

利用GraphTR模型，让每条看一看视频推荐都“投你所好” | CIKM-2020

原创 谢若冰 微信AI 12月2日

收录于话题

#微信看一看背后的技术架构

8个



本文基于CIKM-2020论文《Graph Neural Network for Tag Ranking in Tag-enhanced Video Recommendation》，由腾讯微信和清华大学合作完成。

导语

标签增强的视频推荐系统（tag-enhanced video recommendation system）是一种较新颖的视频推荐形态。在标签增强的视频推荐系统中，每个视频下面会附带着一些标签（tag），这些标签能够强调视频中不同粒度的实体和兴趣点。这个系统中标签排序算法的目标是针对不同用户产生个性化的标签，从而使得标签相关的指标（如标签点击等）和视频相关的指标（如观看视频时长等）均得到提升。

好的个性化标签不仅能够帮助用户理解当前视频的内容，快速找到自己的兴趣点，而且能够吸引用户点击标签，进入相应的标签专属视频流中，然后观看更多标签相关的视频。然而，目前很少有针对标签增强的视频推荐系统设计的标签排序算法，同时考虑用户在标签和视频上的行为。

于是在这个工作中，我们提出一个新的基于异质图神经网络的GraphTR模型，构建了一张融合视频、标签、用户、媒体的大图。我们在图聚合中融合了Transformer、GraphSAGE和FM特征聚合器，同时针对我们的任务设计了一种基于邻居相似度的loss，从用户丰富的在视频上的行为信息中，捕捉用户对于标签的偏好信息。在实验中，我们进行了离线和在线实验，也进行了消融实

验，证明了GraphTR模型的有效性和鲁棒性。目前，GraphTR已被部署于看一看标签增强的视频推荐系统，影响千万用户。

一、模型背景与简介

随着信息通信技术的迅猛发展，视频已经成为用户获取信息和娱乐的主要方式之一，个性化视频推荐系统正在影响我们的生活。和文本、图像不一样的是，视频往往蕴含着更多的信息，这些信息很难在封面图和标题中完整展示给用户。于是，为了方便用户快速理解视频内容，找到自己的兴趣点，标签增强的视频推荐系统被提出并广泛使用。图1给出了看一看标签增强的视频流示意图：

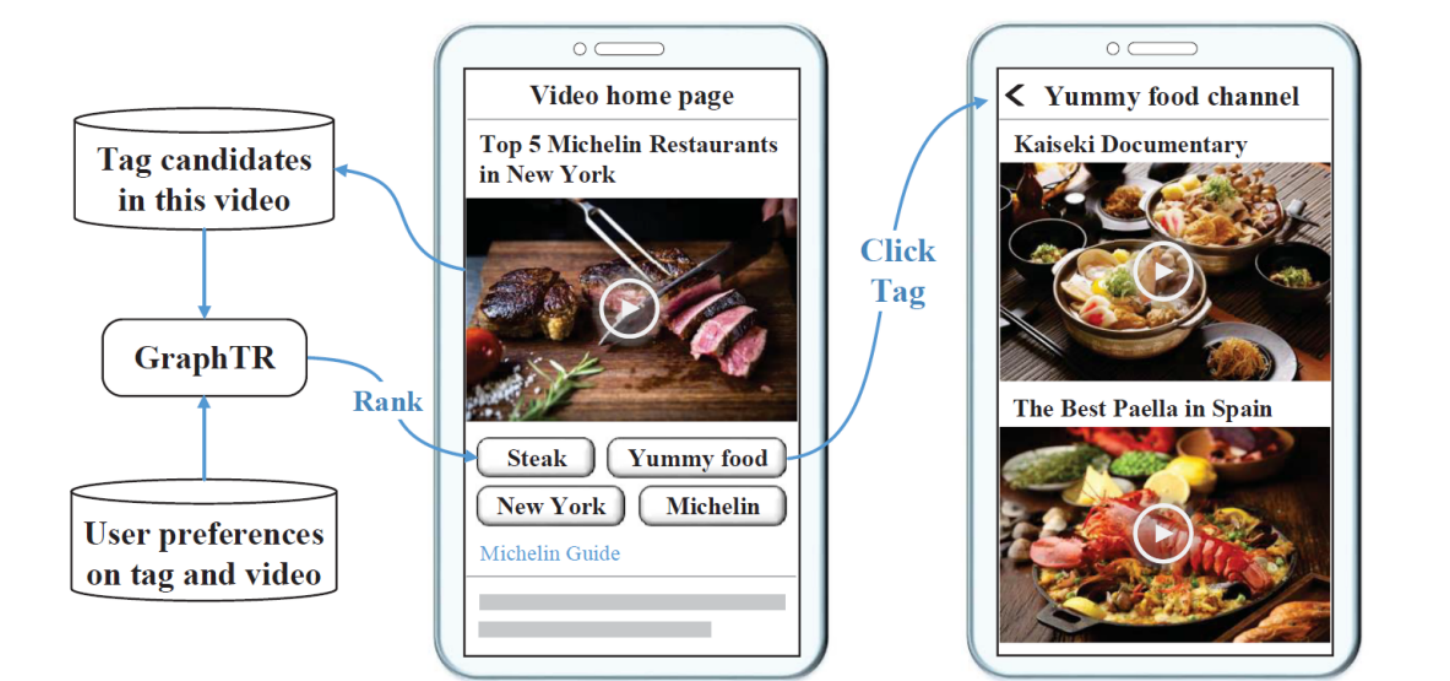


图1：看一看标签增强的视频推荐系统

如图1所示，部分视频底下会附带一些视频标签（如图中的Steak，Yummy food等）。这些标签往往是经过生产者或者编辑预先标注的，用以从不同角度和不同粒度描述视频内容，每个视频拥有几个到十几个标签不等。标签排序算法会根据不同用户的特征，从这些标签池中选择合适的个性化标签展示给用户。当用户点击一个标签时（例如点击Yummy food标签），他/她会进入对应的标签频道，频道内仅包含点击的标签相关的视频（如各种美食视频），进行沉浸式的连续观看。我们认为一个好的个性化标签需要做到以下三点：

- (1) 根据用户的喜好，显式地通过标签展示用户感兴趣的视频内容；
- (2) 吸引用户点击标签；
- (3) 引导用户在标签频道中进行深度阅读。

总体上，标签排序对于标签增强的视频推荐是很重要的。

在这个工作中，我们关注标签增强的视频推荐系统，希望标签排序能够同时提升标签相关的指标（如标签点击等）和视频相关的指标（如观看视频时长等）。标签排序主要挑战来源于以下两点：

- (1) 用户在标签上的行为相对视频行为来说尤为稀疏，影响模型的监督训练；
- (2) 这个工作希望同时提升视频和标签相关的指标，但是这两者之间可能存在着gap。传统的标签排序算法基本只考虑了用户在标签上的行为以及在标签上的指标，没有充分考虑视频相关的信息。

因此，我们提出一种新的Graph neural network based tag ranking (GraphTR)模型。具体地，GraphTR主要包括三个模块：

- (1) **异质图构建模块**，构建一个包括了视频、标签、媒体、用户以及他们之间交互的异质大图。这些异质的信息交互连接了用户在用户不同兴趣维度上相近的异质节点。
- (2) **网络表示学习模块**，使用一个新的HFIN图神经网络模型，融合特征域维度的Transformer，GraphSAGE和FM特征抽取器得到节点表示。我们使用了一个基于邻居相似度的loss，将用户的多样偏好学到节点表示中。
- (3) **线上标签排序模块**，使用学到的节点表示进行线上标签排序。GraphTR成功地从多种异质交互信息（特别是用户丰富的对视频的行为）中学习到了用户对于标签的偏好，从而极大地缓解了用户标签行为稀疏带来的问题，构建用户对视频的偏好和对标签的偏好之间的桥梁。

在实验中，我们在看一看数据集上进行了充分的线上线下实验，**证明了GraphTR模型在标签增强的视频推荐系统中的有效性和鲁棒性。模型在视频和标签相关指标上都获得了提升。**这个工作的

主要贡献点如下：

- 1、我们首次系统研究了标签增强的视频推荐这个新颖的推荐形式，同时提出了一个新的GraphTR框架。据我们所知，这是第一项在标签增强的视频推荐系统中的标签排序任务里引入异质图神经网络的工作。
- 2、我们提出了一种新的图神经网络模型，融合Transformer，GraphSAGE和FM特征抽取器得到节点表示。我们还提出了一个基于邻居相似度的loss，将用户的多样偏好学到节点表示中。
- 3、模型的线上线下效果均得到提升，同时我们也将模型部署到了微信看一看系统上，在真实场景里得到应用。

二、模型结构

如前所述，GraphTR框架主要包括三个模块：（1）异质图构建模块；（2）网络表示学习模块；（3）线上标签排序模块。图2给出了系统的示意图：

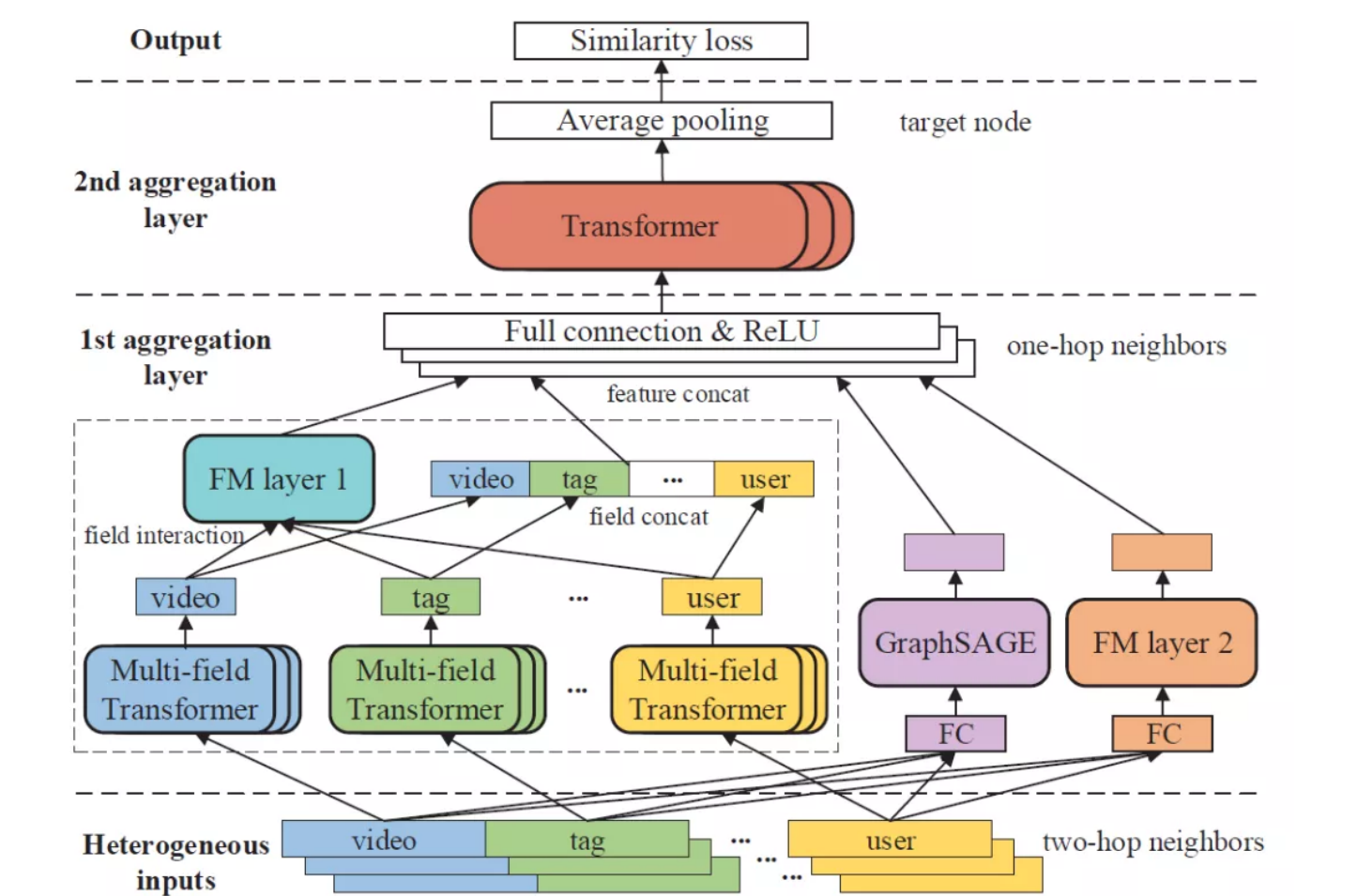


图2：GraphTR模型示意图

首先，在异质图构建模块，我们选择了四种标签增强的视频推荐系统中的重要元素——视频、标签、用户、媒体（视频来源）作为节点，同时选择了这些节点之间的五种重要交互作为边，这些边就是用户多样化偏好的来源。我们的五种交互关系边包括：

- (1) 视频——视频，从用户的视频观看session行为中构建；
- (2) 视频——用户，从用户的视频观看行为中构建；
- (3) 视频——标签，从视频拥有的标签信息中构建；
- (4) 视频——媒体，从视频的所属视频来源中构建；
- (5) 标签——标签，从标签在视频中的的共现信息中构建。

根据具体任务和推荐系统的不同，我们还可以使用更多的节点和边的信息，辅助推荐效果。

然后，我们提出了一个Heterogeneous field interaction network (HFIN)图神经网络，融合Transformer，GraphSAGE和FM特征抽取器得到节点表示。具体地，每一个节点都由视频、标签、用户、媒体特征域组成。我们在每一层聚合的时候基于不同特征域分别进行聚合。其中，Field-level Transformer能够分特征域抓住邻居节点之间的交互信息，GraphSAGE进行了直观整体的邻居级特征聚合，而FM则基于邻居聚合后的特征域的进行特征交互和聚合。我们有：

$$\mathbf{h}_{\text{Trans}} = \text{concat}(\mathbf{h}_{\text{FM1}}, \hat{\mathbf{h}}_v, \hat{\mathbf{h}}_t, \hat{\mathbf{h}}_m, \hat{\mathbf{h}}_u)$$

$$\mathbf{h}_{\text{Graph}} = \text{ReLU}(\mathbf{W}^G \cdot (\sum_{i=1}^n \mathbf{f}_i + \mathbf{f}_s))$$

$$\mathbf{h}_{\text{FM2}} = \sum_{i=1}^4 \sum_{j=i+1}^4 \mathbf{h}'_i \odot \mathbf{h}'_j, \mathbf{h}' = \{\mathbf{h}'_v, \mathbf{h}'_t, \mathbf{h}'_m, \mathbf{h}'_u\}$$

$$\mathbf{h} = \text{concat}(\mathbf{h}_{\text{Trans}}, \mathbf{h}_{\text{Graph}}, \mathbf{h}_{\text{FM2}})$$

这种邻居节点聚合方式强调了不同特征域之间和不同邻居之间的交互，能够促进用户多样化偏好之间的信息交互，从而使得用户的标签偏好能够更好地被学习。

我们通过HFIN得到节点表示之后，使用了一个基于邻居相似度的loss进行学习。由于用户在标签上的行为远比在视频上的行为稀疏（因为视频是标签增强的视频推荐系统中的核心推荐客体），仅仅使用用户在标签上的点击进行监督学习，往往难以使模型学习充分。因此，我们设计了一种基于邻居相似度的新的loss，希望能够通过用户的其它行为（特别是用户在视频上的行为），学习到用户在标签上的偏好。

具体地，我们认为在异质图上相邻的邻居需要有相似的表示（可以看作是一种步长为1的随机游走算法，步长为1缓解了传播噪声的影响）。这种loss的合理性在于，我们构建异质图时使用了多种视频、标签、用户、媒体之间的交互信息，这些交互信息连接的节点往往在用户某种偏好维度（例如社群、视频来源、类别等）上拥有相似点。

例如，视频session能够将在用户观看序列上相似视频（以及视频相关的标签、媒体号、用户等）联系到一起，反映了用户在视频观看序列上的偏好；媒体能够将一个媒体发布的视频联系到一起，反映了用户在视频来源上的偏好；用户也能将相似视频联系到一起，反映了同一类用户群体爱看的视频内容。通过邻居相似度的loss和异质图上的多步路径，在不同用户兴趣维度上相似的视频和标签，会被学习到拥有相似的向量。

$$J = - \sum_{o_i} \sum_{o_k \in N_i} \sum_{o_j \notin N_i} (\log(\sigma(\mathbf{o}_i^T \mathbf{o}_k)) + \log(1 - \sigma(\mathbf{o}_i^T \mathbf{o}_j)))$$

最后，在线上标签排序模块，我们通过用户历史视频观看序列中视频的标签，获得用户的代表性标签，加权组成用户表示（可以看做用户在标签上的偏好表示）。然后，使用此用户表示挖掘相

似的标签候选，基于通过余弦相似度对视频标签池中的标签进行排序，最终展示给用户。

三、实验结果

我们基于看一看中的标签增强视频推荐系统，构造了离线数据集进行实验。在标签点击预测任务中，GraphTR获得了最佳结果：

Model	HIT@2	MRR
Random-based	0.420	0.210
Popularity-based	0.467	0.237
TF-IDF [2]	0.456	0.228
BERT [4]	0.532	0.293
FM [19]	0.476	0.242
DeepFM [8]	0.492	0.259
AutoInt [24]	0.511	0.275
GraphTR (DeepWalk)	0.593	0.330
GraphTR (GraphSAGE)	0.613	0.346
GraphTR (HGAT)	0.623	0.354
GraphTR (HFIN)	0.678	0.384

图3：离线标签点击预测实验结果

另外，我们还在看一看线上系统上进行了A/B实验，在多个标签相关和标签视频流内视频相关指标上，GraphTR均获得了最好的效果。具体结果如下：

	Tag click rate	Tag click number	Tag list-wise click rate	Average video watching time	Video views per capita	Page turns per capita
BERT	+4.64%	+2.27%	+3.47%	+4.49%	+1.46%	+1.89%
GraphTR (DeepWalk)	+5.46%	+2.92%	+3.92%	+3.15%	+1.77%	+2.89%
GraphTR (GraphSAGE)	+5.74%	+3.53%	+4.11%	+6.25%	+4.31%	+4.49%
GraphTR (HGAT)	+7.30%	+4.58%	+5.60%	+7.23%	+4.73%	+4.96%
GraphTR (HFIN)	+8.48%	+5.60%	+6.81%	+9.70%	+5.11%	+6.32%

图4：线上A/B实验结果

消融实验结果也证明GraphTR各个模块的有效性。

Ablation version	HIT@2	MRR
GraphTR (HFIN)	0.678	0.384
– GraphSAGE layer	0.671	0.380
– FM layer	0.662	0.371
– 1st transformer layer	0.617	0.341
– 2nd transformer layer	0.597	0.332
– user nodes and edges	0.643	0.359
– media nodes and edges	0.656	0.367

图5：消融实验结果

最后，我们还给出两个case study的例子，展示了标签向量的K近邻标签结果（图6），以及同一个视频不同用户的个性化标签排序（图7），直观说明GraphTR的效果。图6中，和标签【长城】相近的标签，有同是北京世界遗产的【紫荆城】，有粗粒度的【世界遗产】，也有相关的朝代【秦朝】。这里说明了GraphTR学习到的标签，不仅有标签自身的语义信息，也包含着用户在推荐系统中的不同偏好维度的信息。这样的标签向量更适合标签增强视频推荐中的标签排序任务。

Tag	Nearest tags
The Great Wall	Monument; World Cultural Heritage; Forbidden city; Qin Dynasty; Tourism inventory
Michelin star restaurant	Michelin chef; French red wine; Gourmet show; Spain Seafood Risotto; Japanese food
New energy vehicle	Hydrogen powered vehicle; Fuel consumption; Bao-jun; 4-wheel drive; Foreign car

图6：标签向量的K近邻标签结果

图7中也说明，对于同一个视频，GraphTR会基于不同用户的自身兴趣，推荐不同的个性化标签，展示用户感兴趣的视频的不同亮点（如特定明星、篮球运动员、粗粒度的综艺等）。这样的个性化标签能够辅助提升推荐系统效果。

Video title	Shaquille O’Neal performs his Dream shake and N Jia imitates the movement comically .
User1 tags	N Jia; Go Champion! ; Variety show in China
User2 tags	Shaquille O’Neal; Basketball; Variety show
User3 tags	Imitation; Funny moment; Variety show

图7： 同一个视频不同用户的个性化标签排序

四、总结

在这个工作中，我们探索了标签增强的视频推荐系统中的标签排序问题。我们提出了一个GraphTR框架，基于HFIN图神经网络和邻居相似度的loss，在全局异质大图上，通过用户丰富视频行为学习用户对于标签的偏好，缓解用户标签行为稀疏的问题。模型现在已经部署在看一看线上系统，服务广大用户。在未来，我们将在异质图构建、图神经网络模型设计和线上标签排序等模块继续进行探索。

微信AI

不描摹技术的酷炫，不依赖拟人的形态，微信AI是什么？是悄无声息却无处不在，是用技术创造更高效率，是更懂你。

微信AI关注语音识别与合成、自然语言处理、计算机视觉、工业级推荐系统等领域，成果对内应用于微信翻译、微信视频号、微信看一看等业务，对外服务王者荣耀、QQ音乐等产品。