**北 京 邮 电 大 学**

**本科毕业设计（论文）开题报告**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 学院 | 计算机学院 | 专业 | | 计算机科学与技术 | | 班级 | 2015211307 |
| 学生姓名 | 王睿嘉 | 学号 | | 2015211906 | | 班内序号 | 31 |
| 指导教师姓名 | 石川 | 所在单位 | | 计算机学院 | | 职称 | 教授 |
| 设计（论文）题目 | （中文）异质信息网络中的相似性推荐算法研究 | | | | | | |
| （英文）Research on Similarity Recommendation Algorithm in Heterogeneous Information Network | | | | | | |
| **一、 选题背景和意义**  推荐是解决信息过载的有效方法，被广泛应用于电子商务和互联网服务中。而融合更多信息进行混合推荐是解决数据稀疏性的有效技术。  近年来，大量涌现的社会媒体网站包含许多不同类型对象间的复杂交互，将这些对象建模为同质网络是很困难的。作为半结构化的表示方法，异质信息网络可以有效建模和处理复杂多样的数据。其次，与同质网络相比，异质网络可以融合更多类型的对象及其复杂交互关系，也可以融合多个社交网络平台的信息。最后，异质信息网络包含丰富的语义，将导致更细微的知识发现。因此，异质信息网络作为有效的信息融合方法可以用于整合推荐系统中的对象和关系，可能产生更加准确的推荐结果。  本工作首先调研总结相关异质信息网络推荐算法，了解异质信息网络的数据形式及推荐评价方法。然后，深入挖掘不同元路径下的语义关系相似性，设计基于元路径的推荐算法，使得推荐模型融合多种语义信息，得到更加全面准确结果。   1. **研究现状和目标** 2. **研究现状**   由于大量复杂系统存在广泛的关联数据，信息网络建模与分析成为学术界和工业界的研究热点。很多研究者从不同角度对信息网络展开研究。下面从与本工作密切相关的两个方面，即异质信息网络中的相似性推荐和网络表示学习，分别介绍国内外研究现状。  1.1 异质信息网络中的相似性推荐  传统的推荐系统通常仅利用用户-产品评分反馈信息进行推荐。协同过滤是最普及的技术之一，包括两种方法：基于记忆的方法和基于模型的方法。最近，矩阵分解体现出其在推荐系统中的有效性和效率，它将用户-产品评分矩阵分解为两个低级的用户-特征和产品-特征矩阵，然后利用分解矩阵进行预测。随着社交媒体的普及，越来越多的研究人员通过利用用户间的社交关系研制社交推荐系统。  近年来，异质信息网络（Heterogeneous Information Network），即多种类型节点和链路的建模方式，在数据挖掘领域大放异彩。作为重要的数据挖掘任务，推荐系统包括许多对象类型（如电影推荐中的用户、电影、演员和导演等）及对象间的丰富交互关系（如观看信息、社交信息和属性信息等），可以自然而然地利用异质信息网络建模。而异质信息网络优越的信息综合能力及丰富的语义信息使得更准确的推荐成为可能。  元路径（meta-path），即连接两类对象的关系组合，被广泛用于异质信息网络中的语义信息建模和关系抽取。Shi等人实现了一个基于语义的推荐系统HeteRecom，利用元路径的语义信息来评估电影间的相似性。此外，考虑属性值，如链接上的评分，他们进一步将推荐系统建模为加权异质信息网络，并提出基于语义路径的个性化推荐方法SemRec。为充分利用关系异质性，Yu等人引入基于元路径的潜在特征来表示沿不同类型路径的用户和产品间的连接，然后利用贝叶斯排名优化技术在全局和个性化水平上定义推荐模型。同样基于元路径，Burke等人提出在加权混合中包含多个关系的推荐方法。  许多方法利用异质信息网络来融合各种信息。结合不同的上下文，Jamali等人提出基于上下文的矩阵分解模型，考虑每个实体的一般潜在因子和依赖于上下文的潜在因子。使用用户隐式反馈数据，Yu等人进一步解决全局和个性化实体推荐问题。根据相关的兴趣组，Ren等人提出基于簇的引用推荐框架来预测每个查询在书目网络中的引用。同样，Wu等人结合兴趣组信息，利用图表摘要和基于内容的聚类方法来进行媒体推荐。基于多种异质网络特征，Yang等人使用基于SVMRank的方法将多个特征建模至统一的框架中。利用多种类型的关系，Luo等人提出一种社交协同过滤算法。此外，采用用户和产品的相似性作为正规化，一些工作提出用于推荐的矩阵分解框架。  总的来说，大多数异质信息网络的相似性推荐算法集中在简单的网络结构（如二分或星型模式）和粗糙的语义建模（如元路径）上。但很多实际数据，特别是电子商务数据，所构建的异质信息网络包含很多节点和边类型，没有简单的网络模式。这种无模式异质信息网络带来了很多新的研究问题，如多种类型对象关系的管理及元路径的自动产生等，与此同时也亟需更强大语义建模和分析方法挖掘丰富细致的语义信息。  1.2 网络表示学习  网络表示学习是把网络节点映射到一个低维向量空间，从而用低维向量来表示网络中的节点。网络表示学习使得学习到的节点表示能保持网络中的有效结构及性质等信息，并能够与许多传统机器学习技术实现完美耦合，因此成为近些年数据挖掘研究领域的热点问题。  不同于传统的网络表示学习普遍采用浅层模型，图神经网络采用深度神经网络模型分析图数据，具有更好的学习能力，并有可能融合知识实现推理功能，因此成为近些年的研究热点。图卷积网络作为图神经网络的重要分支被广泛研究，其主要分为Spectral方法和Spatial方法。在Spectral方法中，相关研究对图的拉普拉斯矩阵进行特征值分解模拟离散傅里叶变换，将图信号变换到谱域来构建在图结构数据上的“傅里叶基”，从而在图信号上进行卷积操作。ChebNet模型在图谱论基础上，针对傅里叶基运算耗时问题，设计了基于契比雪夫多项式的滤波器，保证算法的时间复杂度与图规模线性相关。GCN模型将ChebNet进行了一阶近似，将邻接矩阵更换成自环的邻接矩阵；作为ChebNet的一种特殊形式，GCN将谱图卷积和空间结构结合以得到更好的节点表示。在Spatial方法中，相关研究使用图的拓扑结构，将相邻节点看作一个集合，节点之间通过特征信息的相互传递，从而学习每个节点的表示。同时，由于注意力机制在许多任务中取得了效果，研究者也试图在图神经网络中使用注意力机制。  总体来看，目前大多数的网络表示学习重点考虑如何更有效地保持网络的结构以及性质等，并在许多任务上都取得优越的效果。因此，如何将先进的网络表示学习方法更好地运用至异质信息网络中的相似性推荐，将是本工作的关键研究内容。   1. **拟解决的主要问题**   围绕研究目标，学习异质信息网络中的相似性推荐、网络表示学习等相关理论知识，掌握相应算法原理，并尝试在异质信息网络上运用GCN产生更好的推荐结果；  学习并搭建深度学习框架TensorFlow；  编程实现模型，并准备经典数据集对模型进行训练。验证算法的有效性，并对结果进行评估，尝试总结模型优缺点；  对算法的不足之处尽可能进行优化。  **三、 研究方法和措施**  根据本课题的研究内容和拟解决的主要问题，决定采用以下方法进行研究：   1. **文献调研**   异质信息网络中的相似性推荐是未接触过的新问题。首先，阅读相关综述，了解异质信息网络中的重要概念，从而对待研究的问题产生宏观上的总体认知。其次，通过进一步阅读相应论文，学习综述所提及的先进且具有代表性的算法，并根据研究目标认真思索解决方法，尝试对现有模型进行改进。   1. **博客参考**   在了解相似性推荐算法的基本实现方法后，对于论文中阐述不够清晰或模型设置不够明确的部分，通过查找相关视频或博客进一步学习。   1. **源码阅读**   参考GitHub上相似性算法的具体实现，从他人的模型中思考如何实现自身的研究目标，如何在已有的基础上实现改进，在阅读源码过程中寻找灵感。   1. **交流讨论**   在模型设计过程中遇到问题或疑惑，可与导师或师兄交流讨论，在其指导下找到相应的解决方法。   1. **代码实现**   有了详细的设计方案后，可以开始算法代码的编写。在实际操作中进一步熟悉框架的使用方法、参数的优化过程等，并在迭代中观察模型的效果，尝试进一步改进与调整。  **四、 进度安排**  第一阶段（01-02周）：阅读相关论文，了解相似性推荐的实现算法，并提交开题报告；  第二阶段（03-04周）：根据调研结果，对比分析现有的相似性推荐算法，选择合适的算法，并在此基础上改进，进一步设计详细的实现方案；  第三阶段（05-08周）：熟悉数据特点及相应算法框架，完成方案算法的编程实现，进行中期检查；  第四阶段（09-11周）：完成模型实验验证模块的编程实现，并对结果进行评估；  第五阶段（12-15周）：完成毕业设计论文的撰写，进行答辩。  **主要参考文献：**  1. Shi C, Li Y, Zhang J, et al. A survey of heterogeneous information network analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(1): 17-37.  2. Shi C, Zhou C, Kong X, et al. Heterecom: a semantic-based recommendation system in heterogeneous networks[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012: 1552-1555.  3. Shi C, Zhang Z, Ji Y, et al. SemRec: a personalized semantic recommendation method based on weighted heterogeneous information networks[J]. World Wide Web, 2019, 22(1): 153-184.  4. Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: A heterogeneous information network approach[C]//Proceedings of the 7th ACM international conference on Web search and data mining. ACM, 2014: 283-292.  5. Burke R, Vahedian F, Mobasher B. Hybrid recommendation in heterogeneous networks[C]//International Conference on User Modeling, Adaptation, and Personalization. Springer, Cham, 2014: 49-60.  6. Jamali M, Lakshmanan L. HeteroMF: recommendation in heterogeneous information networks using context dependent factor models[C]//Proceedings of the 22nd international conference on World Wide Web. ACM, 2013: 643-654.  7. Ren X, Liu J, Yu X, et al. Cluscite: Effective citation recommendation by information network-based clustering[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014: 821-830.  8. Wu J, Chen L, Yu Q, et al. Trust-aware media recommendation in heterogeneous social networks[J]. World Wide Web, 2015, 18(1): 139-157.  9. Luo C, Pang W, Wang Z, et al. Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations[C]//2014 IEEE International Conference on Data Mining. IEEE, 2014: 917-922.  10. Shi C, Wang R, Li Y, et al. Ranking-based clustering on general heterogeneous information networks by network projection[C]//Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. ACM, 2014: 699-708.  11. Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs[J]. arXiv preprint arXiv:1312.6203, 2013.  12. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Advances in neural information processing systems. 2016: 3844-3852.  13. Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.  14. Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903, 2017. | | | | | | | |
| 指导教师签字 |  | | 日期 | | 2019 年 3 月 11 日 | | |

注：可根据开题报告的长度加页。