

A Survey on Knowledge Graph-Based Recommender System

장석규

Recommendation Systems

- **CF based**: user preference based on the similarity of users or item from the interaction data, no extracting features, suffers from the data sparsity and cold start problem
- **content based**
- **hybrid**: unify the interaction-level similarity and content-level similarity to alleviate problems

Knowledge graph into recommender Systems

- KG is heterogeneous graph that nodes function as entities, and edges represent relations
- users and user side information can also be integrated into the KG
- interpretability of recommendation results
- embedding-based method, path-based method, and the unified method

<DATA>

six of the cross-domain KGs are utilized in recommender systems in this survey, and we briefly introduce them as follows: Freebase [15] was launched in 2007 by Metaweb and was acquired by Google in 2010. It contains more than 3 billion facts and almost 50 million entities by 2015 [26]. Though it is a cross-domain KG, around 77% of its information is in the domain of media [27]. Currently, the data is available at Google's Data Dumps [28]. DBpedia [16] is an open community project, which was started by researchers from the Free University of Berlin and Leipzig University, in cooperation with OpenLink Software. The first version was released in 2007 and is updated yearly. The main knowledge is extracted from different language versions of Wikipedia, and DBpedia combines them in a large-scale graph structure. YAGO [17] (Yet Another Great Ontology) was introduced by the Max Planck Institute in 2007. It contains more than 5 million facts, such as people, locations, and organizations. It automatically extracts and unifies knowledge from Wikipedia and multiple sources, including WordNet [29] and GeoNames [30], then unifies them into an RDF graph. Satori [31] is a KG proposed by Microsoft. Similarly to Google's Knowledge Graph, which empowers the Google search engine, Satori has been integrated into the search engine Bing. Though publicly accessible documents about the Satori KG are limited, it is known that Satori consisted of 300 million entities and 800 million relations in 2012 [32]. CN-DBpedia [33] is the largest Chinese KG.

Process of recommendation system

- First, the system learns a representation u and v for given user u , and an item v
- Second, it learns a scoring function $f : u_i \times v_j \rightarrow \hat{y}_{i,j}$ (preference of u for v)
- Finally, the recommendation can be generated by sorting the preference scores for items

Collaborative Filtering(CF)

- CF assumes that users who share similar interaction records may be interested in same items
- explicit interaction(rating), implicit interaction (click and view)
- user-item interaction matrix is required
- sparsity of user-item interaction data, cold-start problem

- MEMORY BASED CF

- first, learn the user-user similarity from user-item interaction data
- then, unobserved items are recommended to a give user based on the interaction records of people similar to the specific user
- alternatively, some models learn the similarity among items and recommend similar items for a user based on the user's purchase history

- MODEL BASED CF

- attempts to alleviate the sparsity issue by building an inference model
- latent factor model: extracts the latent representation of the user and item from the high dimensional user-item interaction matrix (using matrix factorization) and then computes the similarity between the user and item with the inner product or other methods

Content based Filtering

- depict the user and item from the content of items (c.f CF from user-item interaction)
- assume that users may be interested in items that are similar to their past interacted items
- item representation: extracting attributes from the item's auxiliary information
- user representation: based on the features of personal interacted items

Hybrid Method

- leverage multiple recommendation techniques

IN THIS SURVEY, KG-based recommender systems leverage KG as the side information, combining the CF-based technique for more accurate recommendation.

OVERVIEW

- Heterogeneous Information Network (HIN): directed graph with an entity type mapping function and a link type mapping function
- Knowledge Graph: directed graph whose nodes are entities and edges are subject-property-object triple facts, instance of a HIN
- Meta-path: we can define new composite relation $R_1R_2...R_k$ between type A_0 and A_k if they are connected remotely and have other connected entities between
- Meta-graph: combination of meta-paths, more expressive structural info between entities
- Knowledge Graph Embedding: embed KG into a low dimensional space
- User Feedback: binary user feedback (1: implicit interaction)
- H-hop Neighbor
- Relevant Entity
- User Ripple Set
- Entity Ripple Set

Methods of Recommender Systems with Knowledge Graph

1. Embedding-based methods

- translation distance models, semantic matching models
- item graph: KGs are constructed with items and their related attributes, users are not included

- leverage the KGE algorithms to encode the graph and then integrate the item side info into recommendation framework (latent vector of each item is obtained from KG, user-item interaction matrix, item's attributes and latent vector of each user is obtained from user-item interaction matrix, interacted items; embedding)

- get preference from preference score function that is calculated with embeddings of user and item

- user-item graph: both attribute-level relation and user-related relations serve as edges

-

2. Path-based methods

3. unified methods