task. STEM has been tested on three datasets from different domains (finance, music) using two different threshold-based classifiers (linear and Naive Bayes), significantly improving the quality of the entity matching. STEM has also been applied in the context of the European research project 3cixty in the creation of a tourist knowledge graph encompassing places and events of a city, enhancing the deduplication process, and consequently, the quality of the knowledge graph.

Finally, this thesis deals with the extension of the recommendation problem to temporal sequences, i.e. with Sequence-Aware Recommender Systems (SARS). Specific attention is devoted to the problem of learning to recommend tourist paths, sequences of tourist activities that can be of interest for a user. We propose the Path Recommender, an approach based on a Recurrent Neural Network (RNN) trained on sequences of user check-ins collected from Foursquare. The Path Recommender shows to outperform a set of competing sequence-aware algorithms (bigram, Conditional Random Fields) on a set of relevant metrics. An extended and ad-hoc version of the Path Recommender architecture is devised for the problem of automated playlist continuation and tested in the context of the RecSys2018 challenge, achieving the 14th position out of 33 participants in the creative track and the 36th position out of 113 participants in the main track. The Foursquare dataset that has been collected and the evaluation framework for SARS that has been defined in this work have become public resources available to the research community.