CSIT5210: Data Mining and Knowledge Discovery

(Fall 2022)

Group Project Proposal

Project Type: Implementation

Towards Universal Sequence Representation Learning for Recommender Systems

GROUP 09

20928868 WANG, Jiatong

20881383 YAN, Xi

20927307 CHANG, Zijiang

20929070 ZHONG, Lintao

Sequential Recommendation is a recommendation system paradigm that recommends relevant items to users by modeling user behavior and project patterns in time series. There are two kinds of objects in the recommendation system, namely, user and item, which include several interactions in the time dimension, such as user browsing, clicking and purchase conversion.

Traditional recommendation algorithms, such as matrix decomposition and wide & deep learning algorithm generally regard user's historical behavior as static information and sequence recommendation can model the sequence of user behavior by arranges these interaction behaviors in chronological order, such as the sequence of purchasing items, to learn the changes in user interests and the dependencies of different items in the sequence.

In order to develop effective sequential recommenders, a series of sequence representation learning (SRL) methods are proposed to model historical user performances. From early matrix factorization (e.g., FPMC) to recent sequence neural networks (e.g., GRU4Rec, Caser and Transformer). These approaches have largely raised the performance bar of sequential recommendation.

However, most existing SRL methods rely on explicit item IDs for developing the sequence models to better capture user preference. A major issue with this modeling way is that the learned model is difficult to be transferred to new recommendation scenarios, even when the underlying data forms are exactly same due to the limitation by explicitly modeling item IDs.

Such an issue limits the reuse of the recommendation model across domains. As a common case, we need to re-train a sequential recommender from scratch when adapting to a new domain, which is tedious and resource-consuming.

Recently a number of studies have been proposed by learning either semantic mapping or transferable components , in order to bridge the domain gap and enhance the item representations. However, these existing attempts cannot fully solve the fundamental issue.

To tackle this issue, this paper presents a novel universal sequence representation learning approach, named **UniSRec**. The proposed approach utilizes the associated description text of items to learn transfer-able representations across different recommendation scenarios. Extensive experiments conducted on real-world datasets demonstrate the effectiveness of the proposed approach. Especially, their approach also leads to a performance improvement in a cross- platform setting, showing the strong transferability of the proposed universal SRL method.

In previous studies, item representations are usually learned in transductive learning setting, where item IDs are pre-given and ID embeddings are learned as item representations. Such a way largely limits the transferability of item representations.

To tackle the issue of transferability, author's solution is to learn transferable item representations based on the associated item text, which describes the item characteristics in the form of natural language, which provides a general data form to bridge the semantic gap across different tasks or domains. Based on this idea, author first utilizes the pre-trained language model (PLMs), parametric whitening and mixture-of-experts (MoE) enhanced adaptor to transform the text semantics into a universal form suited to the recommendation tasks.

- **Textual Item Encoding via Pre-trained Language Model.**
  Use BERT model to learn universal text representations for representing items. Given an item i and its corresponding text t_i , author firstly concatenate (1) a special symbol [CLS], (2) the words of item text $\{w_1, w_2, \dots, w_c\}$, in order and derive the input sequence for BERT. Then feed the concatenated sequence into the BERT model,

  $$\boldsymbol{x_i} = \mathrm{BERT}([[\mathrm{CLS}];\ w_1, w_2, \dots, w_c])$$

  where $\boldsymbol{x_i} \in R^{dW}$ is the final hidden vector corresponding to the first input token ([CLS]), and "[;]" denotes the concatenation operation.

  BERT (Bidirectional Encoder Representation from Transformers) model is a pre-trained semantic representations model, which utilize masked language model (MLM), and it can produce deep and bidirectional semantic representation. In NLP domain, BERT has achieved tremendous success, so the authors use the BERT model to learn the universal text representations.

  Recent years, XLNet shot up to fame after it beat BERT in roughly 20 NLP tasks, sometimes with quite substantial margins. We will try to use XLNet to replace the BERT model used by the authors and explore whether the new word embedding model will improve the results.
- **Semantic Transformation via Parametric Whitening.**
  The authors consider that though they can obtain semantic representations from BERT, these representations are not directly suited for the recommendation tasks. Existing studies have found that BERT induces a non-smooth anisotropic semantic space for general texts. To address this problem, author inspired by recent works on whitening-based methods.

In general, when we compare the similarity of two sentences, we will calculate their sentence vector and compare their angle cosine, but using cosine similarity as the similarity measure is built on the foundation of standard orthogonal basis. If basis vector is different, then the meaning of the value in each vector might be different. Considering that the semantic representations (sentence vector) produced by BERT might be at different coordinate system, so their meaning might be different. The solution is to normalize each vector to a standard orthogonal basis, and transform the anisotropic semantic space to isotropic semantic space, so BERT-whitening method is introduced.

The idea of BERT-whitening is after obtaining the sentence vector of each sentence, apply whitening (also known as PCA) to these matrix, so that each dimension holds zero mean and covariance matrix is the identity matrix, then maintain k principe component.

The paper conduct a simple linear transformation to transform original BERT representations for deriving isotropic semantic representations. Different from the original whitening methods with pre-set mean and variance, they incorporate learnable parameters in the whitening transformation for better generalizability on unseen domains. Formally,

$$\widehat{\boldsymbol{x}_i} = (\boldsymbol{x}_i - \mathrm{b}) * \mathrm{W}_1$$

where $\boldsymbol{b} \in R^{dW}$ and $\mathrm{W}_1 \in R^{dW \times \mathrm{dV}}$ are learnable parameters, and $\widehat{\boldsymbol{x}_i} \in R^{dW}$ is the transformed representation. In this way, the anisotropy issue of learned representations can be alleviated, which is useful to learn universal semantic representations.

In this paper, author only utilize the linear method. To achieve the better performance, we will introduce complex non-linear architectures, such as flow-based generative models in this part.

- **Domain Fusion and Adaptation via MoE-enhanced Adaptor.**

  With the above whitening transformation, our model can learn more isotropic semantic representations. In order to learn universal item representations, another important issue is how to transfer and fuse information across domains, since there is usually a large semantic gap between different domains. This paper propose to learned multiple whitening embeddings for an item, and utilize an adaptive combination of these embeddings as the universal item representations.

To implement their idea, they employ the mixture-of-expert (MoE) architecture for learning more generalizable item representations and construct the MoE-enhanced adaptor based on a parameterized router, which brings the follwoing three advantages. First, the representation of a single item is enhanced by learning multiple whitening transformations. Second, we no longer require a direct semantic mapping across domains. Third, more flexible to fine-tuning when adapting to new domains.

Since different domains usually correspond to varying user behavioral patterns, it may not work well to simply mix interaction sequences from multiple domains for pre-training. This paper propose one solution is to introduce two kinds of contrastive learning tasks, which can further enhance the fusion and adaptation of different domains in deriving the item representations. Firstly, paper adopt a widely used self-attentive architecture to construct the sequential patterns based on the learned universal textual item representations instead of item IDs.

Then this paper design the following sequence-item and sequence-sequence contrastive tasks to alleviate the seesaw phenomenon and capture their semantic correlation in the pre-training stage.

- **Sequence-item contrastive task**
  his task aim to aims to capture the intrinsic correlation between sequential contexts (i.e., the observed subsequence) and potential next items in an interaction sequence. For a given sequence, They adopt across-domain items as negatives. Such a way can enhance both the semantic fusion and adaptation across domains.
- **Sequence-sequence contrastive task**
  Besides Item-leve; pre-training task, they further propose a sequence-level pretraining task, by conducting the contrastive learning among multidomain interaction sequences. The object is to discriminate the representations of augmented sequences from multi-domain sequences. They propose two kinds of augmentation strategies: Item drop and Word drop.

At the pre-training stage, paper propose a mixed Opti mization object which jointly optimize the proposed sequence-item contrastive loss and sequence-sequence contrastive loss. Then the pre-trained model is to be fine-tuned for adapting to new domains.

To evaluate the performance of the proposed approach, the author conduct experiments in both cross-domain setting and cross-platform setting, and compare the proposed approach with some baseline methods like SASRec, BERT4Rec, FDSA and CCDR. Besides, to evaluate the performance of the next item prediction task, the author adopt two widelyy used metrics Recall@N and NDCG@N, where N is set to 10 and 50. For each user interaction sequence, the last item is used as the test data, the item before the last one is used as the validation data, and the remaining interaction records are used for training. The paper rank the ground-truth item of each sequence among all the other items for evaluation on test set, and finally report the average score of all test users. As a result, the proposed method $UniSRec_{t+ID}$ achieves the best performance in almost all the cases.

In our implementation, we plan to build a model based on UniSRec by coding, and well follow the settings of the experiments in this paper and evaluate the model by comparing the performance of our implementation with reserchers implementation and other baseline models. Well also conduct some analysis if there is a discrepancy in the exprimental results. In particular, we are very interested in the effect that model structure changes may have on the performance efficiency of the proposed method. In the experimental results, the author pays more attention to the accuracy brought by the method and ignores the comparison of inference speed and efficiency, so we will compare the improved model with the original model and other baseline models for a more comprehensive performance and efficiency comparison , in order to evaluate the model more reasonably.

We plan to conduct experiments in both cross-domain setting and cross-platform setting, and the dataset settings are as follows:

1. Pre-trained datasets: we select five categories from Amazon review datasets 1, “Grocery and Gourmet Food”, “Home and Kitchen”, “CDs and Vinyl”, “Kindle Store” and “Movies and TV”, as the source domain datasets for pre-training.
2. Cross-domain datasets: we select another five categories from Amazon review datasets 1, “Prime Pantry”, “Industrial and Scientific”, “Musical Instruments”, “Arts, Crafts and Sewing” and “Office Products”, as target domain datasets to evaluate the proposed approach in cross-domain setting
3. Cross-platform datasets: we also select a dataset from different platforms to evaluate the pre-trained universal sequence representation model in a cross-platform setting. Online Retail2 contains transactions occurring between 01/12/2010 and 09/12/2011 from a UK-based online retail platform, which does not contain shared users or items with the Amazon platform.

REFERENCES

[1] Yupeng Hou, et al. Towards Universal Sequence Representation Learning for Recommender Systems. 2022 Knowledge Discovery And Data Mining(KDD)

[2] Jianmo Ni, Jiacheng Li, and Julian J. McAuley. 2019. Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects. In EMNLP.

[3] Junjie Huang, Duyu Tang, Wanjun Zhong, Shuai Lu, Linjun Shou, Ming Gong, Daxin Jiang, and Nan Duan. 2021. WhiteningBERT: An Easy Unsupervised Sentence Embedding Approach. In Findings of EMNLP.

[4] Jianlin Su, Jiarun Cao, Weijie Liu, and Yangyiwen Ou. 2021. Whitening Sentence Representations for Better Semantics and Faster Retrieval. arXiv preprint arXiv:2103.15316 (2021).

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In NeurIPS.

This project is done solely within the course but not other scopes.