Modeling High-order Interactions across Multi-interests for Micro-video Recommendation (Student Abstract)

Dong Yao[1], Shengyu Zhang[1], Zhou Zhao[1],
Wenyan Fan[1], Jieming Zhu[2], Xiuqiang He[2], Fei Wu[1]

[1]Zhejiang University
[2]Huawei
{yaodongai, sy_zhang, zhaozhou, wenyan.17, wufei}@zju.edu.cn
jmzhu@ieee.org, hexiuqiang1@huawei.com

Abstract

Personalized recommendation system has become pervasive in various video platform. Many effective methods have been proposed, but most of them didn't capture the user's multi-level interest trait and dependencies between their viewed micro-videos well. To solve these problems, we propose a Self-over-Co Attention module to enhance user's interest representation. In particular, we first use co-attention to model correlation patterns across different levels and then use self-attention to model correlation patterns within a specific level. Experimental results on filtered public datasets verify that our presented module is useful.

Introduction

With the emergence of various micro-videos platforms, the amount of micro-videos and users are growing exponentially. Simultaneously, it also becomes increasingly difficult for users to watch micro-videos that they are interested in. Given this, an intelligent personalized recommendation system is crucial. However, building a personalized recommendation system is a big challenge, owing to the following two reasons:1) **Multi-level interest**. There are different interaction behavior between users and micro-videos. In our dataset, three interaction types are included: "click", "like", and "follow", representing user's different levels of preference. 2) **Correlation and dependencies**. Various micro-videos interacted by one user may have inherent correlation from each, and the pattern of such correlation can vary within and across levels of interest.

In recent years, many effective models have been applied in the personalized recommendation system, such as collaborative filtering based models (Xin 2016), content-based systems (Zhongqi 2014), and hybrid methods (Ferracani A.). Despite the promising performance achieved by such methods, most of them neglect the nature of the user's multi-level interests. To this end, some researchers propose to model the user's multi-level interest and construct level-specific embedding in a pre-processed manner (Yongqi Li 2019). However, they don't consider the correlation of micro-videos. To explicitly capture such correlation, we propose the Self-over-Co Attention module, dubbed as *SoC*

Attention. The overall architecture is displayed in Figure 1. Specifically, SoC Attention first models the correlation patterns across levels using co-attention and investigates the level-specific correlation using the self-attention mechanism. Compared with existing methods that directly learn the level related representation by back-propagation such as ALPINE (Yongqi Li 2019), SoC Attention can quickly learn the level related representation of users beyond the training set without retraining. We construct a model named SCAA based on the proposed SoC Attention module and the ALPINE-Base model which is a modified version of ALPINE without its original multi-level interest module.

Our Method

In this section, we will illustrate the two building blocks of SoC Attention in detail:

Co-attention Layer

Expressly, for a user, we set a set of micro-videos' features as $\mathbf{U}_l$, in which the user likes every video. Analogously, we develop a set of micro-videos' features as $\mathbf{U}_f$, in which the user follows every video. Firstly, we feed $\mathbf{U}_l$ into three full connected layers that each of them has different parameters,

$$\mathbf{Q}_l = \mathbf{W}_{ql}\mathbf{U}_l, \mathbf{K}_l = (\mathbf{W}_{kl}\mathbf{U}_l)^T, \mathbf{V}_l = \mathbf{W}_{vl}\mathbf{U}_l$$

Where $\mathbf{Q}_l$, $\mathbf{K}_l$, $\mathbf{V}_l$ are query, key, value matrix for liked micro-videos. Likewise, We feed $\mathbf{U}_f$ into three various full connected layers,

$$\mathbf{Q}_f = \mathbf{W}_{qf}\mathbf{U}_f, \mathbf{K}_f = (\mathbf{W}_{kf}\mathbf{U}_f)^T, \mathbf{V}_f = \mathbf{W}_{vf}\mathbf{U}_f$$

where $\mathbf{U}_l \in \mathbb{R}^{m \times d}$, $\mathbf{U}_f \in \mathbb{R}^{n \times d}$, and $\mathbf{Q}_f$, $\mathbf{K}_f$, $\mathbf{V}_f$ are query, key, value matrix for followed micro-videos. We set **softmax** funuction as $g(\cdot)$. Then, the enhanced "like" and "follow" visual features $\mathbf{U}_l^e$ are computed as follows,

$$\mathbf{U}_l^e = g(\mathbf{Q}_l\mathbf{K}_f)\mathbf{V}_f + \mathbf{U}_l, \mathbf{U}_f^e = g(\mathbf{Q}_f\mathbf{K}_l)\mathbf{V}_l + \mathbf{U}_f$$

Self-attention Layer

Thereafter, we feed $\mathbf{U}_l^e$ into a group of full-connected layers,

$$\mathbf{Q}_l^e = \mathbf{W}_{ql}^e\mathbf{U}_l^e, \mathbf{K}_l^e = (\mathbf{W}_{kl}^e\mathbf{U}_l^e)^T, \mathbf{V}_l^e = \mathbf{W}_{vl}^e\mathbf{U}_l^e$$

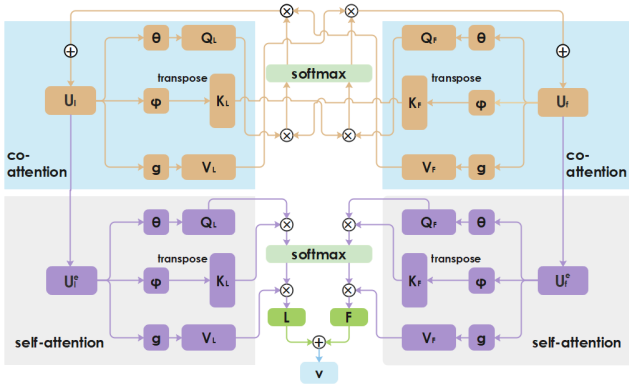Where $\mathbf{Q}_l^e$, $\mathbf{K}_l^e$, $\mathbf{V}_l^e$ denote enhanced query, key, value ma-

Figure 1: SoC Attention module

trix for liked micro-videos. For the $\mathbf{U}_f^e$,

$$\mathbf{Q}_f^e = \mathbf{W}_{qf}^e \mathbf{U}_f^e, \mathbf{K}_f^e = \left(\mathbf{W}_{kf}^e \mathbf{U}_f^e\right)^T, \mathbf{V}_f^e = \mathbf{W}_{vf}^e \mathbf{U}_f^e$$

where $\mathbf{U}_l^e \in \mathbb{R}^{m \times d}$, $\mathbf{U}_f^e \in \mathbb{R}^{n \times d}$, and $\mathbf{Q}_l^e$, $\mathbf{K}_l^e$, $\mathbf{V}_l^e$ denote enhanced query, key, value matrix for followed micro-videos. Then, we get the "like" interest matrix $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_m]^T$ and the "follow" interest matrix $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n]^T$ shown in the following equations,

$$\mathbf{L} = g(\mathbf{Q}_l \mathbf{K}_l)\mathbf{V}_l, \mathbf{F} = g(\mathbf{Q}_f \mathbf{K}_f)\mathbf{V}_f$$

Eventually, we get final improved interest representation $\mathbf{v}$,

$$\mathbf{l} = \frac{1}{m}\sum_{i=1}^m \mathbf{l}_i, \mathbf{f} = \frac{1}{n}\sum_{i=1}^n \mathbf{f}_i, \mathbf{v} = \frac{m}{m+n}\mathbf{l} + \frac{n}{m+n}\mathbf{f}$$

Experiment

Dataset and Evaluation Metric

We use the dataset released by the Kuaishou Competition[1]. Then, we divide it into two parts: i) 80% of the data as the training set; ii) 20% of the data as the testing set. We select Area Under Curve(AUC), R@50, P@50, F@50 as evaluation metric. For a recommendation list computed based on the click probability, R@50, P@50, separately means the recall value, precision of the top50 items, and F@50 is the harmonic average of corrsponding precision and recall.

Comparative Experiment and Ablation Study

To testtify the effectiveness of SCAA, we compare it with: i) ATRank (Chang Zhou 2018) ii) NCF (Xiangnan He 2017) iii) THACIL (Xusong Chen 2018). The results of the aforementioned models are shown in Table 1. It shows that SCAA doesn't achieve better performance compared with ALPINE.

Methods	ATRank	NCF	THACIL	ALPINE	SCAA
AUC	0.722	0.724	0.727	**0.737**	**0.737**

Table 1: Performance of SCAA, ALPINE and other models

[1]https://anonymous1240.wixsite.com/alpine

Methods	AUC	R@50	P@50	F@50
ALPINE	0.696	0.383	0.355	0.368
SCAA_cs	0.701	0.383	0.355	0.368
SCAA_s	0.705	0.385	0.359	0.371
SCAA	**0.712**	**0.390**	**0.364**	**0.376**

Table 2: Performance of the original ALPINE, SCAA and several variants on the filtered dataset

Actually, most of the users in the dataset don't have "like" and "follow" interaction behavior. Obviously, SoC Attention doesn't work on these users. Therefore, we filter these users and retrain models. The result is showed in Table 2. It is worth mentioning that SCAA without SoC Attention is our baseline model. The improvement of SCAA over ALPINE is 2.3%, reflecting that the effectiveness of SoC Attention on gaining user's interest representation is distinct. We design ablation experiments to study how each component in SoC Attention contributes to the final performance: **SCAA_cs:** We removed the Co-attention part and Self-attention part from SCAA; **SCAA_s:** We eliminated the Self-attention part from SCAA. The results are demonstrated in Table 2. By analyzing Table 2, we can conclude:

Compared with SCAA, SCAA_s decreases by 1% in terms of AUC and other indicators all decline. This revealed that the Self-attention part is useful in capturing the dependency of micro-videos of the same interaction behavior.

It can be seen that SCAA_s surpasses SCAA_cs by 0.6% in terms of AUC. Other indicators have a different degree of improvement. This indicates the Co-attention part's effectiveness in catching the correlation of micro-videos of the various interaction behavior.

References

Xin Dong; Lei Yu; Zhonghuo Wu; Yuxia Sun; Lingfeng Yuan; Fangxi Zhang. 2016. A Hybrid Collaborative Filtering Model with Deep Structure for Recommender Systems. In *AAAI*.

Zhongqi Lu; Zhicheng Dou; Jianxun Lian; Xing Xie; and Qiang Yang. 2014. Content-Based Collaborative Filtering for News Topic Recommendation. In *AAAI*.

Ferracani, A.; Pezzatini, D.; Bertini, M.; DelBimbo, A. Item-based video recommendation:An hybrid approach considering human factors. In *ACM*.

Yongqi Li; Meng Liu; Jianhua Yin; Chaoran Cui; XinShunXu; Liqiang Nie. Routing Micro-videos via A Temporal Graph-guided Recommendation System. In *ACM*.

Chang Zhou; Jinze Bai; Junshuai Song; Xiaofei Liu; Zhengchao Zhao. An attention-based user behavior modeling framework for recommendation. In *AAAI*.

Xiangnan He; Lizi Liao; Hanwang Zhang; Liqiang Nie; Xia Hu; Tat-SengChua. Neural collaborative filtering. In *ACM*.

Xusong Chen; Dong Liu; Zheng-Jun Zha; Wengang Zhou; Zhiwei Xiong; YanLi. Temporal hierarchical attention at category and item-level for micro-video click-through prediction. In *ACM*.