

7月9日

一、IJCAI 2021--跨域推荐 (Cross-Domain Recommendation) 的最新综述(<https://zhuanlan.zhihu.com/p/371162633>)

1 背景介绍

- 基于协同过滤技术 (Collaborative Filtering, CF) ([3条消息](#)) [个性化智能推荐\(协同过滤算法\)技术研究](#) [zolalad的专栏-CSDN博客](#)的推荐系统，都多多少少受到数据稀疏的影响
- 特别对于新用户或者新产品 (Cold-Start, 冷启动问题) 来说，由于在系统内还没有产生任何的交互信息 (评分、评论等等)，推荐的精度会比较低
- 跨域推荐的提出就是为了解决这个数据稀疏问题。基本思路是利用**丰富领域 (richer domain又称为source domain)** 的较为丰富训练数据来提升**稀疏领域 (sparser domain又称为target domain)** 的推荐精度。
- 领域 (domain) 的定义：
 - **内容层级相关性 (content-level relevance)**：两个或多个领域中，用户之间或产品之间存在共同的内容或者特征 (例如关键字、标签)。但是这些领域不存在共同的用户或者产品。例如：亚马逊音乐 (Amazon music, 音乐相关) 和奈飞 (Netflix, 电影相关)。
 - **用户层级相关性 (user-level relevance)**：两个或多个领域中，存在共同用户但是产品层级不同。产品层级的不同又可以细分为属性层级 (attribute-level) 的不同 (即类型相同 (例如图书) 但是属性不同，例如教科书、小说、自传等等) 和类型层级 (type-level) 的不同 (例如图书、电影、音乐、服装等等)。
 - **产品层级相关性 (item-level relevance)**：两个或多个领域中，存在共同产品但是用户不同。例如：MovieLens和奈飞，都是电影相关的系统，存在大量相同的电影，但是用户不同，或者很难识别用户一致性。这种类型在一些文献里又被称为cross-system recommendation (跨系统推荐)。

2 综述动机

- 近些年来，在跨域推荐领域出现了一些新的或愈发明显的挑战，例如：
 - 特征映射问题
 - 嵌入 (embedding) 优化问题
 - 负面迁移问题 (negative transfer)
- 随着应用场景的拓展，跨域推荐也出现了一些新的方向：
 - 双目标跨域推荐 (dual-target CDR)
 - 多目标跨域推荐 (multi-target CDR)

3 不同的跨域推荐场景以及挑战

- **场景1. 单目标跨域推荐 (single-target CDR)**：传统的跨域推荐主要是针对的单目标跨域推荐 (single-target CDR)，即利用丰富领域 (源领域, source domain) 来提升稀疏领域 (目标领域, target domain) 的推荐精度。根据上面‘领域’的不同定义，单目标跨域推荐可以细分为以下三个应用场景：

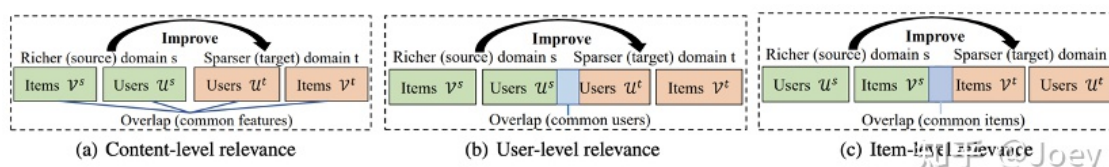


Figure 1: Single-target CDR scenarios

- 这个场景下，研究人员将面临如下三个挑战：
 - 构建基于内容的关系
 - 生成准确的用户/产品嵌入表达 (embeddings) 或者评价模式 (rating patterns)
 - 学习准确的映射关系：领域之间embedding或rating pattern的迁移对应关系。
- 场景2. 多领域推荐 (Multi-Domain Recommendation)**：在单目标跨域推荐中，有一个研究分支，即多领域推荐 (Multi-Domain Recommendation)。这个分支主要针对来之多个领域的特定用户集中用户，推荐来之多个领域的特定产品，本质上这也是single-target CDR，因为推荐的目标是特定一个集合（用户集合）。因此，本综述认为这个研究方向还是single-target CDR的一个分支。

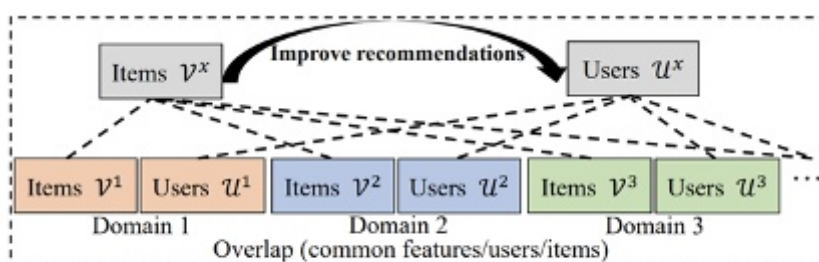


Figure 2: MDR scenario

- 场景3. 双目标跨域推荐 (dual-target CDR)**：这是一个新的跨域推荐场景，即同时利用两个领域的数据来同时提升两个领域的推荐精度。这个场景面临以下两个新的挑战：
 - 构建一个可行的双目标跨域推荐的框架
 - 各领域之间的对应关系，优化用户/产品的嵌入表达 (embeddings)
- 场景4. 多目标跨域推荐 (Multi-target CDR)**：这也是一个新的跨域推荐场景，即同时利用多个领域的数据来同时提升多个领域的推荐精度。这个场景面临以下挑战：
 - 避免负面迁移 (negative transfer)，这个问题随着越来越多领域的加入会愈加严重。

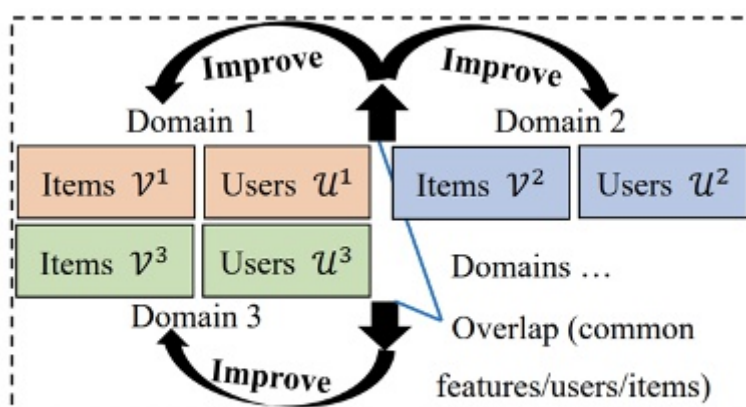
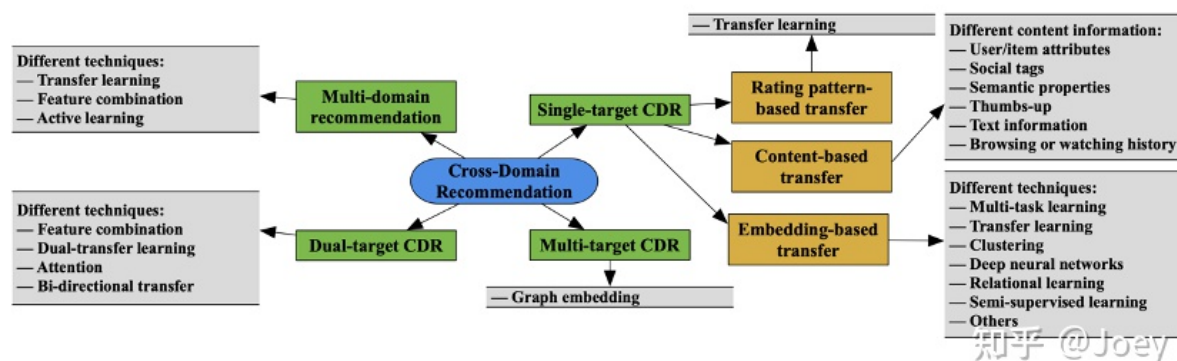


Figure 4: Multi-target CDR scenario

4 相关的研究进展

- 现有的跨域推荐方法主要划分如下图：



5 可用数据集集合

Table 3: Summary of datasets for CDR

Datasets	Domains	Data types	Scale	Website
Arnetminer [Tang <i>et al.</i> , 2012]	Research domains (user-level relevance — attribute-level)	Paper & author & conference name ...	1 million	https://www.aminer.org/collaboration
MovieLens + Netflix [Zhu <i>et al.</i> , 2018]	Movie (item-level relevance)	Rating & tag	25 million & 100 million	https://grouplens.org/datasets/movielens/ https://www.kaggle.com/netflix-inc/netflix-prize-data
Amazon [Fu <i>et al.</i> , 2019]	Book & music & movie ... (user-level relevance — type-level)	Rating & review & side information	100 million+	http://jmcauley.ucsb.edu/data/amazon/
Douban [Zhu <i>et al.</i> , 2018] [Zhu <i>et al.</i> , 2019; Zhu <i>et al.</i> , 2020]	Book & music & movie (user-level relevance — type-level)	Rating & review & side information	1 million+	https://github.com/FengZhu-Joey/GA-DTCDR/tree/main/Data

6 未来研究方向

- 异质化跨域推荐：现有跨域推荐的假设前提是跨域的信息是同质的，但是实际应用场景中，存在跨域的异质信息。
- 序列化跨域推荐：跨域推荐系统也和传统推荐系统一样，面临如何序列化地给用户/产品建模。
- 隐私保护的跨域推荐：现有的跨域推荐方法忽略了信息孤岛的问题。而实际应用场景，用户敏感信息是无法直接跨域分享的。
- 另外，数据集稀疏程度、领域间的重叠规模（overlap scale）、以及领域间的关联度，如何分别影响跨域推荐的性能？这些问题同样也值得进一步研究。

7 结论

近些年，随着深度神经网络以及图学习的发展，跨域推荐越来越受工业界、学术界的关注。本篇综述分别从单目标跨域推荐、多领域推荐、双目标跨域推荐、以及多目标跨域推荐来系统性地分析、总结现有跨域推荐方法。综述的最后，给出了三个未来值得研究方向以供读者参考，并希望这些方向能得到进一步的研究，最终解决困扰推荐领域多年的问题——数据稀疏问题。