

基于网络特征学习融合内容与协同过滤的自适应推荐算法

袁梦祥

安徽大学大数据与云服务工程实验室

摘 要

在推荐系统中有一类方法基于协同过滤,还有一类方法是基于项目内容,前者虽然一般来说效果更好但是无法处理冷启动情况,而后者即使冷启动时还能够工作。为了融合两个算法的优点,本文提出了一种基于网络特征学习的自适应推荐算法 **NRAM**。首先,获取用户的行为数据和项目的标签数据并进行二部图表示,进而,使用网络表示学习的方法获取网络中节点的低维度潜在表示。最后,采用注意力网络,针对不同的用户、不同的项目,学习对应的隐藏权重,自适应的融合协同过滤信息和项目内容信息所占的权重。我们在 **GoodBooks** 数据集上对此方法进行了大量实验,结果表明与经典的 **CF** 推荐算法相比,我们的算法有更好的推荐效果。

关键词: 推荐系统; 协同过滤; 二部图网络; 网络表示学习; 矩阵分解

1 引言

随着互联网的不断发展,信息爆炸已经成为人们面临的一个巨大的挑战,而这正是推荐系统主要解决的问题,使用推荐系统可以帮助我们获得更有效的信息。传统的推荐算法有基于内容 (**CB**) 和基于协同过滤 (**CF**)。它们在学术界被广泛讨论和应用,许多相关的研究都是基于这两种算法。它们分别根据两种不同的信息进行推荐:项目的内容信息和用户的历史反馈信息。

基于内容的推荐方法正是基于内容信息的,内容信息主要有两种类型,即,用户偏好信息,例如用户的年龄、性别、职业等,以及项目的内容信息,例如衣服的颜色、食物的味道、音乐的旋律等。此外,文本评论等其他特定信息可以同时作为用户和项目的内容信息。使用基于内容的推荐方法,即使没有用户的历史反馈信息也可以给用户进行推荐,而且基于内容的推荐算法具备很好的可解释性 [1] 人们可以清楚看到。

目前应用最广泛的推荐算法是协同过滤的推荐算法,学术界有许多关于该算法的研究 [2, 3, 4, 5]。在实际应用中用户的行为数据通常非常稀疏,矩阵分解是最常用的基于模型的协同过滤算法 [6, 7],该算法通过用户-项目二部图网络,同时学习用户和项目的潜在表示。传统的矩阵分解在嵌入表示顶点时,使用线性变换将网络顶点嵌入到低维空间。最近的研究趋势是使用深度学习的方法学习网络上的顶点嵌入 [8] 进行推荐。在为推荐而构建的用户-项目二部网络中,虽然边只存在于不同类型的顶点之间,但同一类型的顶点之间存在隐式关系,例如在用户之间就存在一种隐式的关系,这种关系表明用户对同一项目的偏好。最近有论文指出,对这种隐式关系进行建模可以提高推荐性能 [9]。

然而，现有的网络表示学习方法 [10, 11, 12] 仅仅对二部图中的显式关系进行了建模而忽略潜在的隐式关系。与现有的协同过滤模型用于推荐的工作不同，我们结合网络表示学习的方法，在嵌入表示顶点时充分考虑二部图网络结构的特点。最后在生成推荐时，不仅考虑用户的行为信息，同时利用到项目的内容信息，可以比以前的方法产生更好的预测。

总之，本文的主要贡献有两方面：

- 在嵌入表示顶点时充分考虑二部图网络结构的特点。通过项目用户/项目二部图数据学习用户顶点的嵌入表示，通过项目/标签二部图数据学习项目顶点的嵌入表示，在低维嵌入空间中考虑用户行为和项目内容的领域信息；
- 提出了一种改进的预测方法，该方法基于矩阵分解技术与网络表示学习技术，联合训练用户的行为和项目内容的领域信息，对原始评分矩阵进行填充，为用户产生基于预测评分的推荐。

本文的其余部分安排如下。第二节介绍了相关的工作；第三节详细介绍了推荐算法的整体架构；第四节介绍了实验；最后，第五节总结论文并讨论论文的未来工作。

2 相关工作

我们的工作涉及到协同过滤和基于网络表示学习的推荐。因此，在本节中，我们将简要回顾这些领域的相关工作。

2.1 协同过滤

基于用户的历史行为数据对用户产生推荐的方法主要是基于协同过滤思想 [5]，矩阵分解是实现协同过滤最常用的方法，它可以很好的解决冷启动问题 [?]。但基本的矩阵分解模型，如 [6, 7]，完全通过用户-项目的二部图数据，学习用户和项目的潜在表示，使用用户-项目潜在特性的点乘操作预测用户对项目的评分。[13] 将邻域信息集成到矩阵分解中。它假定用户对某一项的评价不仅由用户对该项的潜在表示决定，而且还由用户对其他项的评价行为构成，即考虑物品的邻域信息。这种方法在许多领域的性能都优于传统的矩阵分解模型，然而这种算法在建模时使用线性变换将网络顶点嵌入到低维空间，没法捕捉用户与项目的非线性关系，而且仅仅考虑了用户的行为信息，没有考虑项目的内容信息。

2.2 基于网络表示学习的推荐

在推荐系统中，用户和项目形成一个二部网络，边包含丰富的协同过滤模式的用户评价行为 [14]。基于网络表示的推荐算法利用网络的结构信息，使用非线性变换将网络顶点嵌入到低维空间，学习网络中顶点的低维表示进行推荐 [15, 16]，可以捕捉用户与项目的非线性关系。开创性的 DeepWalk [10] 和 Node2vec [11] 算法对同质网络进行建模，有一些后续工作是利用同质顶点之间的高阶邻近来嵌入同构网络，例如 LINE [12] 学习了一阶和二阶关系的两个分离嵌入。这类算法的基本思想是通过随机游走将网络转换为顶点序列的语料库。尽管这种方式具备有效性和普遍性，但 [17]

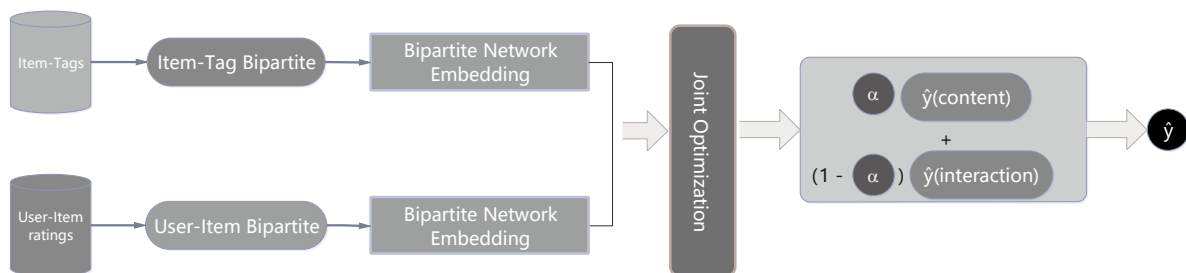


图 1: recommendation framework

指出这些方法忽略了二部图网络的特殊性质，对于嵌入表示二部图网络可能不是最理想的。而且现有的网络表示学习的工作主要集中在嵌入表示同质网络，网络中的顶点都是相同类型的 [11, 10?]，仅仅是对二部网络中的显式关系进行建模，而且在产生推荐时没有考虑项目的内容信息。

3 推荐模型

在本节中，我们将介绍我们的推荐模型，这是一种基于二部图嵌入表示和协同过滤的推荐模型。我们的模型通过利用两个异构信息源，分别是项目的标签信息和用户的评分信息，来预测用户对项目的评分。在下文中，我们首先详细介绍了我们的推荐框架，然后是分别介绍数据的预处理过程，二部图的嵌入表示算法，以及在分别获得用户和项目的嵌入表示后，怎么进行联合训练，获得最终的预测评分矩阵。

3.1 算法框架

推荐系统是根据用户过去的历史行为，为用户提供个性化的推荐服务。因此，在我们的推荐系统中需要使用到用户的历史评分信息，其次为了获得项目的内容信息，我们需要使用项目/标签数据。

图1描述了我们所提出方法的基本思想。首先，对收集来的用户-项目评分文件进行预处理，获得用户-项目的二部图表示，然后使用二部图网络嵌入算法进行获得用户的嵌入表示。接下来，同样处理项目-标签文件以获得项目的嵌入表示，最后使用改进后的矩阵分解模型，联合训练两个向量，使用预测评分填充评分矩阵，对用户进行个性化推荐

3.2 数据预处理

从原始的文本数据到可以进行二部图嵌入表示的二部图数据需要经过数据的预处理过程，我们的算法使用到了 user-item 的评分数据和 item-tag 内容数据。针对用户的评分数据，我们很容易的可以使用二部图进行表示。如图2，左边的图为用户的评分矩阵图，右边的图即为对应的用户物品二部图网络，其中圆形节点分别代表用户和项目，节点之间的边代表用户对物品存在打分行为，边

的权重代表对应的用户项目评分。类似的，对于项目标签数据，我们可以把项目和 tag 当作网络中的节点。

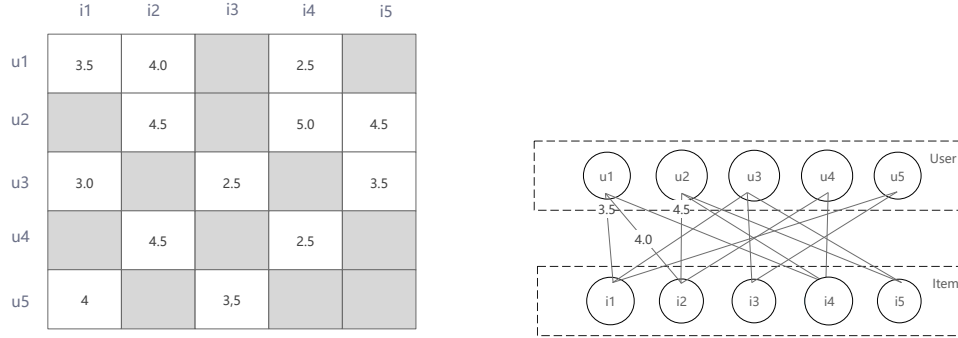


图 2: An example of the bipartite network structure

3.3 二部图嵌入表示

Follow Ming Gao 的工作 [17]，我们使用 BiNE 的算法对二部图网络进行嵌入表示。在本小节我们会对 BiNE 方法进行简要的概述。

良好的网络嵌入应该能够很好地重构原始网络。为了在二部网络实现这一目标，我们考虑从两个特征来重构二部网络，分别是由直接相连的边所代表的显式关系，和由未直接相连的传递链隐含的隐式关系，通过联合优化这两个目标函数来学习二部图顶点的嵌入：

$$\text{maximize } L = \alpha \log O_2 + \beta \log O_3 - \gamma O_1 \quad (1)$$

公式1中 O_1 代表对二部网络中的显示关系进行建模得到的最小化目标函数。 O_2 和 O_3 代表对二部网络中的隐式关系进行建模得到的最大化目标函数，参数 α, β, γ 三个参数分别代表了联合训练时各部分的比例。最后，为了优化联合模型，使用随机梯度上升算法 (SGA) 对模型进行训练，获得最终的嵌入向量表示。

在本论文中，我们采取 [17] 中的默认配置，分别对用户-项目二部图数据和项目-标签二部图数据进行嵌入表示。

3.4 联合训练

矩阵分解是一种广泛采用的基于模型的协同过滤方法。它将用户项目评分矩阵分解为低秩矩阵以进行预测。这种因式分解的思想是同时获得用户和项目的低维潜在表示，使用用户-项目潜在特性的点乘操作预测用户对项目的评分。在本文中，为了同时保留用户和项目嵌入表示后的领域信息，我们基于矩阵分解模型，形成一个联合优化框架。

在本文中，我们定义一个评级矩阵来表示用户对项目的评分，并使用矩阵分解方法通过矩阵点乘来近似评分矩阵。我们的目标函数如下：

$$\min_{U,V} \mathcal{L}(R, U, V) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \left(R_{ij} - U_i^T V_j \right)^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2 \quad (2)$$

上述方法是传统的协同过滤（CF）方法，其使用所有可用的评分值来预测评分矩阵中的缺失值。然而，由于真实的评分矩阵非常稀疏，在大多数情况下传统的矩阵分解不能生成最佳的预测值。因此，在本文中，我们利用嵌入表示，结合用户和项目在嵌入表示空间的领域信息，本论文提出了一种高性能方法，以最小化误差平方和二次正则化项，其公式为：

$$\min I(R, U, V, S_U, S_V) = \frac{1}{2} I(R - (\alpha U^T V S_V + (1 - \alpha)(S_U U^T V)))^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2 \quad (3)$$

公式3中 S_U, S_V 代表的分别是用户和项目的相似矩阵，表示在嵌入表示空间中用户和项目的领域信息。在考虑用户和项目的相似度时，使用如下公式：

$$S_{ik} = \frac{\text{COS}(i, k)}{\sum_{k \in T(i)} \text{COS}(i, k)} \quad (4)$$

4 实验

在本节中，我们使用 MovieLens 数据集和 GoodBooks 数据集进行实验，比较了我们的推荐算法和其他推荐算法的实验效果，最后还分析了不同的参数设置对模型的影响。

4.1 实验设置

为了衡量推荐算法的有效性，我们使用了 MovieLens 和 GoodBooks 两个数据集。MovieLens 数据集被广泛用于评价电影推荐系统 [8]，在构建用户-项目评分二部图时，边的权重表示用户对某一项的评分。类似地，GoodBooksS 数据集包含书籍的评分信息。表1总结了我们的实验数据集的统计数据。注意：为了研究用户自定义标签和项目内容标签的区别，在两个数据集中分别使用不同类型的项目标签数据。

4.2 性能评估

为评价所提出的 NEMF 方法的预测精度，选择以下经典方法在 MAE 和 RMSE 两个指标上进行对比实验。

- UMEAN: 该方法利用用户对项目的平均值来预测用户未评分的数据。
- IMEAN: 该方法利用项目的平均得分来预测项目未评分的数据。

表 1: Statistics of Datasets

Name	GoodBooks	MovieLens
U	9000	610
V	10,000	9,742
E	136,203	80,419
Desity	0.15%	1.30%

表 2: Prediction performance on MovieLens and GoodBooks

Algorithm	MovieLens		GoodBooks	
	RMSE	MAE	RMSE	MAE
UMean	0.9360	0.7277	1.1678	0.8241
Imean	1.1536	0.8356	0.9625	0.7642
UPCC	0.9335	0.7250	1.1657	0.8209
IPCC	1.1425	0.8241	0.9559	0.7584
MF	1.4479	1.0255	1.3978	0.9972
NIMF	1.1660	0.8407	1.5321	1.1113
NEMF	0.8642	0.6622	0.8668	0.6780

- UPCC: 该方法基于相似用户的历史评分来预测。
- IPCC: 基于项目的协同过滤模型。
- UIPCC: 该方法是分别给与 UPCC 和 IPCC 一个权重, 用综合的方法去预测缺失的值
- MF: 该方法是传统的矩阵分解模型。
- NIMF: 该方法是融合领域信息的矩阵分解模型

在我们的实验中, 我们将用户行为数据按照用户均匀分布, 选择每个用户评分数据的 80% 作为训练集, 剩下的 20% 作为测试卷, 保证每个用户都拥有评分数据。然后我们将我们的方法与上述五种方法在同一训练数据集上进行训练, 再在同一个测试集进行比较。我们方法的参数设置为: $\alpha = 0.2$ (MovieLens 为 0.9), $\text{threshold}=0$, 矩阵的维度 = 20. 比较结果在表 2 中。这些参数设置的原因将在下一节中详细解释。

从表 2 可以看出, 我们提出的方法的 MAE 和 RMSE 明显小于其他方法的 MAE 和 RMSE。这表明所提出方法的预测精度更高。我们提出的方法显示出优于其他方法的预测性能。这证明了通过利用用户和项目的嵌入表示, 可以实现更好的预测性能。实验结果验证了可以使用嵌入表示会结合用户和项目的领域信息来准确预测缺少的评分值。

4.3 模型分析

我们现在研究不同模型设置对模型的影响。

参数 α 值的影响。研究 α 值对模型的影响。结果如图3

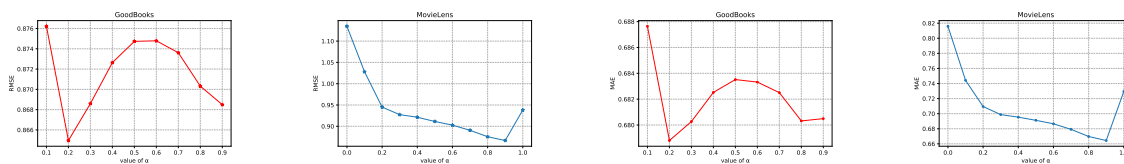


图 3: the inspect of Value α

阈值的影响。研究阈值对模型的影响。结果如图4

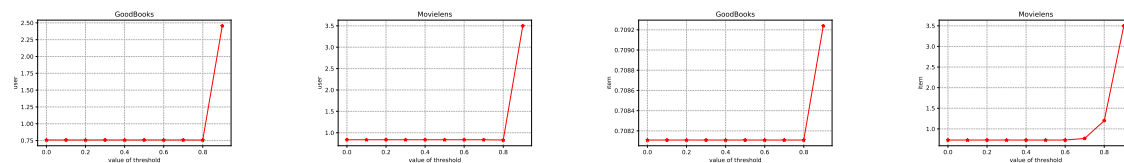


图 4: the inspect of Value Density

联合训练时矩阵维度的影响。联合训练时矩阵维度对模型的影响。结果如图5

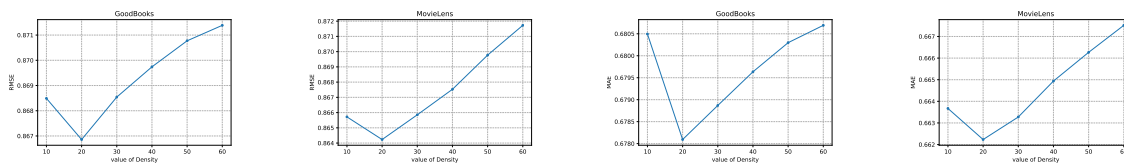


图 5: the inspect of Value Density

5 结论

结论部分，介绍论文的实验结果

参考文献

- [1] Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: Explaining Recommendations Using Tags. *Proceedings of the 13th international conference on Intelligent user interfaces - IUI '09*, 2008.
- [2] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, jan 2003.

- [3] Catarina Miranda and Alípio Mário Jorge. Item-based and user-based incremental collaborative filtering for web recommendations. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5816 LNAI, pages 673–684, 2009.
- [4] Badrul Sarwar, George Karypis, Joseph Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th . . .*, 2001.
- [5] Xiaoyuan Su and Taghi M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009.
- [6] R Salakhutdinov and A Mnih. Probabilistic Matrix Factorization. In *Proc. Advances in Neural Information Processing Systems 2007 (NIPS 07)*, 2007.
- [7] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, aug 2009.
- [8] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural Collaborative Filtering. *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, pages 173–182, 2017.
- [9] Lu Yu, Chuxu Zhang, Shichao Pei, Guolei Sun, and Xiangliang Zhang. WalkRanker: A Unified Pairwise Ranking Model With Multiple Relations for Item Recommendation. *AAAI*, 2018.
- [10] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*, pages 701–710, New York, New York, USA, 2014. ACM Press.
- [11] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, pages 855–864, New York, New York, USA, 2016. ACM Press.
- [12] Marina Bährle-Rapp and Marina Bährle-Rapp. Line: Largescale information network embedding. In *Springer Lexikon Kosmetik und Körperpflege*, pages 323–323. ACM Press, New York, New York, USA, 2010.
- [13] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008.
- [14] Xiangnan He, Ming Gao, Min Yen Kan, and Dingxian Wang. BiRank: Towards Ranking on Bipartite Graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2017.
- [15] Suchit Pongnumkul and Kazuyuki Motohashi. A bipartite fitness model for online music streaming services. *Physica A: Statistical Mechanics and its Applications*, 2018.
- [16] Jie Liu, Mingsheng Shang, and Duanbing Chen. Personal recommendation based on weighted bipartite networks. In *6th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2009*, volume 5, pages 134–137. IEEE, 2009.
- [17] Ming Gao, Leihui Chen, Xiangnan He, and Aoying Zhou. BiNE: Bipartite Network Embedding. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18*, pages 715–724, New York, New York, USA, 2018. ACM Press.