考虑评论质量的自注意力胶囊网络评分预测模型

梁顺攀*①② 刘 伟① 尤殿龙①② 刘泽谦① 张付志①②

①(燕山大学信息科学与工程学院 秦皇岛 066004)

②(燕山大学河北省软件工程重点实验室 秦皇岛 066004)

摘 要：基于评论文档的推荐系统普遍采用卷积神经网络识别评论的语义，但由于卷积神经网络存在“不变性”，即只关注特征是否存在，忽略特征的细节，卷积中的池化操作也会丢失文本中的一些重要信息；另外，使用用户项目交互的全部评论文档作为辅助信息不仅不会提升语义的质量，反而会受到其中低质量评论的影响，导致推荐结果并不准确。针对上述提到的两个问题，该文提出了自注意力胶囊网络评分预测模型(Self-Attention Capsule network Rate prediction, SACR)，模型使用可以保留特征细节的自注意力胶囊网络挖掘评论文档，使用用户和项目的编号信息标记低质量评论，并且将二者的表示相融合用以预测评分。该文还改进了胶囊的挤压函数，从而得到更精确的高层胶囊。实验结果表明，SACR在预测准确性上较一些经典模型及最新模型均有显著的提升。

关键词：推荐系统；胶囊网络；注意力；评论质量；评分预测

中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2021)12-3451-08

DOI: 10.11999/JEIT200932

Self-attention Capsule Network Rate Prediction with Review Quality

LIANG Shunpan①② LIU Wei① YOU Dianlong①②

LIU Zeqian① ZHANG Fuzhi①②

①(*College of Information Science and Engineering, Yanshan University, Qinhuangdao* 066004, *China*)

②(*Key Laboratory for Software Engineering of Hebei Province, Yanshan University, Qinhuangdao* 066004, *China*)

Abstract: Recommendation systems based on reviews generally use convolutional neural networks to identify the semantics. However, due to the “invariance” of convolutional neural networks, that is, they only pay attention to the existence of features and ignore the details of features. The pooling operation will also lose some important information; In addition, using all the reviews as auxiliary information will not only not improve the quality of semantics, but will be affected by the low-quality reviews, this will lead to inaccurate recommendations. In order to solve the two problems mentioned above, this paper proposes a SACR (Self-Attention Capsule network Rate prediction) model. SACR uses a self-attention capsule network that can retain feature details to mine reviews, uses user and item ID to mark low-quality reviews, and merge the two representations to predict the rate. This paper also improves the squeeze function of the capsule, which can obtain more accurate high-level capsules. The experiments show that SACR has a significant improvement in prediction accuracy compared to some classic models and the latest models.

Key words: Recommendation system; Capsule network; Attention; Review quality; Rate prediction

收稿日期：2020-11-02；改回日期：2021-10-25；网络出版：2021-11-10

*通信作者： 梁顺攀 liangshunpan@ysu.edu.cn

基金项目：国家自然科学基金(62072393)， 河北省自然科学基金(G2021203010，F2021203038)

Foundation Items: The National Natural Science Foundation of China (62072393), The Natural Science Foundation of Hebei Province (G2021203010, F2021203038)

1 引言

近年来，许多网站允许用户以评论的形式表达对目标项目的态度。用户的评论中包含了丰富的用户喜好信息以及商品特性信息。使用评论文档作为辅助信息的推荐模型可以缓解评分矩阵的数据稀疏问题。

根据基础结构，将基于评论的推荐模型分为两类，分别是基于主题模型的方法和基于深度模型的方法。基于主题模型的方法[1,2]使用LDA(Latent Dirichlet Allocation)模型[3]抽取评论文本的主题，再根据主题预测评分。由于主题模型会丢失文本词序，并且学习到的是浅层特征，所以这类方法逐渐被基于深度模型的方法所替代。在基于深度模型的推荐算法中，ConvMF(Convolutional Matrix Factorization)模型[4]使用卷积神经网络(Convolutional Neural Networks, CNN)处理项目评论文档；DeepCoNN(Deep Cooperative Neural Networks)模型[5]使用两个CNN并行处理用户和商品评论，通过融合层计算用户对项目的评分；NARRE(Neural Attentional Regression model with Review-level Explanations)模型[6]在DeepCoNN的基础上考虑了评论的质量，使预测精度进一步上升。TARMF(Topical Attention Regularized Matrix Factorization)[7]采用基于注意力的循环神经网络(Recurrent Neural Network, RNN)提取主题信息，可以将词序信息保留，更大程度地理解文本。

CNN存在“不变性”、RNN常会出现梯度消失或者梯度爆炸问题，它们均不能准确地提取评论文本的细粒度特征，影响预测精度。胶囊网络[8]将神经网络中的神经元扩展成“胶囊”向量，胶囊的长度代表某个属性是否存在，胶囊的方向代表属性的具体描述，这些特性使其可以保存细粒度文本语义，在关系提取[9,10]和文本分类[11,12]任务中有效地改善了CNN的缺点。此外，胶囊网络仅通过少量迭代就可以获得较好的结果，避免了RNN的训练缓慢和梯度问题。使用胶囊网络的推荐模型CARP(CApsule network based model for Rating Prediction with user reviews)[13]仅将胶囊网络设置在模型预测层， MIND(Multi-Interest Network with Dynamic routing)[14]忽略了项目文档中隐含的项目特性信息，它们都没有准确地提取出文本的细粒度特征。

在推荐系统中，有些用户对项目要求过高，习惯给出负面评论，甚至对项目恶意评论，这样的评论内容是低质量的。目前已经出现了一些使用数据集中用户对评论的“有用/无用”标记来预测每条评论的有用性的研究[15]，但当数据集中不存在这样的标记时，现有的基于评论文档的推荐算法[1,2,4,5,7,13,14,16]都无法考虑到评论质量的因素，低质量评论混在大量的评论文档中会降低挖掘出语义特征的质量，进而影响预测的精度。

为解决CNN挖掘文本特征不准确以及低质量评论影响结果精度的问题，本文提出考虑评论质量的自注意力胶囊网络评分预测模型SACR。本文模型的主要贡献总结如下：

(1)本文使用自注意力胶囊网络同时挖掘用户和项目评论文档，并且改进了胶囊的挤压函数，可以获取到细粒度级别的用户偏好和项目特性，解决了使用CNN挖掘文本特征不准确的问题。

(2)为了解决低质量评论会对预测精度产生影响的问题，本文模型使用用户和项目的编号信息标记那些质量不高的评论，在模型学习时可以避免其带来的负面影响，从而提升模型的评分预测精度。

(3)将本文模型和基准模型在8个不同规模的现实世界数据集进行实验，结果证明了SACR避免低质量评论影响的能力以及在评分预测上的准确性。

2 模型结构

为充分挖掘用户的细粒度偏好以及项目的细粒度特性，SACR设置了一个双塔结构网络，分别用来处理用户评论文档和项目评论文档，并通过融合层将两个网络的输出和评论质量表示融合，从而预测用户对项目的评分。SACR模型的3层结构分别是：(1)将评论文档嵌入表示，并挖掘文本注意力表示的编码注意力层；(2)挖掘细粒度级别的用户偏好和项目特性的卷积胶囊层；(3)将评论质量表示与上层输出连接、变换后预测评分的融合层。本文提出的SACR模型结构如图1所示。SACR模型中主要用到的符号及其定义如表1所示。由于用户网络和项目网络在前两层的结构相同，所以以下文主要叙述用户网络的详细结构。

2.1 编码注意力层

第1层是编码注意力层，本层首先对评论文档进行词嵌入，然后依据自注意力权重对词嵌入矩阵重新赋权，得到用户或项目文本注意力表示矩阵。以用户i的评论文档为例，模型的输入为用户i对所有项目的评论文本序列：$\boldsymbol{R}_i=[\boldsymbol{R}_{i1},\cdots,\boldsymbol{R}_{ik},\cdots,\boldsymbol{R}_{iL}]$，其中$R_{ik}$是第$k$个单词在字典中的索引，$L$是评论文档的长度。词嵌入模型输出得到词嵌入矩阵$\boldsymbol{X}_i=[\boldsymbol{X}_{i1},\cdots,\boldsymbol{X}_{ik},\cdots,\boldsymbol{X}_{iL}]$，其中$\boldsymbol{X}_{ik}\in\mathbb{R}^{\mathrm{d}}$是第$k$个词向量。词嵌入矩阵并不能表示出用户的表达重点，则需要继续对词嵌入矩阵进行自注意力变换，获取表达重点。

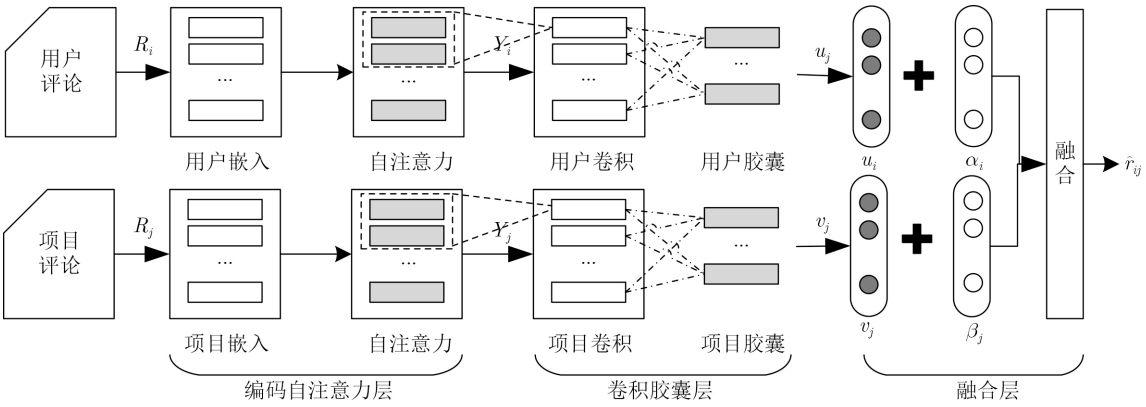


图 1 SACR模型结构

表 1　模型符号定义

符号	定义	符号	定义
N,M	数据集中用户和项目的数量	α_i,β_j	用户i、项目j的编号嵌入
R	用户或项目的评论文档	u_i,v_j	用户i、项目j的胶囊
r_{ij}	评分矩阵中用户i对项目j的评分	C,e^1	输入胶囊的通道数、维度
$\hat{r}_{ij}$	用户i对项目j的预测评分	D,e	输出胶囊的数量、维度
$\boldsymbol{W}$,$\boldsymbol{b}$	模型中的权重矩阵、偏置向量	λ	路由迭代的次数
d	词嵌入维度	F	评分数量

首先将词嵌入矩阵X_i进行3次线性变换：$\boldsymbol{Q}_i=\boldsymbol{X}_i\boldsymbol{W}^Q,\boldsymbol{K}_i=\boldsymbol{X}_i\boldsymbol{W}^K,\boldsymbol{V}_i=\boldsymbol{X}_i\boldsymbol{W}^V$，其中$\boldsymbol{W}^Q,\boldsymbol{W}^K,\boldsymbol{W}^V\in\mathbb{R}^{d\times d}$是转换矩阵，$\boldsymbol{Q}_i,\boldsymbol{K}_i,\boldsymbol{V}_i\in\mathbb{R}^{L\times d}$分别代表注意力函数中的查询、键和值。

计算用户的自注意力表示分为3步，第1步是对$\boldsymbol{Q}_i,\boldsymbol{K}_i$进行相似度计算得到用户在评论文档中的注意力权重$\boldsymbol{W}^{\mathrm{Attn}}\in\mathbb{R}^{L\times L}$，本文采用的相似度计算方式是点积运算。第2步是使用softmax函数将注意力权重归一化。第3步是将注意力权重对相应的$\boldsymbol{V}_i$计算加权和，得到用户i在编码注意力层的输出矩阵$\boldsymbol{Y}_i\in\mathbb{R}^{L\times d}$。$\boldsymbol{Y}_i$即为包含了用户的文本注意力表示矩阵。自注意力计算中用到的公式为

$$\boldsymbol{W}^{\mathrm{Attn}}=\mathrm{softmax}\left(\frac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d}}\right)\tag{1}$$

$$\boldsymbol{Y}_i=\boldsymbol{W}^{\mathrm{Attn}}\boldsymbol{V}_i\tag{2}$$

2.2　卷积胶囊层

第2层是卷积胶囊层。胶囊网络可以挖掘出文本中的细粒度语义，该层使用胶囊网络改进传统的CNN。本层改进了传统胶囊网络[8]的挤压函数，可得到更准确的上层胶囊，进而获取细粒度的用户偏好和项目特性。

卷积胶囊层分为两部分，第1部分是对上层输出进行两次卷积，初始化用户和项目胶囊。以用户i为例，首先选用A个宽度为d，窗口大小为w_1的卷积核初步提取文本特征，第a个卷积核$\boldsymbol{W}^a$对$\boldsymbol{Y}_i$的卷积定义为

$$\boldsymbol{Z}_{ia}=\mathrm{ReLU}(\boldsymbol{Y}_i*\boldsymbol{W}^a+\boldsymbol{b}^a)\tag{3}$$

其中，$*$代表卷积操作，$\boldsymbol{b}^a$是bias向量。将A个卷积核卷积的结果进行拼接，得到用户i的文本卷积映射矩阵$\boldsymbol{Z}_i=[\boldsymbol{Z}_{i1},\boldsymbol{Z}_{i2},\cdots,\boldsymbol{Z}_{iA}]$，$\boldsymbol{Z}_i\in\mathbb{R}^{A\times(L-w_1+1)}$包含用户粗粒度偏好，继续使用1维卷积将$\boldsymbol{Z}_i$初始化成胶囊。选用$B$个宽度为$A$，窗口大小为$w_2$的卷积核再次提取文本特征，将结果进行拼接并变换维度得到初始胶囊组成的2维矩阵：$\boldsymbol{P}_i=[\boldsymbol{p}_1,\cdots,\boldsymbol{p}_k,\cdots,\boldsymbol{p}_{(L-w_1-w_2+2)\times C}]$。$\boldsymbol{p}_{\mathrm{k}}\in\mathbb{R}^{e^1}$，其中$C=B/e^1$代表初始胶囊通道数。

卷积胶囊层的第2部分是将胶囊输入到胶囊网络，进一步挖掘用户偏好和项目特性。首先对低层胶囊特征映射，得到高层胶囊的模糊表示$\boldsymbol{p}_{j|i}$，然后通过高低胶囊的连接权重c_{ij}计算出高层胶囊的准确表示$\boldsymbol{s}_j$，c_{ij}可以通过动态路由[8]过程计算得出，最后将$\boldsymbol{s}_j$通过本文改进的胶囊挤压函数标准化。以上过程公式为

$$\boldsymbol{p}_{j|i}=\boldsymbol{W}_{ij}\cdot\boldsymbol{p}_i\tag{4}$$

$$c_{ij}=\mathrm{softmax}(b_{ij})\tag{5}$$

$$\boldsymbol{s}_j=\sum_{i=1}^{(L-w_1-w_2+2)\times B/e^1}c_{ij}\cdot\boldsymbol{p}_{j|i}\tag{6}$$

$$\boldsymbol{o}_j = \text{squash}(\boldsymbol{s}_j) = \frac{\boldsymbol{s}_j}{||\boldsymbol{s}_j||} \cdot \frac{||\boldsymbol{s}_j||^2}{0.5 + ||\boldsymbol{s}_j||^2} \tag{7}$$

其中，$\boldsymbol{p}_i$代表第i个低层胶囊，$\boldsymbol{W}_{ij}$是特征映射矩阵，代表低层胶囊i和高层胶囊j的关系，b_{ij}表示每对低层和高层胶囊之间一致性的分数，初始值为0。在动态路由过程的每次迭代中，b_{ij}通过式(8)进行更新，公式为

$$b_{ij} = b_{ij} + \text{squash}(\boldsymbol{s}_j) \cdot \boldsymbol{p}_{j|i} \tag{8}$$

本文对胶囊网络[8]的挤压函数进行了改进，挤压函数第1项的作用是将$\boldsymbol{s}_j$单位化，第2项的作用是将$\boldsymbol{s}_j$挤压到[0,1]区间，即当$\boldsymbol{s}_j$很长时，将其拉长为1，当$\boldsymbol{s}_j$很短时，将其压扁为0。第2项分母的数字1代表挤压程度，挤压程度越大总体数值越小，胶囊就会获得更大的挤压效果。图片分类任务中，模型需要扩大特征之间的差别，但在推荐任务中，体现出特征的相似性有助于学习用户和项目的特征，故本文将其第2项的挤压程度1改为了0.5：$0.5 + ||\boldsymbol{s}_j||^2$，使第2项的权重扩大，胶囊网络可以更多保留胶囊间的相似性。

经过λ次迭代的路由过程，胶囊网络输出得到D个高层胶囊，拼接组成用户i的胶囊矩阵：$\boldsymbol{O}_i = [\boldsymbol{o}_1, \cdots, \boldsymbol{o}_k, \cdots, \boldsymbol{o}_D]$，$\boldsymbol{O}_i \in \mathbb{R}^{D \times e}$编码了文本的细粒度特征信息。为了在融合层将细粒度特征与评论质量信息融合，还需计算每个胶囊的长度，将胶囊矩阵映射到与编号嵌入同维度的向量空间，最终，得到卷积胶囊层的输出：$\boldsymbol{u}_i = [||\boldsymbol{o}_1||, \cdots, ||\boldsymbol{o}_k||, \cdots, ||\boldsymbol{o}_D||]$，$\boldsymbol{u}_i \in \mathbb{R}^D$，同理可得项目$j$在卷积胶囊层的输出$\boldsymbol{v}_j \in \mathbb{R}^D$。

2.3 融合层

在融合层中，本文使用用户和项目编号的嵌入对评论质量进行建模，将编号嵌入分别与用户偏好胶囊和项目特性胶囊相融合，在模型训练时通过这样的融合标记可以给恶意用户和受害项目的胶囊赋予较低的权重，从而降低低质量评论对模型的影响。为了使上文得到的用户偏好胶囊和项目特性胶囊中包含表示评论质量的因素，将用户编号嵌入表示α_i、项目编号嵌入表示β_j与卷积胶囊层获得的$\boldsymbol{u}_i$, $\boldsymbol{v}_j$相加，公式为

$$\boldsymbol{s}_i = \boldsymbol{u}_i + \alpha_i \tag{9}$$

$$\boldsymbol{t}_j = \boldsymbol{v}_j + \beta_j \tag{10}$$

其中，$\boldsymbol{s}_i$,$\boldsymbol{t}_j$为融合了评论质量的用户i的细粒度偏好和项目j的细粒度特性。对$\boldsymbol{s}_i$和$\boldsymbol{t}_j$再次进行变换映射，计算用户i的偏好与项目j的特征的相关程度$\tilde{r}_{ij}$。

$$\tilde{r}_{ij} = \text{sum}\left[\text{ReLU}(\boldsymbol{s}_i \odot \boldsymbol{t}_j) \cdot \boldsymbol{W}^{\text{pre}}\right] \tag{11}$$

其中，$\odot$代表逐元素相乘，$\boldsymbol{W}^{\text{pre}}$为权重矩阵。计算评分的偏置项，并将以上结果进行整合为

$$b^{\text{r}} = \eta_{\text{i}} + \theta_{\text{j}} \tag{12}$$

$$\hat{r}_{\text{ij}} = \tilde{r}_{\text{ij}} + b^{\text{r}} + b^{\text{pre}} \tag{13}$$

其中，η_i和θ_j为用户偏置和项目偏置，b^{pre}为全局偏置，$\hat{r}_{ij}$为模型输出：用户i对项目j的预测评分。

2.4 模型训练及优化

本文将SACR模型训练过程的损失函数设置为

$$L_r = \frac{1}{2} \sum_{i=1,j=1}^{N,M} (r_{ij} - \hat{r}_{ij})^2 \tag{14}$$

模型使用Adam[17]优化器对目标函数进行优化，Adam优化器可以在训练过程中自动调整各个参数的学习速率，并且比普通的SGD优化器更准确、收敛更快。

本文在SACR的融合层使用Dropout[18]方法，Dropout可以在模型训练时随机删除ρ百分比的神经元，在每次模型训练只更新Dropout保留的部分参数，可以缓解模型产生过拟合的问题，提高模型的性能。

3 实验

3.1 数据集和评估标准

本文实验使用来自Amazon5-core[19]的不同领域的8个数据集。数据集中包含用户编号、项目编号、评分、评论、评论有用性等9种属性，本文使用其中的用户编号、项目编号、评分和评论。Amazon已将数据设置为每个用户和项目都至少有5条评论，保证了足够的评论文档用以提取特征。数据集统计如表2所示。

对每个数据集随机以8:1:1的比率构建训练集、验证集和测试集，在实验过程中选取均方误差(Mean Square Error，MSE)作为模型实验效果的评判标准。为了验证模型的有效性，实验选取以下5个模型与SACR进行对比：

(1) PMF[20]：概率矩阵分解模型，仅使用评分矩阵进行矩阵分解预测用户对项目的评分。

(2) ConvMF[4]：使用项目评论文档作为辅助，利用CNN挖掘项目特征，结合PMF进行预测评分。

(3) DeepCoNN[5]：使用CNN分析项目评论和用户评论的深度模型。

(4) NARRE[6]：使用CNN分析项目评论和用户评论，并考虑了评论质量的深度模型。

(5) CARP[13]：使用注意力机制细粒度分析评论文档，并利用胶囊网络预测评分的深度学习模型。

3.2 超参数设置与调整

3.2.1 胶囊数量和胶囊维度

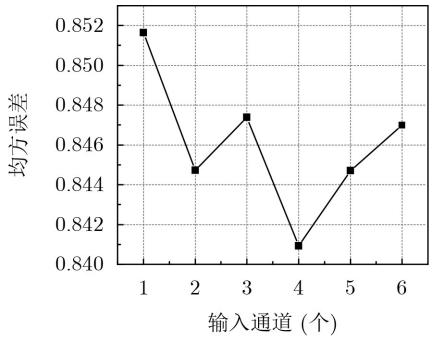
输入胶囊的通道数C和维度e^1决定了胶囊网络处理文本特征的范围和方向，输出胶囊的数量D和维度e决定了特征的维度和细粒度。为确定以上参数，本文在Musical Instruments数据集上测试了使用不同参数的情况下模型预测评分误差的变化。由于胶囊网络的动态路由是一个逐步求精的过程，故输出胶囊的数量和维度要少于输入胶囊的数量和维度，在研究某个参数时，将其他参数设定为固定值。实验结果如图2所示。从图2(a)和图2(c)观察到，当胶囊数量取值较小时，胶囊网络能处理的特征数目相应较少，挖掘到的用户偏好和项目特性不完整，导致误差较大，当胶囊数量取值过大时，又会使特征产生冗余，同样会增高模型误差；从图2(b)和图2(d)观察到，当胶囊维度取值较小时，特征细节体现得不完整，取值过大会使胶囊的细粒度信息产生冗余。根据结果，本文在实验中设定使模型MSE达到最小的参数值：$C=4, e^1=16, D=64, e=4$。

3.2.2 动态路由迭代次数

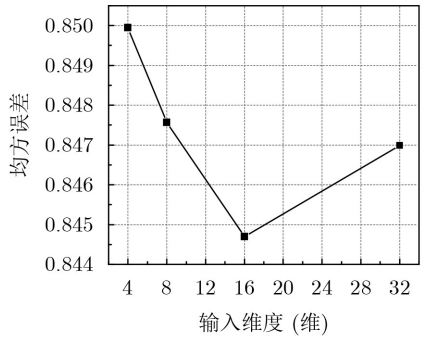
胶囊网络中动态路由的迭代次数λ决定了胶囊网络输出特征的精度。本文在Musical Instruments, Office Products数据集上测试了模型使用不同动态路由的迭代次数时MSE的变化情况。实验结果如图3所示，当迭代次数从1次增加到3次时，模型在两个数据集上的MSE均明显降低，且当$\lambda=3$时模型的MSE达到最低，说明胶囊网络在超过两次迭

表 2　对每个数据集的统计

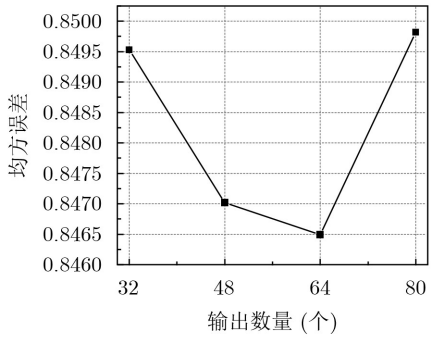
数据集	用户	项目	评论	用户平均评论	项目平均评论
Musical Instruments	1429	900	10261	7.2	11.4
Office Products	4905	2420	53258	10.9	22.0
Digital Music	5540	3568	64706	11.7	18.1
Tools and Improvement	16638	10217	134476	8.1	13.2
Video Games	24303	10672	231780	9.5	21.7
Toys and Games	19412	11924	167597	8.6	14.1
Kindle Store	68223	61935	982619	14.4	15.9
Movies and TV	123960	50052	1679533	13.5	33.6



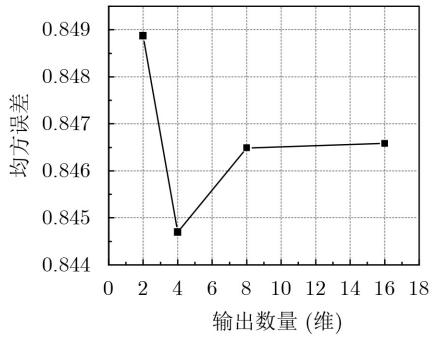
(a) 不同输入通道数对模型均方误差的影响



(b) 不同输入维度对模型均方误差的影响



(c) 不同输出数量对模型均方误差的影响



(d) 不同输出维度对模型均方误差的影响

图 2 不同数量和维度的胶囊对模型均方误差的影响

代之后，就已经找到了较精确的高层胶囊。由于迭代次数选为3次时已经可以挖掘出准确的细粒度特征，所以在继续迭代后，MSE并没有继续降低，反而在升高后趋于平滑。根据结果，本文在实验中选择使模型MSE最低时的迭代次数：$\lambda = 3$。

3.2.3 挤压函数调整

为证明本文改进的挤压函数可以学习胶囊间的相似性，进而更易学习用户的偏好和项目的特性，本节将挤压程度作为参数，设置挤压程度为1.00, 0.75, 0.50, 0.25, 0，并将模型在Musical Instruments, Office Products数据集上进行实验。由于挤压程度设置为0可能会造成挤压函数中分母为0的错误，故使用极小值1e-7替代。实验结果如图4所示。当挤压程度为0时，挤压函数就变成了普通的归一化函数，失去了挤压的作用，模型误差最高；挤压率为0.25时，函数具有一定的压扁和拉长胶囊的效果，模型误差降低；挤压程度为0.50时模型误差最低；挤压程度增加到0.75和1.00时，挤压效果最强，但扩大了胶囊的差别，忽略了胶囊的相似性，故学习到的特征间联系变小，模型误差反而上升。实验证明使用本文挤压函数挖掘的特征更准确。

3.3 模型对比

本文同时在Amazon5-cores的8个数据集上对SACR以及基准模型进行实验，结果如表3所示。

首先，PMF仅使用评分矩阵学习用户和项目的特征，在评分数据稀疏时模型受影响较大，所以相对于其他基于评论文档的模型性能是最差的；在基于评论文档的基准模型中，ConvMF引入了评论文档作为辅助信息，并且在PMF的基础上使用CNN处理项目评论，得到的项目特征更加完整，模型性能也较PMF有较大的提升，但其用户特征还不准确，故与其他模型相比性能较差；DeepCoNN使用CNN处理用户和项目评论，并且使用全连接层代替矩阵分解模型，模型的性能显著高于基于矩阵分解的ConvMF，但CNN不能准确挖掘评论中的细粒度特征，模型性能仍不是基准模型中最高的；NARRE在DeepCoNN的基础上使用了注意力机制，并考虑了评论质量因素，在其工作[6]中选取的Toys and Games, Kindle Store, Movies and TV上模型性能是基准模型中最高的，但在其他的数据集上仅强于PMF，说明模型基于CNN挖掘特征不够准确，在稀疏的数据集上泛化能力较弱；CARP使用注意力机制挖掘评论文档的细粒度特征，并使用胶囊网络进一步分析用户项目交互的情感，在前5个数据集中的表现是基准模型中最高的，在后3个数据集中性能与NARRE持平，但其使用注意力挖掘到的特征仍有限，且模型没有考虑到评论质量对结果的影响；本文模型SACR在各个

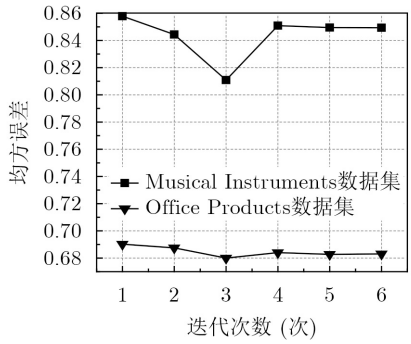


图 3 不同的迭代次数在两个数据集上对模型均方误差的影响

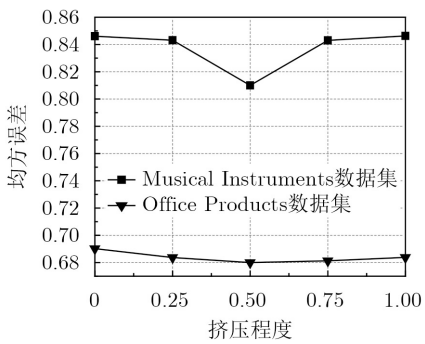


图 4 不同的挤压程度在两个数据集上对模型均方误差的影响

表 3 各模型实验结果对比

数据集	PMF	ConvMF	DeepCoNN	NARRE	CARP	**SACR**
Musical Instruments	1.398	0.903	0.893	1.004	**0.800**	0.810
Office Products	1.092	0.767	0.698	0.931	0.728	**0.680**
Digital Music	1.206	0.876	0.809	1.270	0.889	**0.796**
Tools Improvement	1.566	1.056	0.958	1.196	0.964	**0.924**
Video Games	1.672	1.174	1.134	1.205	1.166	**1.104**
Toys and Games	1.711	0.877	0.810	0.796	0.800	**0.780**
Kindle Store	0.984	0.623	0.621	0.605	0.610	**0.590**
Movies and TV	1.669	1.010	1.026	0.994	0.997	**0.930**
平均MSE	1.412	0.911	0.866	1.001	0.869	**0.826**

数据集的平均MSE显著低于其他模型，且在Office Products, Digital Music, Movies and TV上相对于CARP取得了6.5%，10.5%，6.7%的显著改进，在Musical Instruments中性能与CARP持平的原因是设置了不同于其他数据集的迭代次数和特征维度，虽然准确率提高，但结果导致模型泛化能力降低。在其他数据集中SACR的性能显著高于CARP。

综上所述，在使用相同评论数据集的情况下，本文模型SACR与使用CNN的其他模型相比取得了最低的预测误差，这说明胶囊网络在模型训练的过程中可以比CNN挖掘出更准确的文本特征，由于挖掘出的特征更准确，故可以获得更高的准确率。

3.4 模型有效性分析

为了验证SACR避免低质量评论影响的有效性以及使用胶囊网络进行细粒度特征挖掘的合理性，使用SACR的两个子模型进行对比实验，两个子模型的定义如下：

(1) SACR-base：不包含SACR的融合层中结合评论质量的部分，用于证明SACR可以消除低质量评论对预测结果的负面影响。

(2) SACR-cnn：将SACR的卷积胶囊层替换为连续3次卷积和池化操作的CNN，用于证明SACR的自注意力胶囊网络可以更细粒度地提取特征。

SACR-cnn相当于在DeepCoNN的基础上增加了考虑评论质量的融合层，SACR-base相当于将DeepCoNN的CNN替换成本文的卷积胶囊层。分别将SACR-base和SACR-cnn在Amazon5-cores的5个数据集进行实验，并与SACR以及基准模型DeepCoNN进行对比。实验结果如表4所示。

首先，使用胶囊网络的SACR和SACR-base的性能均高于使用CNN的SACR-cnn和DeepCoNN，说明将神经元扩展成“胶囊”的方式使模型在训练过程中能比CNN携带更丰富的信息，证明SACR能捕获用户和项目的细粒度特征，解决了CNN挖掘文本特征不准确的问题。其次，融入编号信息标记评论质量的SACR和SACR-cnn的性能均高于没有考虑评论质量的DeepCoNN，证明SACR通过对用户项目进行标记，在模型训练时可以赋予恶意用户和受害项目的评论较低权重，进而消除低质量评论对预测结果的负面影响。

表 4　子模型预测准确率实验结果对比

模型	Musical Instruments	Office Products	Digital Music	Tools and Home Improvement	Video Games
SACR-base	0.853	0.681	0.805	0.924	1.105
SACR-cnn	0.886	0.691	0.808	0.936	1.106
DeepCoNN	0.893	0.698	0.809	0.958	1.134
SACR	**0.810**	**0.680**	**0.796**	**0.921**	**1.104**

本节实验证明了SACR可以解决使用CNN挖掘文本特征不准确以及低质量评论影响结果精度的问题，SACR可以给用户生成更加准确的推荐结果。

4　结束语

本文提出了一种考虑评论质量的自注意力胶囊网络评分预测模型SACR。使用自注意力胶囊网络处理评论文档，并改进了胶囊的挤压函数，使其能更准确地获取评论文本中的细粒度特征，通过将用户、项目标识信息与细粒度特征相融合，进一步消除低质量评论对特征挖掘的负面影响。经过与各种基准模型的实验对照，证明SACR可以更有效地提升评分预测的准确率。此外，本文还通过SACR与子模型的对比实验，证明了使用自注意力胶囊网络挖掘特征可以解决使用CNN挖掘文本特征不准确的问题，将用户和项目信息标记评论可以避免低质量评论对结果的负面影响。

在未来的工作中考虑将时间因素加入到模型中，分析模型在用户评论的时间分布不同的情况下，对应的细粒度特征的变化；考虑通过其他上下文信息完善模型对推荐结果的可解释性。

参 考 文 献

[1] TAN Yunzhi, ZHANG Min, LIU Yiqun, *et al*. Rating-boosted latent topics: Understanding users and items with ratings and reviews[C]. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, USA, 2016: 2640–2646.

[2] ZHANG Wei, YUAN Quan, HAN Jiawei, *et al*. Collaborative multi-level embedding learning from reviews for rating prediction[C]. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), New York, USA, 2016: 2986–2992.

[3] BLEI D, CARIN L, and DUNSON D. Probabilistic topic models[J]. *IEEE Signal Processing Magazine*, 2010, 27(6): 55–65. doi: 10.1109/MSP.2010.938079.

[4] KIM D, PARK C, OH J, *et al*. Convolutional matrix factorization for document context-aware

recommendation[C]. Proceedings of the 10th ACM Conference on Recommender Systems, Boston, USA, 2016: 233–240. doi: 10.1145/2959100.2959165.

[5] ZHENG Lei, NOROOZI V, and YU P S. Joint deep modeling of users and items using reviews for recommendation[C]. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, 2017: 425–434. doi: 10.1145/3018661.3018665.

[6] CHEN Chong, ZHANG Min, LIU Yiqun, *et al*. Neural attentional rating regression with review-level explanations[C]. Proceedings of the 2018 World Wide Web Conference, Lyon, Italy, 2018: 1583–1592. doi: 10.1145/3178876.3186070.

[7] LU Yichao, DONG Ruihai, and SMYTH B. Coevolutionary recommendation model: Mutual learning between ratings and reviews[C]. Proceedings of the 2018 World Wide Web Conference, Lyon, Italy, 2018: 773–782. doi: 10.1145/3178876.3186158.

[8] SABOUR S, FROSST N, and HINTON G. Dynamic routing between capsules[C]. NIPS, Los Angeles, USA, 2017: 3859–3869.

[9] ZHANG Ningyu, DENG Shumin, SUN Zhanling, *et al*. Attention-based capsule networks with dynamic routing for relation extraction[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018: 986–992. doi: 10.18653/v1/d18-1120.

[10] ZHANG Min and GENG Guohua. Capsule networks with word-attention dynamic routing for cultural relics relation extraction[J]. *IEEE Access*, 2020, 8: 94236–94244. doi: 10.1109/ACCESS.2020.2995447.

[11] ZHAO Wei, YE Jianbo, YANG Min, *et al*. Investigating capsule networks with dynamic routing for text classification[C]. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018: 3110–3119. doi: 10.18653/v1/D18-1350.

[12] KIM J, JANG S, PARK E, *et al*. Text classification using capsules[J]. *Neurocomputing*, 2020, 376: 214–221. doi: 10.1016/j.neucom.2019.10.033.

[13] LI Chenliang, QUAN Cong, PENG Li, *et al*. A capsule network for recommendation and explaining what you like and dislike[C]. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, 2019: 275–284. doi: 10.1145/3331184.3331216.

[14] LI Chao, LIU Zhiyuan, WU Mengmeng, *et al*. Multi-interest network with dynamic routing for recommendation at tmall[C]. The 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 2019: 2615–2623. doi: 10.1145/3357384.3357814.

[15] OLATUNJI I E, LI Xin, and LAM W. Context-aware helpfulness prediction for online product reviews[C]. 15th Asia Information Retrieval Societies Conference on Information Retrieval Technology, Hong Kong, China, 2019: 56–65. doi: 10.1007/978-3-030-42835-8_6.

[16] 丁永刚, 李石君, 付星, 等. 面向时序感知的多类别商品方面情感分析推荐模型[J]. 电子与信息学报, 2018, 40(6): 1453–1460. doi: 10.11999/JEIT170938.

DING Yonggang, LI Shijun, FU Xing, *et al*. Temporal-aware multi-category products recommendation model based on aspect-level sentiment analysis[J]. *Journal of Electronics & Information Technology*, 2018, 40(6): 1453–1460. doi: 10.11999/JEIT170938.

[17] KINGMA D P and BA L J. ADAM: A method for stochastic optimization[C]. International Conference on Learning Representations (ICLR), San Diego, USA, 2015: 1–15.

[18] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, *et al*. Dropout: A simple way to prevent neural networks from overfitting[J]. *The Journal of Machine Learning Research*, 2014, 15(1): 1929–1958. doi: 10.5555/2627435.2670313.

[19] MCAULEY J, TARGETT C, SHI Qinfeng, *et al*. Image-based recommendations on styles and substitutes[C]. The 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, USA, 2015: 43–52. doi: 10.1145/2766462.2767755.

[20] SALAKHUTDINOV R and MNIH A. Probabilistic matrix factorization[C]. Proceedings of the 20th International Conference on Neural Information Processing Systems, Vancouver, Canada, 2007: 1257–1264. doi: 10.5555/2981562.2981720.

梁顺攀：男，1976年生，副教授，研究方向为推荐系统.
刘 伟：男，1996年生，硕士生，研究方向为推荐系统.
尤殿龙：男，1981年生，副教授，研究方向为特征选择.
刘泽谦：男，1996年生，硕士生，研究方向为推荐系统.
张付志：男，1964年生，教授，研究方向为推荐系统.

责任编辑：陈 倩