

A Bipartite Network Based Recommendation System via Network Embedding Augmented Collaborative Filtering Model

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,因此研究怎么将用户和用户可能感兴趣的项目关联起来的推荐算法是整个推荐系统的核心。协同过滤算法是推荐系统中的常用算法 [1, 2, 3]。已经证明协同过滤算法拥有很好的推荐效果,协同过滤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 [4]。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,相似度的计算结果不稳定,存在冷启动问题。基于内容的推荐

与现有的工作不同,本文使用网络表示学习的方法,将项目 and 用户映射到低维向量空间。使用二部图网络来表示数据,不仅考虑项目或者用户之间的显性关系,同时考虑项目和用户之间的隐性关系,最后该系统集成了矩阵分解模型,一个好的推荐系统不仅能给用户推荐,也能反应出用户喜欢的程度,结合矩阵分解模型,可以给用户的推荐产生,同时解决冷启动问题 [5]。

总之，本文的主要贡献有三方面：

- 本论文采取表示学习的方式，学习到用户和行为的低维度向量表示，再结合矩阵分解和协同过滤的方法，实现评分预测和用户个性化推荐。
- 使用二部图表示用户的行为数据和物品的内容信息。然后，通过网络表示学习的算法，将用户根据行为数据，物品根据内容信息分别映射到不同的嵌入向量空间。最后在嵌入向量空间中比较用户之间和物品之间的相似度，利用协同过滤的思想对原始评分矩阵进行填充。
- 不仅能给用户推荐，也能反应出喜欢的程度。

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、基于二部图网络的推荐和基于网络表示学习的二部图。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

基于用户的历史交互行为，对用户产生推荐的方法主要是基于协同过滤 (CF) 思想的 [3]，其目的是用静态的潜在特征向量表示用户和条目。矩阵分解 (MF) 是实现协同过滤最常用的方法。基本的 MF 模型，如 [13,14] 尝试学习用户和项目的潜在功能完全通过匹配用户-项目的交互 (即，二进制指标或用户-项目评分) 矩阵与点积 (DP) 操作。然后，通过对给定的用户-项目对使用派生的用户/项目潜在特性的 DP 操作，计算评级预测。大量的工作试图通过基于用户-项目交互的建模来提高 MF 的性能。例如，[14] 在 MF 中引入了用户和项目偏差。[15] 将邻域建模集成到 MF 中。它假定用户对某一项的评价不仅由用户对该项的潜在特征构成，而且还由用户对其他项的评价行为构成。这些方法在许多领域的性能都优于传统的 MF 模型。然而，所有这些基于 MF 的模型都使用 DP 操作作为其评级预测器。DP 操作的一个固有限制是潜在特征相互独立。也就是说，DP 只支持潜在特征的线性组合，而不考虑高阶特征交互。实验结果表明，现有的基于 MF 的方法性能较差。

2.2 网络表示学习

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。例如，在搜索引擎中，查询和网页形成一个二部图网络，边缘可以指示用户点击行为，提供有价值的关联信号 [1,2]；在推荐系统的另一个应用中，用户和项目形成一个二部网络，边缘可以编码包含丰富的协同过滤模式 [3] 的用户评价行为。要对网络数据执行预测分析，首先获得表示形式 (即，特征向量) 表示顶点。传统的向量空间方法，如袋装词表示方法，在处理大规模动态网络的实际应用中，语义捕获太少，效率低下。近年来，数据挖掘和信息检索领域的研究主要集中在从数据中学习表示 [4-7]。特别是，它们将顶点嵌入到低维空间

中，即，将顶点表示为可学习的嵌入向量。基于顶点嵌入的标准机器学习技术可以应用于各种预测任务，如顶点标记、链接预测、聚类等。

1. 协同过滤：介绍传统的协同过滤算法
2. LFM 矩阵分解：使用机器学习的方式解决推荐问题
3. item2Vec：更进一步使用深度学习的方式解决推荐问题

3 推荐模型

使用图嵌入的方式，将用户-物品的二部图表示转化为低维度的向量，在低维度向量空间中，考虑相似性，最后结合矩阵分解和协同过滤模型，通过机器学习的方式填充评分矩阵，完成个性化推荐。

融合领域信息的矩阵分解的 loss 函数如下：

3.1 算法

这里介绍模型使用的算法

Algorithm 1 Conjugate Gradient Algorithm with Dynamic Step-Size Control

Input: $f(x)$: objective function; x_0 : initial solution; s : step size;

Output: optimal x^*

- 1: initial $g_0 = 0$ and $d_0 = 0$;
 - 2: **repeat**
 - 3: compute gradient directions $g_k = \nabla f(x_k)$;
 - 4: compute Polak-Ribiere parameter $\beta_k = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}$;
 - 5: compute the conjugate directions $d_k = -g_k + \beta_k d_{k-1}$;
 - 6: compute the step size $\alpha_k = s / \|d_k\|_2$;
 - 7: **until** ($f(x_k) > f(x_{k-1})$)
-

3.2 算法模型图

介绍算法的模型

4 实验

实验部分，介绍论文的实验

我强烈建议你使用 `booktabs` 宏包，这个宏包有三个命令 `\toprule`、`\midrule` 和 `\bottomrule` 能方便你制作三线表。图1 是一个示例：

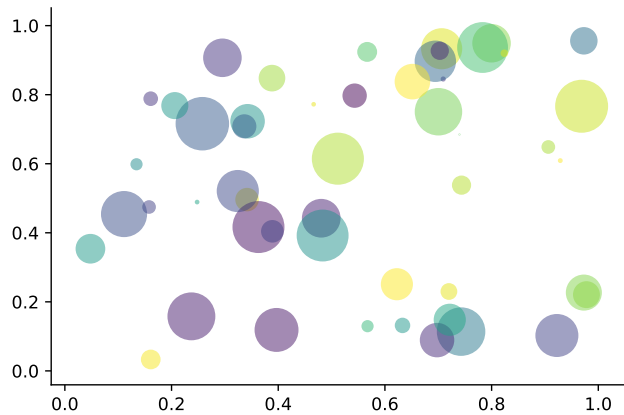


图 1: Scatter Plot Example

5 结论和未来工作

结论部分，介绍论文的实验结果

代表显示所有的论文文献包括文中没有引用的

参考文献

- [1] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, jan 2003.
- [2] 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.
- [3] Xiaoyuan Su and Taghi M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. pages 3111–3119, 2013.
- [8] 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.

- [9] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, mar 2002.
- [10] Walter Carrer-Neto, María Luisa Hernández-Alcaraz, Rafael Valencia-García, and Francisco García-Sánchez. Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, 39(12):10990–11000, apr 2012.
- [11] Libing Wu, Cong Quan, Chenliang Li, Qian Wang, and Bolong Zheng. A Context-Aware User-Item Representation Learning for Item Recommendation. 2017.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Badrul Sarwar, George Karypis, Joseph Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th . . .*, 2001.
- [18] Ximeng Wang, Yun Liu, Guangquan Zhang, Yi Zhang, Hongshu Chen, and Jie Lu. Mixed Similarity Diffusion for Recommendation on Bipartite Networks. *IEEE Access*, 5:21029–21038, 2017.
- [19] 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.
- [20] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, aug 2009.
- [21] 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.
- [22] Suchit Pongnumkul and Kazuyuki Motohashi. A bipartite fitness model for online music streaming services. *Physica A: Statistical Mechanics and its Applications*, 2018.
- [23] R Salakhutdinov and A Mnih. Probabilistic Matrix Factorization. In *Proc. Advances in Neural Information Processing Systems 2007 (NIPS 07)*, 2007.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] Koren Y, Bell R, and Volinsky C. Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8):30–37, 2009.
- [28] Yong Zheng. Matrix Factorization In Recommender Systems. *Journal of Physics A: Mathematical and Theoretical*, 44(8):085201, 2011.
- [29] Manolis Vozalis and Konstantinos G. Margaritis. On the combination of user-based and item-based collaborative filtering. *International Journal of Computer Mathematics*, 81(9):1077–1096, 2004.