

A Network Embedding Augmented Collaborative Filtering Model For Recommendation

袁梦祥

安徽大学大数据与云服务工程实验室

摘 要

使用推荐系统为用户提供个性化的推荐服务,有助于提高用户的满意度,更好的发掘物品的“长尾”。推荐算法是推荐系统的核心,基于协同过滤的推荐算法是目前应用比较广泛的推荐算法,但传统的协同过滤算法在推荐时存在数据稀疏量大,数据维数高,无法利用内容信息等问题。为了解决协同过滤算法的缺点,本文引入了网络表示学习的方法,使用二部图网络来表示用户的显式行为数据和项目的内容数据,再通过针对二部图结构设计的网络表示学习方法,学习网络中节点的低维度潜在表示,以获得用户和项目的领域信息。为了结合用户和项目的领域信息对用户进行推荐,我们进一步提出了一种融合领域信息的矩阵分解技术。本文结合协同过滤和嵌入算法的优点,提出了一种新的二部图推荐系统,该系统充分考虑了不同项目的内容特征和用户的行为偏好。我们在 GoodBooks 和 Movielens 数据集上对此方法进行了实验验证,结果表明,与其他推荐算法相比,基于网络表示学习的加强协同过滤算法有更好的推荐效果。

关键词: 推荐系统, 二部图网络, 网络表示学习, 矩阵分解, 协同过滤

1 介绍

在当今竞争激烈的市场环境下,产品的个性化程度已经成为影响顾客产品选择和满意度的重要因素。如果要给用户提供个性化的商品或服务,就必须充分研究用户的兴趣,而这正是推荐系统主要解决的问题,通过挖掘用户的历史行为数据,推荐系统可以自动发现用户的个性化需求。

推荐系统的本质是通过一定的方式将用户和项目联系起来,研究怎么将用户兴趣和项目关联起来的推荐算法是整个推荐系统的核心。GroupLens[4] 提出可以使用项目的内容信息将用户和物品联系起来,基于内容的推荐算法 [1, 2, 3] 是常用的推荐算法,该算法通过分析用户对内容的喜好和物品的内容属性来进行推荐,结果非常直观且便于理解,该算法通常用来处理文本信息 [5],处理非文本信息的难度比较高,比如音乐、图像等。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 [6, 7, 8, 9]。该算法的核心是相似度计算, Badrul Sarwar 等在论文里做了详细的研究 [10]。传统的相似度计算依赖用户的行为,然而在实际应用中,用户的行为矩阵通常比较稀疏,且容易受项目流行度影响,这导致相似度的计算结果并不稳定。基于模型的协同过

滤是目前最主流的协同过滤类型，针对显式的评分数据集，矩阵分解算法 [11, 12, 13] 是实现协同过滤算法中最常用的模型，该算法能够很好的解决冷启动问题 [14]。

二部网络是一种普遍存在的数据结构，用于对两类实体之间的关系进行建模。它在推荐系统、搜索引擎、问答系统等方面得到了广泛的应用。在推荐系统中，用户和项目形成一个二部网络，边可以包含丰富的协同过滤模式的用户评价行为 [15]。矩阵分解算法的本质，是基于用户和项目构成的二部图，学习到用户和项目顶点的低维表示，最近的研究趋势是使用深度学习的方式来学习用户和项目的低维度潜在表示 [16]。但上述方法仅仅是对二部网络中的显式关系进行建模，[17, 18] 指出，通过考虑二部图中的隐式关系，可以很好的改进推荐的效果。

迄今为止，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 [19, 20, 21]。根据 DeepWalk[20] 的开创性工作，这些方法通常采用两步解决方案：首先在网络上执行随机游走以获得顶点的“语料库”，然后应用词嵌入方法，如 word2vec[22] 来获取顶点的嵌入表示。尽管这种方式具备有效性和普遍性，但 Gao, Ming[23] 认为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。

与现有的网络表示学习用于推荐的工作不同 [24]，我们的工作充分考虑了二部图网络结构的特点和推荐系统中领域存在的长尾分布现象，对顶点进行嵌入表示。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，使用考虑领域信息的矩阵分解模型给用户产生基于预测评分的推荐。

总之，本文的主要贡献有三方面：

- 本论文使用二部图来表示项目的标签数据，学习网络中项目顶点的嵌入表示，利用项目的内容信息，不需要大量的用户行为数据就能获得很好的推荐效果。
- 本论文采取 BiNE[23] 的方式，嵌入表示用户项目二部图和项目标签二部图，同时考虑二部图的结构特点和推荐系统存在的长尾分布现象，分别学习到用户和项目的低维度向量表示。
- 本论文在嵌入向量空间中比较用户之间和物品之间的相似度，基于考虑领域信息的矩阵分解模型，联合训练用户的行为信息和项目的内容信息，对原始评分矩阵进行填充，给用户产生基于预测评分的推荐。一个好的推荐系统不仅要能产生好的推荐效果，还要能表现出用户对物品感兴趣的程度。

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模型的整体架构；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、网络表示学习和基于二部图网络的推荐。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 [9]。矩阵分解是实现协同过滤最常用的方法。基本的矩阵分解模型，如 [11, 12]，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，仅仅考虑到二部图的显式关系，没法利用二部图顶点的隐式关系，此外，这种方法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 网络表示学习

网络嵌入方法（Network Embedding）旨在学习网络中节点的低维度潜在表示，所学习到的特征表示可以用作基于图的各种任务的特征，例如分类，聚类，链路预测和可视化。基于神经网络的方法是当前最先进的顶点表示学习技术。开创性的 DeepWalk[20] 和 Node2vec[19] 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 [25] 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。然而，对同质网络的处理不能有效地保持二部图网络中的显式和隐式关系。而且随机生成的语料不符合二部图网络中顶点长尾分布的特点，传统的网络表示算法无法针对二部图网络的进行更好的建模。

2.3 基于网络结构的推荐

基于网络表示的推荐算法将用户与物品抽象为网络中的节点，用户与物品的行为关系保存在连边中，利用网络的结构信息进行推荐 [18, 26]。与基于网络结构的排序任务不同，基于网络结构的推荐目的是学习顶点的低维表示，这个低维表示能保存更多的信息，而不仅仅是一个排序的得分。传统的基于网络结构的推荐算法没法利用物品的内容信息，而且仅仅只能对二部图中的显式关系进行建模，没有考虑二部图中的隐式关系。

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，来估计用户-项目对的评分。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。

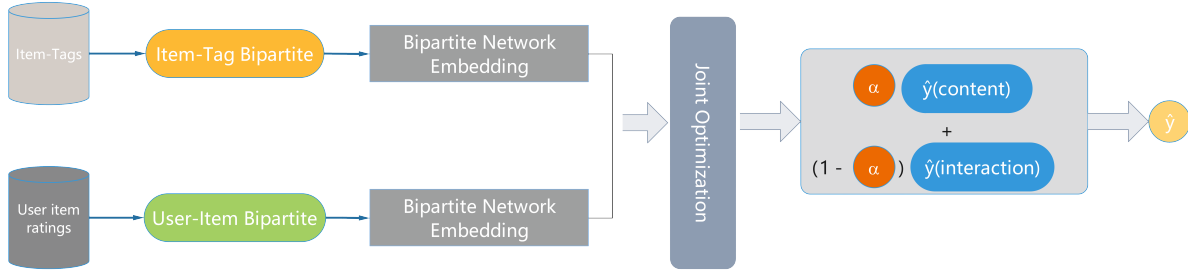


图 1: recommendation framework

3.1 推荐架构

介绍算法的架构如图1。

3.2 数据预处理

这里介绍对数据进行的预处理，将 user-item 评分数据和 item-tag 内容数据表示成二部图。

3.3 二部图嵌入表示

这里介绍二部图的嵌入表示算法 BiNE。

3.4 联合训练

这里介绍模型中联合训练用户和项目嵌入表示的算法。

$$\mathcal{L}(R, S_u, S_i, U, V) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \left(R_{ij} - \left(\alpha U_i^T V_j + (1 - \alpha) \sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j \right) \right)^2 + \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2, \quad (4)$$

4 实验

在本节中，我们使用 MovieLens 数据集和 GoodBooks 数据集进行实验，比较了我们的模型和其他传统模型的实验效果，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

使用 MovieLens 数据集和 GoodBooks 数据集。

4.2 性能评估

这里比较

表 1: Prediction performance on MovieLens and GoodBooks

| Algorithm | MovieLens | | GoodBooks | |
|-------------|---------------|---------------|---------------|---------------|
| | RMSE | MAE | RMSE | MAE |
| User-CF | 0.9335 | 0.725 | 1.1657 | 0.8209 |
| Item-CF | 1.1425 | 0.8241 | 0.9559 | 0.7584 |
| MF | 1.4479 | 1.0255 | 1.3978 | 0.9972 |
| User-NEMF | 1.1534 | 0.8354 | 0.9529 | 0.7587 |
| Item-NEMF | 0.9381 | 0.7294 | 0.9018 | 0.7081 |
| NEMF | 0.8657 | 0.6636 | 0.8684 | 0.6804 |

4.3 模型分析

我们现在研究不同模型设置对模型的影响。

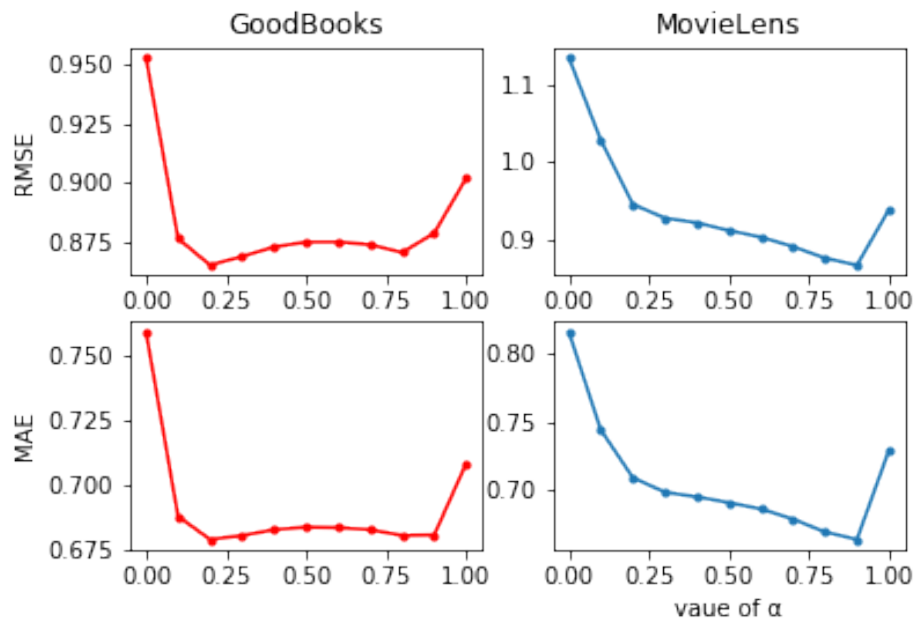


图 2: the inspect of Value α

参数 α 值的影响。研究 α 值对模型的影响。结果如图2

阈值的影响。研究阈值对模型的影响。

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。

二部图网络嵌入模型的影响。研究二部图网络嵌入模型中的参数对模型的影响。

5 结论

结论部分，介绍论文的实验结果

参考文献

- [1] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, mar 2002.
- [2] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. Content-based recommendation. In *Recommender Systems*, pages 51–80. 2013.
- [3] Robin Van Meteren and Maarten Van Someren. Using Content-Based Filtering for Recommendation. In *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop*, pages 47–56, 2000.
- [4] 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.
- [5] Chandra Bhagavatula, Sergey Feldman, Russell Power, and Waleed Ammar. Content-Based Citation Recommendation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 238–251, 2018.
- [6] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, jan 2003.
- [7] 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.
- [8] Badrul Sarwar, George Karypis, Joseph Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th . . .*, 2001.
- [9] Xiaoyuan Su and Taghi M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009.
- [10] Xiao Peng Yu. As-index based collaborative filtering recommendation algorithm. In *Proceedings of the 2009 International Conference on Machine Learning and Cybernetics*, volume 3, pages 1570–1576, New York, New York, USA, 2009. ACM Press.
- [11] R Salakhutdinov and A Mnih. Probabilistic Matrix Factorization. In *Proc. Advances in Neural Information Processing Systems 2007 (NIPS 07)*, 2007.
- [12] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, aug 2009.
- [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] Tian Qiu, Guang Chen, Zi Ke Zhang, and Tao Zhou. An item-oriented recommendation algorithm on cold-start problem. *EPL*, 95(5):58003, sep 2011.
- [15] Xiangnan He, Ming Gao, Min Yen Kan, and Dingxian Wang. BiRank: Towards Ranking on Bipartite Graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2017.

- [16] 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.
- [17] 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.
- [18] Suchit Pongnumkul and Kazuyuki Motohashi. A bipartite fitness model for online music streaming services. *Physica A: Statistical Mechanics and its Applications*, 2018.
- [19] 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.
- [20] 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.
- [21] Lizi Liao, Xiangnan He, Hanwang Zhang, and Tat Seng Chua. Attributed Social Network Embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2018.
- [22] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. pages 3111–3119, 2013.
- [23] 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.
- [24] Oren Barkan and Noam Koenigstein. ITEM2VEC: Neural item embedding for collaborative filtering. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, sep 2016.
- [25] 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.
- [26] 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.