

基于深度学习加强的混合推荐方法*

丁弼原, 张敏, 谭云志, 刘奕群, 马少平

(清华信息科学与技术国家实验室(筹), 清华大学计算机系, 智能技术与系统国家重点实验室, 北京, 100084)

z-m@tsinghua.edu.cn

摘要: 近年来基于矩阵分解的协同过滤算法在评分预测上取得的显著成果, 但冷启动、数据稀疏等问题仍然未能得到很好的解决, 因此如何将评论信息引入推荐系统以缓解上述问题, 开始成为研究的热点之一。另一方面深度学习在图像处理、自然语言处理等领域已经发挥重要作用, 其对文本的编码与分析能力恰好适应推荐领域当前正在研究的问题。本文尝试基于深度学习来加强个性化推荐, 提出将层叠降噪自动编码器 (SDAE) 与隐含因子模型 (LFM) 相结合的混合推荐方法, 综合考虑评论文本与评分, 以提升推荐模型对潜在评分预测的准确性。在常用大规模公开数据集 Amazon 上进行的测试结果表明, 与传统推荐模型相比, 所提出方法可有效提高评分预测的准确性, 最高可达到 64.43% 的性能提升。

关键词: 深度学习; 协同过滤; 混合推荐

Hybrid Recommendation Approach Enhanced by Deep Learning

DING Biyuan, ZHANG Min, TAN Yunzhi, LIU Yiqun, MA Shaoping

(State Key Laboratory of Intelligent Technology and Systems Tsinghua National Laboratory for Information Science and Technology Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Collaborative filtering based on matrix factorization has achieved great success, while cold-start and data sparseness problems have not been well-solved. Hence it has been a significant trend to involve review information into rating prediction as one of the dominant solutions. On the other hand, deep learning has played a vital part in image and natural language processing recently, which meets the requirement of current recommendation research. This paper proposes a hybrid model that introduces deep learning in recommendation system with collaborative filtering. It combines Stacked Denoising Auto Encoder (SDAE) with Latent Factor Model (LFM) to make use of both review and rating information so as to improve the performance on rating prediction. Comparative evaluation has been made on large-scale commonly-used Amazon dataset. Results show that the proposed approach improves the rating prediction precision greatly in comparison with traditional models, revealing a up to 64.43% performance improvement.

Key words: deep learning; collaborative filtering; hybrid recommendation

1 引言

推荐算法的研究来源于互联网公司的实际需要。在互联网上, 根据用户的行为记录向用户推荐可能感兴趣的商品或者是常见的需求。协同过滤是目前常见的推荐算法之一。协同过滤基于所有用户对商品的评分记录, 分析出用户之间的关系和商品之间的关联, 从而预测用户对某个商品的潜在评分。协同过滤算法不关注商品的内容和用户的个人信息, 仅关注商品和用户、商品和商品以及用户和用户之间的关联, 因此, 协同过滤可以看作是根据其他用户的行为和当前用户的历史行为来预测当前用户的未来行为的算法。

在实际条件下, 协同过滤算法面临的挑战主要是冷启动问题和数据稀疏性问题。

* **基金项目:** 国家 973 计划(2015CB358700), 国家自然科学基金(61672311, 61622208, 61532011, 61472206) 资助

冷启动问题是指面对新用户时推荐算法准确率低的问题。协同过滤算法基于用户的历史记录,而对于新用户来说,历史记录很少,因此,协同过滤算法也就不能准确地估计用户的个人偏好,从而导致预测准确性下降。从实际的数据库中我们可以发现,在美国亚马逊商城购物的用户中,有 94%的用户只有不超过三条历史评论。在如此稀疏的评论下,依赖历史评论的协同过滤算法的预测性能会受到影响。

数据稀疏性问题是相对于用户数和商品数,真正存在的评论数量只占有极少的一部分。以美国亚马逊的电影类别评论为例,在这个类别下,一共有 827806 个用户购买过电影,电影商品总数为 241599 个,因此,我们可以得出电影类别中的用户评分矩阵有 2×10^{12} 个元素,而实际的评分数只有 6330033 条,只占全部可能评分的 0.0003%,可见实际的评分矩阵是非常稀疏的。因此,协同过滤算法必须能够处理如此稀疏的矩阵。

为了克服协同过滤算法的缺陷,人们将基于内容的推荐方法与协同过滤方法融合起来,形成混合推荐模型。进一步地,考虑到在很多互联网服务平台例如电子商务网站上,用户对商品的评论不仅包括评分,也包括评论文本。与评分相比,评论文本蕴含了更多的信息。在实际情况下,即使有两个用户对同一个商品打出相同的分数,他们也可能出于完全不同的理由。传统的协同过滤方法仅仅考虑用户的评分,而不考虑评论文本。对于绝大多数评论少的沉默用户和新用户而言,协同过滤的预测准确度并不高。因此,综合利用评论文本和评分信息,一起来帮助提高推荐系统的效果,成为当前个性化推荐研究的一个热点方向。

深度学习方法已经在图像处理、自然语言处理等领域发挥了重要作用,其对文本的编码与分析能力恰好适应推荐领域当前正在研究的方向:充分利用评论文本信息,将其与用户评分之间建立起关联关系,综合起来进行推荐。当前将深度学习方法引入推荐系统的相关研究还很少,特别是算法效率很低,无法在大规模数据上的真实推荐场景中得到应用。因此,研究提出一个基于深度学习方法、融合评论与评分信息的、高效的混合推荐模型就成为一个前沿的有价值的研究课题,也是本文研究工作的核心。

2 相关工作

2.1 协同过滤及基于内容与协同过滤的混合推荐方法

传统的协同过滤推荐模型可以分为三种类别,分别是:基于记忆的推荐[1]、基于模型的推荐[2,3,4]和混合型推荐[5,6]等。

基于记忆的推荐早期研究很多,随着推荐系统的发展,基于模型的算法成为协同推荐的主流,其中矩阵分解方法[2,4]是最主要的一类方法。目前最成功的基于模型的协同过滤算法之一是 LFM[2]算法。LFM 算法使用矩阵分解方法降低数据的维度,提取隐含信息,将用户评分矩阵分解为用户偏好矩阵和商品特征矩阵。在用户偏好矩阵中,每个用户被看作其中的一个向量。在商品特征矩阵中,每个商品被看作其中的一个向量。在实际预测评分是,LFM 算法使用给定用户的偏好向量和给定商品的特征向量去预测用户对商品的潜在评分。其他基于模型的算法中的矩阵分解模型包括 SVD 算法[4]和 NMF 算法[7]。

另一些协同过滤算法结合了基于内容的推荐[8,9,10,11,12]。与仅仅基于用户对商品的评分的协同过滤算法相比,融合了基于内容的混合推荐算法从商品的内容和用户个性化的偏好设置出发,提高了对新用户和新商品的推荐准确性。

如何对内容加以理解,是这一类混合方法的核心之一。这方面近年来一个经典工作 HFT[8]将主题模型 LDA[13]与协同过滤相结合,在推荐准确率上比单纯的 LFM 模型有明显提升。相比于 LFM 算法通过矩阵分解得到的商品特征,从 LDA 学习到的商品特征具有更加明显的语义,即模型中每个维度都是一个主题。其他同时考虑评分与文本的推荐算法还包括 TopicMF[12]等,该模型使用主题模型中的 NMF 方法分析评论,并与使用矩阵分解提取

隐含特征的协同过滤模型相结合。BoWLF[9]模型根据商品特征向量生成商品的评论文档的词概率分布，并将在给定商品特征参数下生成文档的似然度作为 LDA 部分的正则化项，使得模型在训练过程中能够反映评论文本的信息。

2.2 结合深度学习的推荐模型

深度学习已经在图像处理、自然语言处理等领域发挥了重要的作用，相比于传统算法取得了巨大的成功，如卷积神经网络（Convolutional Neural Network, CNN）在图像识别上取得很大的进展[14,15,16]。在自然语言处理领域，基于长短期记忆网络（Long Short-Term Memory, LSTM）[17]在情感分析[18]、双语翻译[19]等任务上也取得了很好的效果。

而当前将深度学习应用于推荐系统中的研究才刚刚起步。在国内，陈达提出了基于 DBN 的推荐系统[20]，杨宇提出了基于深度学习的图像推荐系统[21]。国际上当前相关领域的研究工作中主要有两个重要进展：一是 LMLF[22]模型，它将 LSTM 与协同过滤相结合，使用 LSTM 从商品特征向量生成商品评论的词概率分布，并计算在当前参数下该商品评论的似然度，将该似然度作为 LFM 模型的损失函数的一部分，使得 LFM 模型在训练的过程中可以反映评论文本的特征。然而 LMLF 模型在评分预测中的表现低于更为简单的 BoWLF 模型[9]，而且运行效率差，使得它的实际可用性不高。另一个工作 CDL[23]模型将贝叶斯 SDAE 模型[24]与概率化 LFM 模型相结合，预测用户对商品评分的数学期望，通过物品的描述和标签提升模型的性能。然而 CDL 模型并没有考虑沉默用户问题，而在真实应用环境下（如前文所述的亚马逊数据）我们可以发现沉默用户占有所有用户中的绝大多数。

由此可见，利用深度学习网络来加强对内容的分析，设计融合评论文本与评分数据的推荐方法是相关研究领域的一个趋势所在，而当前相关研究还很不成熟，仅有的几个相关工作存在效果不明显、算法效率低、很难在大规模数据上的真实推荐场景中得到应用等问题。这也是本文提出语义增强的隐含因子模型的所要研究和试图解决的问题。

3 语义增强的隐含因子模型的设计与学习

综合深度学习网络和协同过滤方法，我们提出语义增强的隐含因子模型（Semantics Enhanced Latent Factor Model，缩写为 SELFM），通过引入商品评论文本的自然语言模型来增强矩阵分解结果的一般性。模型中需要用到深度学习方法层叠降噪自动编码器（Stacked Denoising AutoEncoder, SDAE）和协同过滤方法隐含因子矩阵分解（LFM）算法。

3.1 模型框架设计

基于 SDAE 和 LFM 提出的语义增强的隐含因子模型（SELFM）主要框架如图 1 所示。

我们将每个商品都表示为一个向量，称之为商品内容向量，并将这个向量作为文本编码模型的输入。相比于将每个评论当作一个文档，将每个商品的所有评论当成一个文档具有较大的优点：因为用户在评论文本中通常阐述了其评分的具体原因和对商品的描述，所以，如果将一个商品的所有评论合并成一个文档，那么这个文档就是一个对商品的自然语言层面的描述。此外，因为不同用户对同一商品的评价可能存在差异甚至矛盾信息，因此如果只用单条评论作为文档，则会带来输入信息的矛盾，而我们这里将多条评论合并为一个文档，通过词的权重以及 SELFM 中对文本特征的编码降噪，使得文档中的主流评价观点得以体现，这也在较大程度上增加了模型的抗噪性。在 SELFM 模型中，商品的自然语言描述对应于商品的特征向量，通过这种方式，我们增强了矩阵分解模型中商品特征向量的语义性信息。所以，我们将每个商品的所有评论合并成一个文档。

该模型由三部分组成，分别是选词模型、文本编码模型和隐含因子矩阵分解模型。

其中选词模型的作用确定文本编码模型所用的词空间。SDAE 只能接受定长的输入，而评论文本的词汇量却相对巨大而且变动，因此不能像传统的评论文本分析方法那样简单地使

用全部词来构成内容的表示。在评论文本所包含的所有词中，有大量的词只出现极少次，因此难以作为判断商品特征的关键词，而有少量常见词则在评论文本中大量出现，并且没有区分度。因此，很自然地，我们可以采用文本处理中常用的 TFIDF 来确定词的重要性。我们将词按照 TFIDF 大小排序，选出值最高的一组，组成一个词空间。对于每个商品，我们将该商品的所有评论合并为一个文档，并且用一个与词空间同维度的向量来表示这个文档。

文本编码模型的用途是将高维度的商品内容向量压缩为低维度的商品特征向量，并且尽可能多地保留文本的信息。对于这类任务，我们使用 SDAE 作为压缩向量的模型。

在使用选词模型处理了所有的评论后，我们得到每个商品的评论内容向量。这些商品内容向量用来训练 SDAE 中的参数。训练 SDAE 时，我们按顺序训练 SDAE 中的每一层降噪自动编码器，通过最小化公式 (3.1) 来估计每层降噪解码器的参数。

在文本编码模型的参数被估计后，我们估计隐含因子矩阵分解模型的参数，并且微调文本编码模型的参数。

文本编码模型和隐含因子矩阵分解模型在本框架中的应用和学习过程在如下两个小节中分别描述。

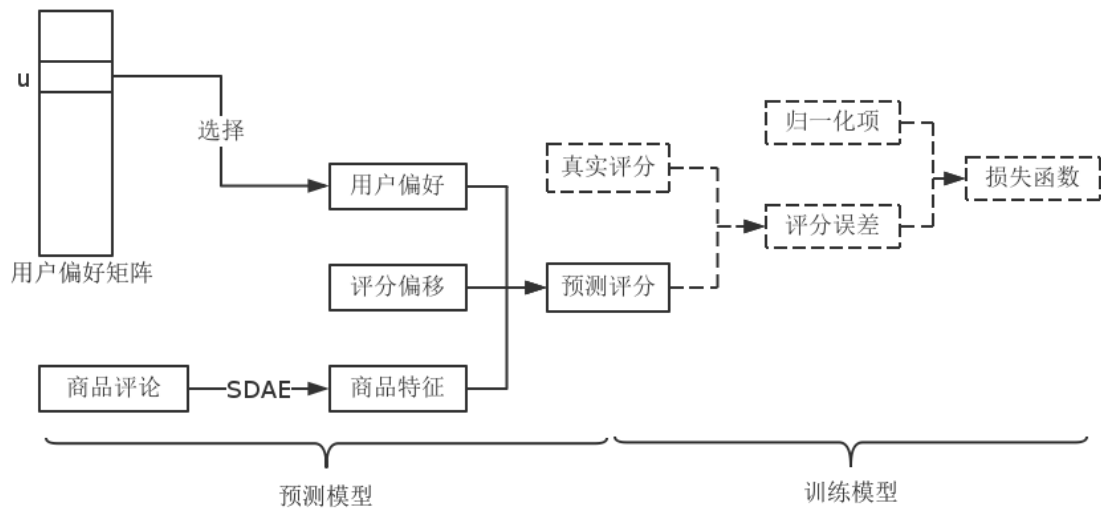


图 1 SELFM 模型框架

3.2 基于层叠降噪自动编码器（SDAE）的文本编码模型

DAE（降噪自动编码器）是一个以尽量少的损失将高维向量压缩为低维向量的网络。自动编码器的网络结构可以看作由一个编码器层和一个解码器层组成。编码器层将输入的高维向量压缩为输出的低维向量，解码器层将低维的向量还原成高维向量。编码器的优化目标是使得解码器还原出来的高维向量尽量接近输入的原始高维向量。

假设输入的高维向量为 x 其维度为 N ，压缩成的低维向量为 y ，其维度为 M ，解码器还原出的高维向量为 \tilde{x} ，其维度为 N 。编码器层需要两个参数，分别是 $N \times M$ 维的矩阵 W_e 和 M 维的偏移向量 b_e 。解码器层需要两个参数，分别是 $M \times N$ 维的矩阵 W_d 和 N 维的偏移向量 b_d 。自动编码器的工作原理为输入原始向量 x ，输出压缩后的向量 y ，压缩方法为：

$$y = W_e x + b_e$$

自动解码器的工作原理为输入压缩后的向量 y ，输出还原的向量 \tilde{x} ，解码方法为：

$$\tilde{x} = W_d y + b_d$$

自动编码器的压缩损失为 L_e ，可以定义为还原的向量与原始向量的 L2 距离的平方，即：

$$L_e = \sum_{i=1}^{i \leq N} (\tilde{x}_i - x_i)^2 \quad (3.1)$$

在训练自动编码器时，以减少压缩损失 L_e 为优化目标。

在实际情况中，输入数据会含有噪音，输入数据少的时候，也会导致编码器过拟合到训练数据集上，因此，在实际的模型中，我们增加一个降噪层，用于提高自动编码器的编码能力。降噪层是一个由随机的二元分布的元素组成的层。降噪层的维数与输入层相同，均为 N 维，每一维的元素都是一个符合二元分布的随机变量 μ_i ，即 $\mu_i \sim B(p)$ ，其中 p 是二元分布的参数。输入数据与降噪层的按元素积为编码器的输入，即加入了降噪层的编码器层计算公式为¹：

$$y = W_e(\mu \otimes x)$$

在实际使用中，解码器的参数矩阵 W_d 通常为编码器的参数矩阵 W_e 的转置。此时学习出的 W_e 矩阵接近正交矩阵。

由于一层 DAE 网络的编码能力有限，因此，我们考虑将多层 DAE 网络堆叠起来，组成编码能力更强的 SDAE 网络。多个 DAE 编码器按顺序首尾相连，相邻的 DAE 层之间插入 sigmoid 层，组成 SDAE 网络。Sigmoid 层作为活化层，使得 SDAE 网络具有比 DAE 网络更强的处理能力。

训练 SDAE 网络由分层训练和微调两部分组成。首先训练最外层 DAE。输入的数据经过最外层编码器编码后的结果直接通过最外层的解码器解码，通过训练最外层的编码器，使得编码后的损失尽量小。当训练次数达到指定次数后，停止最外层编码器的训练。在最外层的编码器训练完成后，训练第二层编码器。在第二层编码器训练完成后，再按照相同的方法训练内部的编码器，直到所有编码器均分别训练完成。

在所有编码器层均训练完成后，就不再考虑解码器部分，只保留编码器连接成的网络，以最外层编码器的输入为整个网络的输入，以最内层编码器的输出为整个网络的输出。之后，再微调整个网络以适应实际的需求。

3.3 隐含因子矩阵分解方法

如果将用户对商品的评分看作一个矩阵，其中每一行代表一个用户对所有商品的评分，每一列代表一个商品的所有用户评分，那么，矩阵的每个元素都是某个用户对某个商品的评分。假设用户数为 U ，商品数为 I ，评分矩阵为 R ，即 $R = (r_{u,i})_{U \times I}$ 。我们将评分矩阵 R 分解为两个低维度矩阵 $P_{U \times N}$ 与 $Q_{I \times N}$ ，其中 N 是远小于 U 和 I 的整数，表示用户和商品的隐含特征空间的维数。在矩阵分解模型下，用户 u 对商品 i 的预测评分为：

$$\tilde{r}_{u,i} = p_u q_i^T$$

从用户和商品的整体范围内考虑，我们也可以使用基于平均偏差的预测模型。其中最基础的是所有评分的平均值 α 。在 α 之上，我们考虑每个用户有自己的评分习惯，因此对每个用户引入一个用户评分偏移量 b_u 。每个商品的质量不同，我们也对每个商品引入一个商品评分偏移量 b_i 。将平均偏差模型与矩阵分解模型相结合，可以得到带有平均偏差的评分预测公式：

$$\tilde{r}_{u,i} = \alpha + b_u + b_i + p_u q_i^T$$

模型训练的目标是使预测的评分尽量接近真实评分。如果所有评论的集合为 $H = \{(u, i, r)\}$ ，用户 u 对商品 i 的预测评分为 $\tilde{r}_{u,i}$ ，那么，对我们通过最小化 L_H 来估计参数：

$$L_H = \sum_{(u,i,r) \in H} (r - \tilde{r}_{u,i})^2$$

基于矩阵分解的协同过滤算法在稀疏的评分矩阵上容易过拟合，导致模型的一般性较低。

¹ \otimes 运算符代表按元素乘法

我们通过文本编码模型来修正矩阵分解的过程，并提高模型在预测评分上的准确率。我们的方法是将通过矩阵分解得到了商品特征向量替换为文本编码模型输出的商品特征向量。矩阵分解得到的商品特征向量在语义上缺乏意义，而文本编码模型输出的商品特征向量则反映了商品的特征。因此，文本编码模型输出的商品特征向量更加具有代表性。

除了预测准确性，我们希望避免模型过拟合的情况，在避免模型过拟合的方法中，最常见的方法是引入参数衰减项（Weight Decay）。对于隐含因子的矩阵分解模型来说，除了参数 α ，其他参数 b_u 、 b_i 、 p_u 和 q_i 都是数值上在零上下波动的偏移量。对于这些在零上下波动的偏移量参数，参数的方差越大，那么参数所对应的模型就越复杂，因此也就更容易过拟合。通过限制参数矩阵的范数，我们可以减小模型中参数的方差，从而使得模型更加简单。然而，如果参数的方差过小，会发生欠拟合现象，也就是模型过于简单，以至于不能够有效地反映用户对商品预测的隐含因子。所以，能够影响参数方差大小的参数衰减项系数会影响到模型的性能，需要谨慎地选取。我们在实验中也测试了参数衰减项系数对模型性能的影响。

因此，优化隐因子矩阵分解模型的目标就是最小化：

$$L_H = \sum_{(u,i,r) \in H} (r - \hat{r}_{u,i})^2 + \lambda_b (\|B_u\| + \|B_i\|) + \lambda_p \|P_u\|$$

在优化隐含因子矩阵分解模型时，文本编码模型的参数也会根据隐含因子矩阵分解模型任务的独特性进行微调，以便达到更好的评分预测效果。

4 实验结果与分析

4.1 实验数据集及评价指标

我们在美国亚马逊评论数据集上测试 SELFM 模型的准确度。美国亚马逊评论数据集采集自 1995 年到 2013 年间，在美国亚马逊网站上用户的真实评论。这些评论按照亚马逊的商品类别被划分为 24 个子数据集，包括了艺术、汽车、手机、电影、音乐和视频游戏等多个领域，不但具有丰富的评论内容，而且各个子数据集之间沉默用户数、评论规模均不同，对于测试推荐模型在不同场合下的性能很有帮助。

表 1 亚马逊数据集

序号	子数据集	用户数	商品数	评论数	沉默用户率	序号	子数据集	用户数	商品数	评论数	沉默用户率
1	艺术	24071	4211	27980	97.69%	13	indle 商店	116191	4372	160793	95.50%
2	机动车	133256	47577	188728	93.89%	14	乐器	67007	14182	85405	95.00%
3	婴儿用品	13930	1660	17332	97.12%	15	办公用品	110472	14224	138084	96.42%
4	化妆品	167725	29004	252056	91.51%	16	庭院	166832	19531	206250	96.47%
5	手机	68041	7438	78930	97.82%	17	宠物用品	160496	17523	217170	93.21%
6	服装	128794	66370	581933	78.84%	18	鞋	73590	48410	389877	70.86%
7	电子产品	811034	82067	1241778	91.36%	19	软件	68464	11234	95084	96.80%
8	食品	112544	23476	154635	93.46%	20	体育	329232	68293	510991	92.53%
9	健康	311636	39539	428781	93.72%	21	日用工具	283514	51004	409499	92.81%
10	厨房用具	644509	79006	991794	91.90%	22	游戏玩具	290713	53600	435996	94.03%
11	工业	29590	22622	137042	93.90%	23	视频游戏	228570	21025	463669	90.95%
12	首饰	40594	18794	58621	93.24%	24	手表	62041	10318	68356	98.52%

从表 1 可以看出,各个数据集之间有很大差距。在艺术品类别中,沉默用户高达 97.69%,而在鞋类别上,沉默用户只有 70.86%。在这个数据集上的测试结果能够反映出模型在不同情况下的准确率。

我们将实验数据集按照 9:1 的比例随机分成训练集和测试集,对 SELFM 模型和其对比模型均在相同的训练集上测试,并在同样的测试集上运行测试,以确保结果可以比较。我们只考虑训练集中的评论文本,而忽略测试集中的所有评论文本。

在对于性能评价,我们采用推荐领域的经典评价指标 MSE。假设测试集为 $T = \{(u_k, i_k, r_k)\}$,其中 u_k 为第 k 个评论的用户 ID, i_k 为第 k 个评论的商品 ID, r_k 为第 k 个评论的真实评分,并且假设推荐模型对用户 u 在商品 i 上的预测评分为函数 $\tilde{r}(u, i)$,那么 MSE 的定义为:

$$MSE = \sum_{(u_k, i_k, r_k) \in T} (\tilde{r}(u_k, i_k) - r_k)^2$$

MSE 的值越小越好,该值越接近于 0,模型的预测准确率越高。

4.2 对比模型

我们选择 SVD++和 LFM 模型作为协同过滤的代表模型,选择 BoWLF 和 LMLF 模型作为结合评论文本的推荐模型,作为 SELFM 的对比模型。

1. SVD++: SVD++方法使用矩阵分解模型将高维度的用户评分矩阵分解为低维度的用户偏好矩阵和商品特征矩阵,提取在用户评分矩阵中蕴含的信息。SVD++方法在 2007 和 2009 年的 Netflix 竞赛中发挥了重要的作用。
2. LFM: LFM 基于矩阵分解和评分偏移量模型,是目前为止最成功的协同过滤模型之一。
3. LDA-LFM: 相比于 LFM 方法中使用随机均匀分布初始化参数,本模型使用 LDA 分析商品的所有评论,为每个商品对应的文档生成一个主题分布,作为商品特征矩阵的初始值。
4. BoWLF: BoWLF 方法将词袋模型与 LFM 模型相结合,是一种将神经网络与基于协同过滤的推荐系统整合的推荐模型。词袋模型根据商品的特征向量估计评论中词出现的概率,并最大化评论文档的整体概率。通过词袋模型,BoWLF 可以正则化 LFM 模型中的商品特征向量,将商品特征向量与商品评论相联系,从而提高模型的一般性。
5. LMLF: LMLF 方法将 BoWLF 方法中的词袋模型替换为循环神经网络中的长短期记忆网络(LSTM),将深度学习与基于矩阵分解的协同过滤方法相结合,在对评论的模型化能力上强于 BoWLF,但是在评分预测的能力上并没有超过 BoWLF。

4.3 实验参数与模型实现

协同过滤的隐参数维度为 5,与 SVD++、LFM、LDA-LFM 和 BoWLF 使用相同的维度。

SVD++使用开源的 MyMediaLite 实现²来预测评分。LFM、LDA-LFM、BoWLF 使用 Theano³实现,并且与 SELFM 中的 LFM 部分保持实现细节一致。

对于 SELFM 模型,我们在使用 TFIDF 方法选择出的 5000 个 TFIDF 值最高的词上使用 SDAE 方法提取商品特征。对于 SDAE,我们使用 4 层 DAE,每层的输出维度分别是 1000、100、20 和 5,每层的降噪参数都是 0.95。我们选择 Adam 方法训练文本编码模型,使用随机梯度下降方法训练矩阵分解模型。

4.4 评分预测结果

同其他相关研究工作一致,我们使用预测评分与测试集真实评分的平方误差 MSE 作为评价指标来衡量模型的效果。

² <http://www.mymedialite.net/>

³ <http://deeplearning.net/software/theano/>

我们在相同的训练集/测试集切分上运行 SELFM 模型和 SVD++模型，我们在另一份相同的训练集/测试集切分上运行 SELFM 模型和 LFM、LDA 初始化的 LFM(简称 LDA-LFM)模型、BoWLF 模型，在测试集上的评分预测性能见表 2。

从表 2 可以发现，与 SVD++相比，SELFM 在 24 个子数据集的 23 个数据集上均有提升（仅在“首饰”数据集上性能略有下降），总体上较大程度优于 SVD++模型。其中在鞋类商品上有 64.43%的准确率提升，在服装商品上有 51.24%的提升，提升效果明显，整体上 SELFM 比 SVD++在预测准确率上提升 7.90%。在 17 个数据集上 SELFM 模型的预测结果优于 LFM 模型和 LDA-LFM 模型。相比于 LFM 模型，SELFM 最高有 3.33%的预测准确率提升；相比于 LDA-LFM，SELFM 最高有 1.36%的预测准确率提升；在 20 个数据集上，SELFM 方法的预测准确率都高于 BoWLF 方法，最高提升了 11.76%。根据提出 BoWLF 方法的论文中的实验[9]，与 BoWLF 方法相比，使用了循环神经网络中的长短期记忆网络（Long Short-Term Memory）的 LMLF 方法虽然在自然语言表示层面取得了更好的结果，但是在评分预测层面 LMLF 方法落后于 BoWLF 方法。而我们的 SELFM 模型在预测评分上比 BoWLF 更准确，因此可以合理预计 SELFM 模型能够比 LMLF 方法取得更好的预测准确性。

表 2 SELFM 与 SVD++/LFM/LDA-LFM/BoWLF 的性能 (MSE) 对比

数据集	实验 1			实验 2						
	SVD++	SELFM	相对提升	LFM	LDA-LFM	BoWLF	SDAE	相对提升		
								LFM	LDA-LFM	BoWLF
艺术	1.4663	1.3106	10.62%	1.4649	1.4506	1.4765	1.4500	1.02%	0.05%	1.80%
机动车	1.5732	1.4527	7.66%	1.4116	1.4077	1.4153	1.4120	-0.03%	-0.30%	0.24%
婴儿用品	1.6436	1.5398	6.32%	1.5357	1.5397	1.7655	1.5579	-1.45%	-1.17%	11.76%
化妆品	1.6262	1.3420	17.48%	1.3185	1.2989	1.3223	1.3055	0.98%	-0.51%	1.26%
手机	2.1765	2.1230	2.46%	2.1144	2.1187	2.1654	2.1156	-0.06%	0.15%	2.30%
服装	0.7158	0.3490	51.24%	0.3390	0.3468	-	0.3523	-3.94%	-1.57%	-
电子产品	1.8037	1.7107	5.16%	1.7178	1.7349	1.7176	1.7117	0.35%	1.36%	0.34%
食品	1.5284	1.4405	5.75%	1.4206	1.4046	1.4199	1.3966	1.69%	0.57%	1.64%
健康	1.6438	1.5012	8.68%	1.5147	1.5040	1.5177	1.4995	1.00%	0.30%	1.20%
厨房用具	1.6860	1.5449	8.37%	1.5295	1.5416	1.5391	1.5278	0.11%	0.90%	0.73%
工业	0.4643	0.3771	18.78%	0.3800	0.3828	0.3789	0.3777	0.60%	1.34%	0.31%
首饰	1.3550	1.3945	-2.92%	1.2672	1.2584	1.2744	1.2259	3.26%	2.66%	3.81%
Kindle 商店	1.4362	1.3945	2.90%	1.4309	1.4127	1.4220	1.4101	1.46%	0.18%	0.84%
乐器	1.4945	1.4447	3.33%	1.4310	1.4317	1.4527	1.4529	-1.53%	-1.46%	-0.01%
办公用品	1.7602	1.7054	3.11%	1.6954	1.6567	1.6536	1.6390	3.33%	1.08%	0.88%
庭院	1.7819	1.7045	4.34%	1.6751	1.6647	1.6836	1.6722	0.17%	-0.45%	0.67%
宠物用品	1.6713	1.5842	5.21%	1.5609	1.5681	1.5900	1.5632	-0.15%	0.31%	1.68%
鞋	0.6291	0.2238	64.43%	0.2333	0.2313	-	0.2282	2.17%	1.32%	-
软件	2.3573	2.2249	5.62%	2.3299	2.3117	2.3171	2.3063	1.02%	0.23%	0.47%
体育户外	1.3366	1.1455	14.30%	1.1479	1.1415	1.1504	1.1346	1.16%	0.61%	1.37%
日用工具	1.6068	1.4904	7.24%	1.4981	1.4973	1.5004	1.4959	0.15%	0.10%	0.30%
游戏玩具	1.4530	1.4049	3.31%	1.3958	1.4058	1.4049	1.4315	-2.55%	-1.79%	-1.89%
视频游戏	1.5304	1.4908	2.59%	1.5426	1.5165	1.5286	1.5123	1.96%	0.28%	1.06%
手表	1.5164	1.4915	1.64%	1.4880	1.4756	1.4927	1.4742	0.93%	0.10%	1.24%

即使对于同样的模型 SELFM，根据表 2 和表 3 中的结果，在同一个数据集的不同训练集/数据集的切分上的预测结果也有很大的差异，例如在视频游戏类别上，SELFM 在两个不同切分上的 MSE 是 1.4908 和 1.5165，差别达到了 1.7%。这说明推荐算法的预测准确率对数据集的切分方法十分敏感，因此，从实验中我们可以得出结论，只有在相同的切分上的实验结果之间才具有可比较性。

4.5 模型超参数对模型性能的影响

4.5.1 SDAE 结构的影响

SDAE 的模型压缩能力和 SDAE 的层数有关联。随着层数的增多，SDAE 可以表示的逻辑的复杂性也会增加，因此，我们使用单层 DAE 与 4 层 SDAE 对比，研究 SDAE 的层数对模型预测准确性的影响程度。对于单层 DAE，我们直接将 5000 维的输入向量压缩为商品特征的 5 维向量，与 4 层 SDAE 的输出维数保持相同，其余正则项系数与训练方法、模型超参数均保持一致。在艺术、自动车、婴儿用品、化妆品等 4 个子数据集上的实验结果见表 3 其他数据集上的结果也是类似的。可以看出，四层 SDAE 相对于单层 DAE 来说，推荐准确率有所上升，但是在部分数据集上提升的幅度很小。在四个数据集上的最大提升幅度有 1.26%，最小的提升幅度只有 0.21%。

表 3 SDAE 层数对模型性能 (MSE) 的影响

子数据集	单层 DAE	4 层 SDAE	提升	子数据集	单层 DAE	4 层 SDAE	提升
艺术	1.3274	1.3106	1.26%	婴儿用品	1.5431	1.5398	0.21%
机动车	1.4642	1.4527	0.78%	化妆品	1.3558	1.3420	1.01%

4.5.2 正则项系数的影响

正则项是用于防止协同过滤模型过拟合到训练数据从而导致模型在实际情况下的表现变差。模型过拟合的一种情况是随着训练过程的继续，模型变得更加复杂，导致参数的绝对值偏大，从而使得在实际场景下的准确率下降。为了避免这种现象，我们对协同过滤中的 b_u 、 b_i 和 p_u 引入正则项，作为模型损失函数的一部分。参数越复杂，正则项的数值越大，从而模型的损失也越大，从而阻止参数向复杂化发展。然而正则项的弊端在于如果正则项系数过大，在损失函数中占据主要部分，那么在训练过程中，简化模型的趋势就会大于学习规律的趋势，从而导致模型完全训练失败。为了验证正则项对模型的预测性能的影响，我们在艺术数据集上以多对正则项系数组合训练协同过滤模型，实验结果见表 4。

从表 4 中我们可以发现，正则项系数对预测准确性的影响很小， λ_p 的波动几乎不会对模型结果造成影响。但是， λ_b 过大（等于 1 时）确实会导致因为正则化项的系数过高，从而使得模型学习结果变差。

表 4 正则项系数对模型性能 (MSE) 的影响

λ_b	0.01	0.01	0.01	0.1	0.1	0.1	1	1	1
λ_p	0.01	0.1	1	0.01	0.1	1	0.01	0.1	1
MSE	1.3222	1.3223	1.3228	1.3219	1.3219	1.3227	1.3285	1.3287	1.3287

5 结论与未来工作

本文提出语义增强的隐含因子模型 (Semantics Enhanced Latent Factor Model, SELFM)，该混合推荐模型将深度学习用于个性化推荐系统，并尝试将评论内容结合到协同过滤的算法

之中，作为加强商品特征分析的一种渠道。**SELFM** 模型通过将深度学习方法中的层叠降噪自动解码器(SDAE)与基于矩阵分解的隐含因子模型(LFM)相结合，以此提升推荐模型在预测评分上的准确率。

从运行性能角度上看，**SELFM** 模型，对商品评论的分析模型不需要反复运行，而且不需要每次都从初始状态训练，相反，可以使用微调的方式将新的数据加入模型，因此文本分析模型的平均训练时间是很短的。而基于 **LFM** 的协同过滤模型的训练时间很短，因此能够保证在合理的时间内处理大规模的数据。

与传统的协同过滤模型相比，我们的 **SELFM** 模型通过结合深度学习算法提升了评分预测的准确性，与 **SVD++**、**LFM**、**LDA-LFM**、**BoWLF** 模型相比均有所提升，分别最高提升了 64.43%、3.33%、1.36% 和 11.76%。根据[11]中的实验结果，**BoWLF** 模型在评分预测效果上优于使用了 **LSTM** 的 **LMLF** 模型，所以，我们认为 **SELFM** 模型在评分预测任务上也可能优于 **LMLF** 模型。相比于 **LSTM**，**SDAE** 的训练速度更快，因此，在运行效率上也优于 **LMLF** 模型。

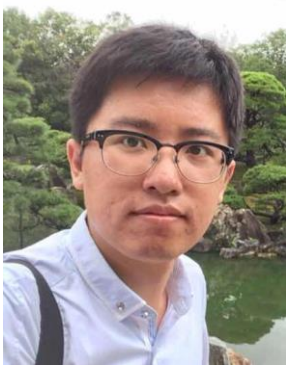
目前我们提出的 **SELFM** 模型在使用 **SDAE** 处理文本时使用了基于词袋的模型，并没有考虑词的相互关系和上下文语义。在未来的工作中，我们将考虑把上下文相关的深度学习算法与协同过滤算法相结合，以提高推荐系统的推荐能力。

参考文献

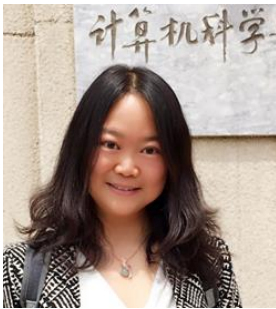
- [1] Resnick P, Iacovou N, Suchak M, et al. An open architecture for collaborative filtering of netnews[C]// Conference on Computer Supported Cooperative Work. 1994.
- [2] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems[J]. IEEE Computer, 2009, 42(8):30-37.
- [3] Deshpande M, Karypis G. Item-based top-N recommendation algorithms. ACM Trans Inf Syst[J]. Acm Transactions on Information Systems, 2004, 22(1):143-177.
- [4] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]// ACM SIGKDD International Conference on Knowledge Discovery and
- [5] Su X, Khoshgoftaar T M. A survey of collaborative filtering techniques[J]. Advances in Artificial Intelligence, 2009, 2009(12). Data Mining, Las Vegas, Nevada, Usa, August. 2008:426-434.
- [6] Melville P, Mooney R J, Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations[C]// Eighteenth National Conference on Artificial Intelligence. 2009:187-192.
- [7] Lee D D. Algorithms for Non-negative Matrix Factorization[J]. Advances in Neural Information Processing Systems, 2015, 13(6):556-562.
- [8] McAuley J, Leskovec J. Hidden factors and hidden topics: understanding rating dimensions with review text[C]// ACM Conference on Recommender Systems. ACM, 2013:165-172.
- [9] Almahairi A, Kastner K, Cho K, et al. Learning Distributed Representations from Reviews for Collaborative Filtering[C]// The, ACM Conference. 2015:147-154.
- [10] Wang C, Blei D M. Collaborative topic modeling for recommending scientific articles[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2011:448-456.
- [11] Sang S L, Chung T, Mcleod D. Dynamic Item Recommendation by Topic Modeling for Social Networks[C]// Eighth International Conference on Information Technology: New Generations, Itng 2011, Las Vegas, Nevada, Usa, 11-13 April. 2011:884-889.
- [12] Bao Y, Fang H, Zhang J. Topiccmf: Simultaneously exploiting ratings and reviews for recommendation[C]// the proceedings of Twenty-Eighth Conference on Artificial Intelligence (AAAI-2014). Québec, Canada. 2014.2-8.
- [13] Blei D M, Ng A Y. Latent Dirichlet Allocation[C]// Journal of Machine Learning Research, 2003, 3(1): 993-1022.
- [14] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach. [J]. IEEE Transactions on Neural Networks, 1997, 8(1):98-113.
- [15] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [Arxiv]. 2015. 4. 10. arXiv:1409.1556. <http://arxiv.org/abs/1409.1556>

- [16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[J]. 2014:1-9.
- [17] Graves A. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-80.
- [18] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[J]. Studies in Computational Intelligence, 2012, 385.
- [19] Sutskever I, Vinyals O, Le Q V. Sequence to Sequence Learning with Neural Networks[J]. Advances in Neural Information Processing Systems, 2014, 4:3104-3112.
- [20] 陈达, 基于深度学习的推荐系统研究[D], 北京邮电大学, 硕士学位论文, 2014.
- [21] 杨宇, 基于深度学习特征的图像推荐系统[D], 电子科技大学, 硕士学位论文, 2015.
- [22] Almahairi A, Kastner K, Cho K, et al. Learning distributed representations from reviews for collaborative filtering. Proceedings of the 9th ACM Conference on Recommender Systems. ACM, 2015. 147-154
- [23] Wang H, Wang N, Yeung D Y. Collaborative Deep Learning for Recommender Systems[J]. Eprint Arxiv, 2014:1235-1244.
- [24] Vincent P, Larochelle H, Lajoie I, et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. [J]. Journal of Machine Learning Research, 2010, 11(6):3371-3408.

作者简介:



丁弼原 (1993—), 男, 清华大学本科, 主要研究领域为个性化推荐。
 Email: thomasbyding@gmail.com;
 Biyuan Ding, born in 1991, Bachelor. His research interest is Personalized Recommendation. Email: thomasbyding@gmail.com



张敏 (通讯作者) (1977—), 女, 清华大学副教授, 主要研究领域为信息检索与挖掘、个性化推荐与用户建模。Email: z-m@tsinghua.edu.cn;
 Min Zhang (Corresponding author), born in 1977, Ph.D., Associate Professor. Her research interests include Web Information retrieval and Data mining, Personalized recommendation and User modeling. Email: z-m@tsinghua.edu.cn



谭云志 (1991—), 男, 清华大学硕士, 主要研究领域为个性化推荐与情感分析。Email: cloudcompute09@gmail.com。
 Yunzhi Tan, born in 1991, Master. His research interests include Personalized Recommendation and Sentiment Analysis. Email: cloudcompute09@gmail.com