Notes taking - Build is ctrl alt b

Navey Mavey

August 30, 2023

Chapter 1

Knowledge Graph

1.1 Definition

A knowledge graph, also known as a semantic network, represents a network of real-world entities—i.e. objects, events, situations, or concepts—and illustrates their relationship. This information is usually stored in a graph database and visualized as a graph structure, prompting the term knowledge "graph."

1.1.1 Ontologies

- serves to create a formal representation of the entities in the graph - based on a taxonomy (since there are multiple, ontologies has its own definition) - Ontology example: Madison Square Garden - distinguish between events at the location using variables such as time

1.1.2 how it works

- schemas -¿ Provide the framework for the knowledge graph - identities -¿ classify the underlying nodes appropriately - context -¿ determines the setting in which that knowledge exists - help distinguish words with multiple meanings

- fueled by machine learning, NLP to construct a comperhensive view of nodes, edges and labels through a process called semantic enrichment

1.2 Combining Graph Neural Networks and Sentence Encoders for Knowledge-aware Recommendations

- first exploited **graph neural networks** to encode both **collaborative features**, such as the interactions between users and items, and structured properties of the items - Next, we used a **sentence encoder** that relies on trans-

formers to learn a representation based on textual content describing the items.
- Finally, these **embeddings are combined** by **exploiting a deep neural network** where both *self-attention* and *cross-attention* mechanisms are used to learn the relationships between the initial embeddings and to further refine the representation

1.3 Interactive Recommender System via Knowledge Graph-enhanced Reinforcement Learning

- https://arxiv.org/pdf/2006.10389.pdf
- https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da
- Reinforce learning
  - a process in which an agent leans to make decision through trial and error
  - modelled mathematically as a Markox decision process
  - there is a learner and a decision maker called agent and the surrounding with which it interacts is called evironment
  - environment, in return, provides rewards and a new state based on the actions of the agent
- Q learning
  - Q-learning finds an optimal policy in the sense of maximizing the expected value of the total reward over any and all successive steps, starting from the current state.
  - Reinforcement learning in IRS

    model based techniques -¿ utilize pol -icy iteration to search for the optimal recommendation polcy where an action is defined to bean item and a state is represented as n-gram of items
  - model free techniques - more popular lately (including policy gradient based) - pg based learn a stochastic policy as a distribution over the whole item space and sample an item according to such distribution
- KGQR (Knowledge Graph Enhanced Q-Learning Framework for Interactive Recommendation)
  - integrate graph learning and sequential decision making as a whole to facilitate knowledge in KG and pattern mining in IRS

  - to alleviate data sparsity, the user feedback is modeled to propagte via strcture information of KG so tha the userÄs preference can be transited among correlateditems
  - This way one interctive record can affect multiple connected items, thus the sample efficiency is imporved
  - Aggregating the semantic correlations among items in KG, the item embedding and the user's preference are effectively represented -¿ leads to more accurate Q-value approximation
  - 4 main components
    * graph convolution module
    * state representation module
    * candidate selection module
    * Q-learning network module
  - at each timestep, the IRS squentially recommends item to user and correspondingly updates its subsequent recommendation strategy based on user's feedback
  - IRS models user's preferece via graph covolution module and state representation
    * interaction history
    * knowledge graph
  - state representation - Graph convolutional embedding layer - Behavior aggregation layer

1.4 Knowledge graph embedding based on semantic hierarchy

- The knowledge graph is mapped to a polar coordinate system, where concentric circles naturally reflect the hierarchy, and entities can be divided into modulus parts and phase parts, and then the modulus part of the polar coordinate system is mapped to the relational vector space through the relational vector, thus the modulus part takes into account the semantic information of the knowledge graph, and the phase part takes into account the hierarchical information.
-

1.5 Building Knowledge Graphs and Recommender Systems for suggesting Reskilling and Upskilling Options from the Web

- the adaptation of knowledge extraction methods to the human resources and continuing education domain

- developing a knowledge-driven recommender system that draws upon this background knowledge to support users in identifying useful reskilling and upskilling options
- slot filling -¿ combines multiple pieces of information (eg. skills, learning outcomes) into a single knwledge base entries
- When applied to open-world scenarios, slot filling is very challenging, as demonstrated by competitions such as the TAC 2017 Cold Start Slot Filling Task in which even the winning systems only obtained F-measures below 20 %

1.6 Knowledge Graph Convolutional Networks for Recommender Systems

- we propose a knowledge graph convolutional netwrok (KGCN) - end to end framework that captures inter-tiem relatedness effectively by mining their associated attributes on KG
- a recent study talks about that attributes are not isolated but linked up with each other, which forms knowledge graph
- benefit of incorperating KG into recommendation benefit
  - rich semantic relatedness among items in a KG can help explore their latent connections and improve the precision of results
  - various types of relation in a KG are helpful for extending a user's interests reasonably and increasing the diversity of recommended items
  - KG connects a user's historically-liked and recommended items, thereby bringing explainability to recommender system and increasing the diversity of recommended items
  - KG connects a user's historically liked and recommended items, thereby bringing explainability to recommender system.
- idea of KGCN - aggregate and incorporate neighbourhood information with bias when calculating the representation of a given entity in the KG
- three types of aggregators
  - sum aggregator

    $$agg_{sum} = \sigma(W \cdot (v + v^u_{S(v)}) + b)$$

  - concat aggregator

    $$agg_{concat} = \sigma(W \cdot concat(v, v^u_{S(v)}) + b)$$

- neighbor aggregator

$$agg_{neighbour} = \sigma(W \cdot v^u_{S(v)} + b)$$

- Results
  - KGCN can well address sparse scenarios
  - neighbour sampling size is best at K=4 or 8
  - depth of receptive field is best at 1 or 2

1.7 Attentional Graph Convolutional Networks for Knowledge Concept Recommendation in MOOCs in a Heterogeneous View

- for knowledge concept recommendation in MOOCs
- we look into ACKRec but it suffers from sparsity issue so we use attention based CGN to learn the representation of different entities
- considering other type of entities and construct a Heterogeneous information network to capture the corresponding fruitful semantic relationships among diffenet types of entities and incorporate them into the representation learning process
- PROBLEM STATEMENT: curren MOOCs recommendation might overlook students' interest to specific knowledge concepts (eg. computer vision courses taught by different instructors may be quite different in a microscopic views)
- benefit of heterogeneous relationships system
  - semantic relatedness among knowledge concepts can be introduced and help to identify the latent interaction
  - user's interests can be rasonably extended and the diversity of recommended knowledge concet can be increased
  - user's interest can be interpreted by tracking a user's historical records along these relationships.
    - use meta paths as the guidance to capture the heterogeneous context information in a HIN via GCN
- System architecture
  - feature extraction
    - content feature - generate the word embedding of the name of the knowledge concept and use it as a content feature for knowledge concept

    * context feature - word embedding of knowledge concept names can be used to represent information of a knowledge concept
  - meta-path selection
    * We model MOOCs data as HIN including 5 entities (user(U), course(C), video(V), teacher (T), knowledge concept (K))
    * Network schema - represent semantic and relations information comprenehsively in the MMOCs dataset
    * Meta-path - describes a composite relation between object N and N_{l+1}
  - representation learning of heterogeneous entities - proprsed to learn the lower dimensional representations of the entities in a heterogeneous view
  - Rating prediction - dense vectors of entities are fed to an extended matrix factorization to learn the parameters of the model

1.8 Recommending Knowledge Concepts on MOOC Platforms with Meta-path-based Representation Learning

- focuses on task of remmoending knowledge concets of interest to users, which is challenging due to the sparsit of user-concept interactions given a large number of concepts. We discuss HIN to learn user and concept representations using GCN based on user-user and concept-concept relationships via meta-paths
- understanding a user's learning needs from a micoscopic view and predicting knowledge concepts that the user might be interested in are important
- Based on different indirect paths chosen, we can derive various user (concept) representations, and those representations of users (concepts) regarding different paths can be aggregated, e.g., using the mean of those representations.
- contribution in this works
  - an end-to-end framework for predicting and recommending knowledge concepts of auer's interestinvestigate 2 attention mechanisms for aggregating information from different meta-paths to dervie user and cencept representations
  - we evaluate our approach with several baselines and state-of-the-art approaches in terms of well-established evaluation metrics
- related work

- YouEDU - pipeline for classifying MOOC forum posts and recommending instructional video clips that might be helpful to reseove the confusion
- LeCORe - exploits user interest modelling for recommendig courses as well as similar users for promoting peer learnign in enterprise environment
- to use recommendation with HIN - using pre-trained user and item embedings based on meta-path infromation with random walk
- in this paper,
  * they formulate interacted concepts for each user as implicit feedback while treated the number of clicks as ratings and formulated the problem as rating prediction for recommending top-k unkown concepts with highr ratings
  * investigate different attention mechanism including the one incorporating the latent features of users(items) from matrix factorization
- proposed approach (MOOCIR - MOOC Interest Recommender)
- We extend the MF with user (concept) representations/embeddings that are learned by applying GCNs to meta-path-based-Graphs
- Meta-path selection
  * for each meta path. a homogeneous graph with respect to users(concepts) can be extracted, which is depicted as its corresponding adjacency matrix
- GCNs
  * learn the representation for each meta-path
  * the output representarion of the last layer can be used as user(concept) representation
- Attentions
  * attention mechanism is motivated by how we pay visual attention to different regions of an image or relevant words in one sentence.
  * Different meta-paths can have different importance with resepect to each user
  * incorporating the importance with resepct to each user and incorporating the importance of each meta path differently for each user can be beneficial when aggregating user represantations from different meta paths.
- prediction

• evaluation
  - Evaluation matrix - Hit ratio of top-k concepts (HR@k), Normalised discounted cumuative gain (nDCG@k), Mean reciprocal rank (MRR)

Chapter 2

Natural Language Processing

- Rule based modelling of human language

2.1 Tasks

- Speech Recognising - Part of speech tagging (whether its a noun or a verb etc. as some words can be both - word sense disambiguation (semantic analysis of word) - Named entity recognition -¿ eg. Fred is a name, Germany is a location, etc - Co-reference resolution -¿ 2 words refer to the same entity. eg. Lea = she but could also be a metaphor or idiom - Sentiments analysis: attempt to extract subjective quality (sarcasm, confusion) - Natural language generation

2.2 BERT

- train deep bidirectional representation of unlabeled text by joint conditioning on both left and right context in all layers

Chapter 3

Learning as a Network (LaaN)

builds upon connectivism, complexity theory, and double-loop learning. It promotes a theory of change, movement, dynamism, self-organization, emergence, and effectiveness, which puts the learner at the center and represents a knowledge ecological approach to learning.

Chapter 4

GCN

4.1 Neural Graph Collaborative Filtering

- 2 componentsin learnable CF model
  - embeddong - transforms users and items to vectorized representations
  - interaction modelling - reconstructs historical interactions based on the embedding
- most embedding function lacks an explicit encoding of the crucial collabotaritve signal(important in user-item interaction to reveal the behavioral similarity between users)
- we tackle this problem by exploiting high-order connectivity from user-item interaction
- a neural network method to propagate embeddings recursively on the graph. By stacking multiple embedding prpagation layers, we can enforce the embeddings to capüture the collaborative signal in high-order connectivities (eg stacking 4 layers we get $u_1 \leftarrow i_2 \leftarrow u_2 \leftarrow i_4$
- Discussion
  - NGCF generalises SVD++ as SVD++ is a special case of NGCF with no high-order proagation layer

4.2 LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation

- 2 most common designs in GCNs, feature transformation and nonlinear activation, contributes little to the performance of collaborative fitlering, instead they add difficulty

- simplify the design of GCN to make it more concise and appropriate for recommendation
- Basic idea: Collaborative filtering (matrix factorization, SVD++, Neural attentive item similarity (NAIS))
- NGCF (Neural graph collaborative fitlering) - deepens the use of subgraph structure with huge-hop neighbours.
  - takes inspiration from GCN
  - follows propagation rule to refine embeddings
  - heavy and burdensome - not useful for CF
  - semi supervised node classification
    * each node has rich semantic features as input, such as the title and abstract words of a paper
  - it works by propagating L layers, nGCF obtains L+1 embedding to describe user and an item
  - then it concatenates these L+1 embeddings to obtain the final user embedding and item embedding, using inner product to generate the prediction score
  - σ (not linear activation function) and feature transformation matrix W1 and W2 are not as useful in collaborative filtering
  - it might be useful in semi-suprevised learning for features learning (eg. of title and abstract words of a paper)
  - in collaborartive filtering each node of user item interaction graph only has an ID
- LightGCN includes the most essential GCN component - Neighbourhood aggregation - for collaborative filtering
- BASIC IDEA : learn representation for nodes by sommthing features over the graph
$$e_u^{kt1} = AGG(e_u^{(k)}, \{e_i^{(k)} : i \in \mathcal{N}_u\})$$
- AGG : aggregation function - considers the kth layer's representation
- we aggretgate only the connected neighbours and not the target node itself (self connection) since the layer combination operation essentially does the same thing
- Layer combination and model prediction - ak is importance of the kth layer embedding -¿ treated as hyperparameter
- Perfomance comparison with NGCF
  - in all cases Light GCN outperforms NGCF by a large margin.

- increasing the number of layers can improve the performance but the benerfit diminish -¿ using 3 layers leads to satisfactory performance in most cases
- LightGCN fits training data better than NGCF (lower training loss)
- even comparing with state of the art (see table) its still better
- even using 4 layers, the performan e does not degrade

4.3 Interest-aware Message-Passing GCN for Recommendation

- to tackle over-smoothing problem in LightGCN in the domain of recommendation
- existing GCN-based recommendation models have not distinguished the high-order neighbors, and just simply aggregate the messages from all those neighbors to update user embeddings.
- As a result, the embeddings of dissimilar users are also involved in the embedding learning of a target user, negatively affecting the performance.
- result graph: LightGCN gets worse aas the number of layers increases whereas IMP-GCN does not