

论文阅读报告

Neural Attentional Rating Regression with Review-level Explanations

简介

评论未必每一个都是有用的，不太有用的评论会损害模型的性能，同时对用户的参考意义也不大，有一些模型会利用评论来提高推荐系统的性能，但是这些模型很少将评论的有用性考虑进来，并且也没有考虑到评论的有用性可能会影响到推荐系统的推荐质量，因此该论文引入了注意力机制来探索评论的有用性，并提出了NARRE模型。

目录 CONTENT

1

背景知识

2

模型概述

3

实验结果



PART

01



背景知识

PART 1

● LFM(Latent Factor Model)

LFM是一种主要基于矩阵分解技术的算法，LFM算法一个重要的应用就是预测用户 u 对一个物品 i 的评分 $R_{u,i}$ ，具体如下：

$$\hat{R}_{u,i} = q_u p_i^T + b_u + b_i + \mu$$

在该等式中， μ 表示全局平均评分， b_u 表示用户偏差(user bias)， b_i 表示物品偏差， q_u 表示用户偏好，而 p_i 则表示物品的特征



PART

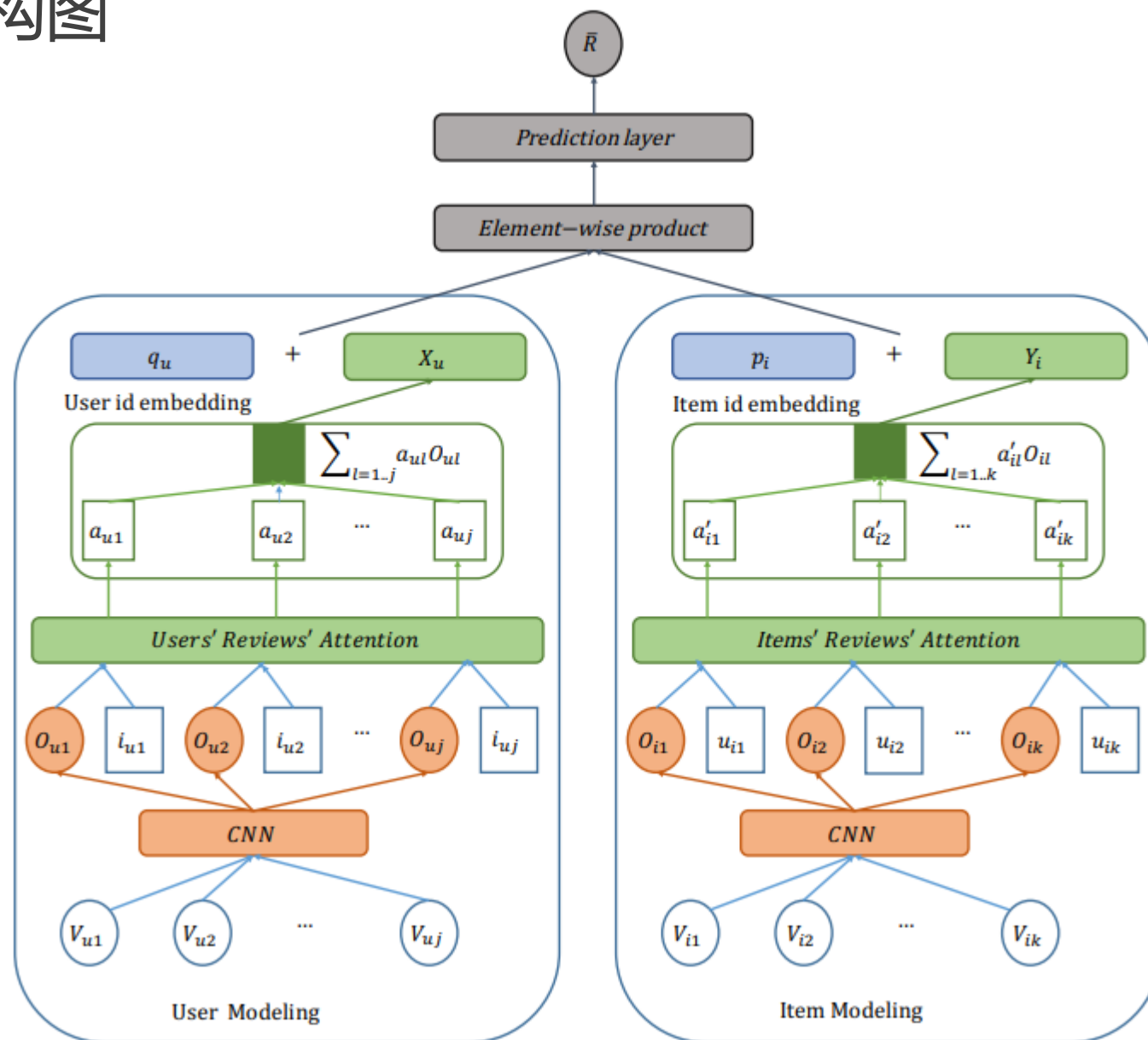
02



模型概述

PART 2

● 模型结构图



该模型由两个并行神经网络组成，一个用于用户建模（ Net_u ），一个用于物品建模(Net_i)，在这两个网络的上方增加了一个预测层，让用户和物品的隐藏潜在因素相互作用，从而计算出模型的最终结果，由于这两个网络仅输入不同，因此下面将重点说明 Net_i 。

在 Net_i 的第一阶段，使用CNN处理项目i的文本评论，每个评论首先被转换成一个单词向量矩阵，表示为 V_{i1} 、 V_{i2}，然后将这些矩阵送入卷积层，输出即为它们的特征向量，记这些特征向量为 O_{i1} 、 O_{i2} O_{ij} 。

由于这些向量在相同的特征空间中，一般的想法是将这些向量累加从而得到物品i的表示：

$$O_i = \sum_{l=1, \dots, k} \frac{1}{k} O_{il}$$

然而在这个等式下，每个评论对物品*i*有相同的贡献，这并不可取，因而在模型中引入了注意力机制，这有助于以远程监督的方式学习每个评论的权重。这里应用了两层网络来计算注意力得分 a_{il} ，输入包含了评论的特征向量 O_{il} 和编写它的用户的id嵌入 u_{il} ，之所以引入id嵌入是为了对用户的质量进行建模，这有助于识别总是写不太有用的评论的用户，注意力网络的定义如下：

$$a_{il}^* = h^T \text{ReLU}(W_O O_{il} + W_u u_{il} + b_1) + b_2$$

其中 W_O ， W_u ， b_1 ， h ， b_2 都是模型的参数。

评论的最终权重是通过使用softmax函数对前面得到的注意力得分 a_{il} 进行归一化来获得的：

$$a_{il} = \frac{\exp(a_{il}^*)}{\sum_{l=0}^k \exp(a_{il}^*)}$$

在得到每个评论的注意力权重后，物品 i 的特征向量计算方法如下：

$$O_i = \sum_{l=1, \dots, k} a_{il} O_{il}$$

输出的特征向量是 k 维的向量，将其发送到具有权重矩阵 W_0 和偏移量 b_0 的全连接层，该全连接层计算物品 i 的最终表示：

$$Y_i = W_0 O_i + b_0$$

然后就是预测层，提出了一种用于预测评分的神经网络形式的LFM，具体来说，首先将用户和物品的潜在因素映射到公共的隐藏空间，通过引入从评论中学习到的隐表示，用户 u 和物品 i 之间的交互（interaction）被建模成：

$$h_0 = (q_u + X_u) \odot (p_i + Y_i)$$

其中 q_u 表示基于评分的用户偏好， p_i 表示基于评分的物品特征，而 X_u 和 Y_i 是之前两个神经网络获得的用户偏好和项目特征，但这里的id嵌入与之前引入注意力机制时用到的id嵌入不同，因为用户的质量和他们的偏好是不同类型的对象，具有不同的特征，如果在同一个向量空间对它们进行建模存在局限性。输出的 h_0 是一个n维向量，需要将其传递到预测层以获得最终的评分

$$\hat{R}_{u,i} = W_1^T h_0 + b_u + b_i + \mu$$

其中 W_1 表示预测层的边缘权重（edge weights）， b_u ， b_i ， μ 分别表示用户偏差、项目偏差和全局偏差，如果将 W_1 固定为1，并且去掉 X_u 和 Y_i ，那么就恢复到了一般的LFM，因此可以在 h_0 和预测层之间添加更多的非线性转换隐藏层，这可以作为未来的一个探索方向。

● 数据集介绍

实验共使用了四个数据集，其中三个来自于Amazon 5-core，分别是Toys_and_Games、Kindle_Store和Movies_and_TV，还有一个数据集来自Yelp Challenge 2017，这是一个由餐厅评级和评论组成的大型数据集，具体信息如下：

Table 1: Statistical details of the datasets.

	Toys_and_Games	Kindle_Store	Movies_and_TV	Yelp_2017
<i>users</i>	19,412	68,223	123,960	199,445
<i>items</i>	11,924	61,935	50,052	119,441
<i>ratings & reviews</i>	167,597	982,619	1,679,533	3.072.129



PART

03



实验结果

PART 3

● 评估指标介绍

实验使用的评估指标是均方根差(RMSE)，RMSE分数越低表示模型性能越好，具体计算方法如下：

$$RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (\hat{R}_{u,i} - R_{u,i})^2}$$

● 实验结果展示

	Toys_and_Games	Kindle_Store	Movies_and_TV	Yelp-2017
PMF	1.3076	0.9914	1.2920	1.3340
NMF	1.0399	0.9023	1.1125	1.2916
SVD++	0.8860	0.7928	1.0447	1.1735
HFT	0.8925	0.7917	1.0291	1.1699
DeepCoNN	0.8890	0.7875	1.0128	1.1642
NARRE	0.8769**	0.7783**	0.9965**	1.1559*

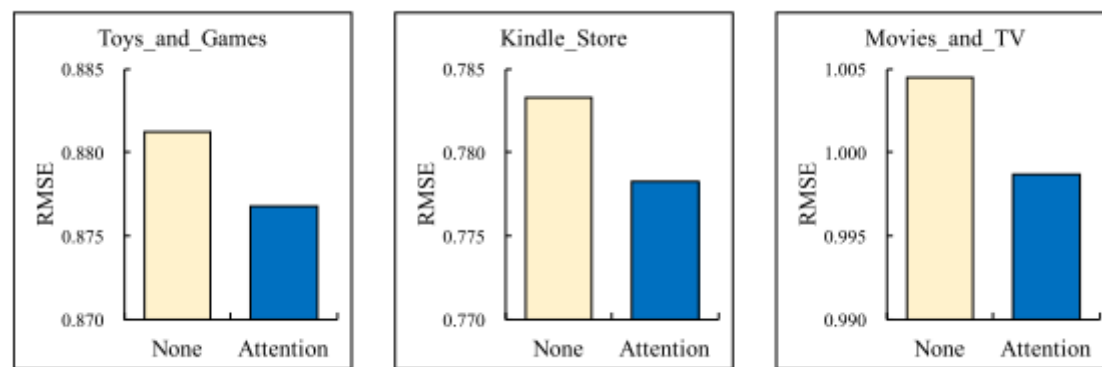
各方法在四个数据集上的表现

● 实验结果展示


Item 1	a ($a_{ij}=0.1932$)	These brushes are great quality for children's art work. They seem to last well and the bristles stay in place very well even with tough use.
	b ($a_{ij}=0.0161$)	I bought it for my daughter as a gift.
Item 2	a ($a_{ij}=0.2143$)	From beginning to end this book is a joy to read. Full of mystery, mayhem, and a bit of magic for good measure. Perfect flow with excellent writing and editing.
	b ($a_{ij}=0.0319$)	I like reading in my spare time, and I think this book is very suitable for me.

NARRE模型选择的高权重和低权重评论的例子（ a_{ij} 表示注意力权重）

● 实验结果展示



有无注意力机制对模型的影响



谢谢观看

T H A N K Y O U