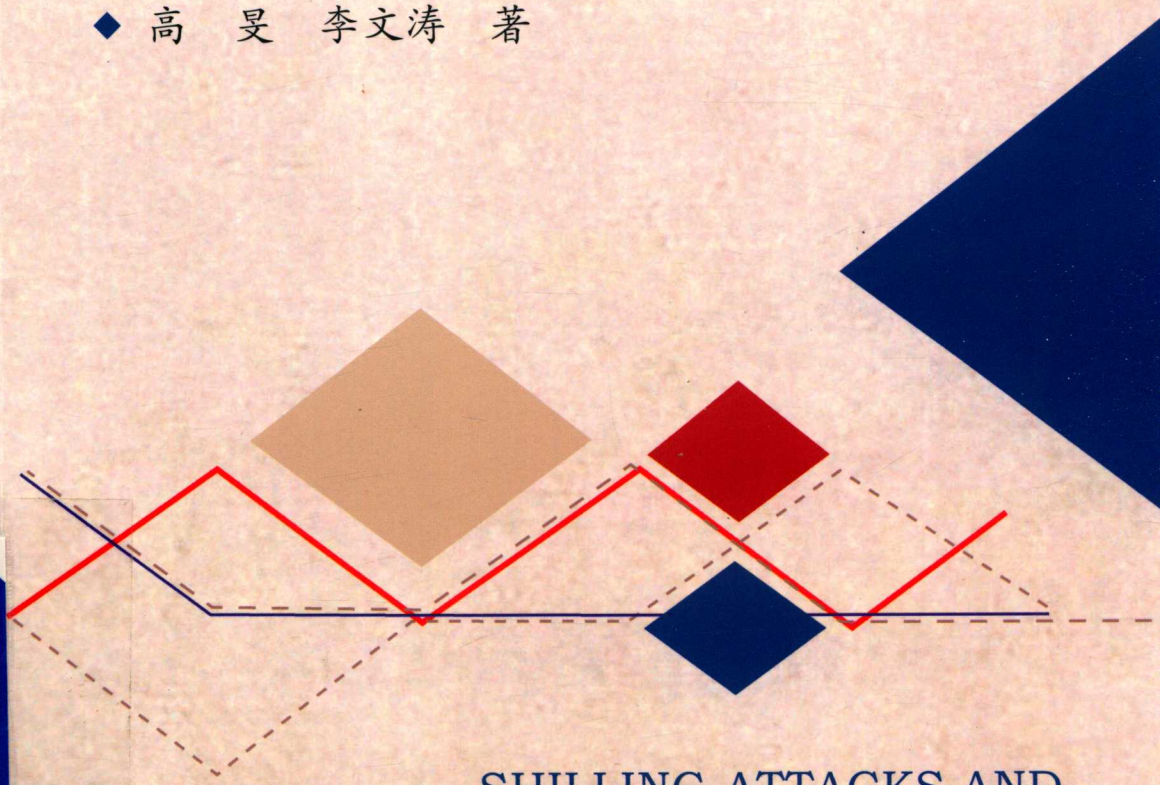
面向社会化推荐的托攻击及检测研究

◆ 高 旻 李文涛 著



SHILLING ATTACKS AND DETECTION IN SOCIAL RECOMMENDER SYSTEMS

科学出版社

面向社会化推荐的托攻击及检测研究

Shilling Attacks and Detection in Social Recommender Systems

高 旻 李文涛 著

科学出版社

北 京

内 容 简 介

社会化推荐利用社交关系缓解基于评分驱动的推荐系统中存在的稀疏性与冷启动等问题，然而推荐系统开放性的特点使其易受托攻击的严重影响。托攻击者通过注入虚假信息操纵推荐结果，影响推荐系统的公正性。针对此问题，本书完成四方面工作：一是分析社会化推荐中可能的托攻击形式，提出托攻击模型；二是在检测注入评分的攻击时，从选择行为分析入手，提出基于流行度的分类特征；三是在检测注入关系的攻击者时，使用基于拉普拉斯的特征提取方法，对用户的高维特征进行无监督提取；四是在评分与关系特征上分别训练分类器，基于半监督协同训练实现社会化推荐中的托攻击检测。

本书适合作为相关专业研究生、本科生及业界人员的参考书。

图书在版编目 (CIP) 数据

面向社会化推荐的托攻击及检测研究 / 高旻，李文涛著. —北京：科学出版社，2016.11

ISBN 978-7-03-050336-7

Ⅰ. ①面… Ⅱ. ①高…②李… Ⅲ. ①电子商务－商业管理 Ⅳ. ①F713.36

中国版本图书馆 CIP 数据核字 (2016) 第 257105 号

责任编辑：阚 瑞 / 责任校对：桂伟利

责任印制：张 倩 / 封面设计：迷底书装

科 学 出 版 社 出版

北京东黄城根北街 16 号

邮政编码：100717

http://www.sciencep.com

北京盛通印刷股份有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 12 月第 一 版 开本：720×1000 1/16

2016 年 12 月第一次印刷 印张：7 1/2

字数：153 000

定价：45.00 元

（如有印装质量问题，我社负责调换）

前　言

随着电子商务零售业的迅猛发展和社交网络营销的兴起，以用户间社交关系作为额外输入的社会化推荐系统成为新的研究方向。社会化推荐系统基于社交关系体现用户间相似性这一假设，对解决传统推荐系统中存在的冷启动问题和提高推荐结果的准确性具有重要作用。但社会化推荐系统天然开放性的特点，使其容易受到托攻击者注入虚假欺骗信息（虚假评分或虚假关系等）的影响。此类攻击称为“托攻击”，托攻击严重影响了推荐结果的公正性和真实性，降低了用户对系统的信任度。

社会化推荐系统可以看成传统推荐系统与在线社交网络结合的产物。现有研究大多关注评分驱动的推荐系统或关系驱动的社交网络中托攻击的检测问题，而较少关注同时受评分和关系驱动的社会化推荐系统可能受到的攻击形式与检测手段。针对现有研究的不足，本书首先对社会化推荐系统中的托攻击者的行为方式进行建模，然后提出用于检测推荐系统与社交网络中虚假欺骗信息的特征提取方法，进而得到社会化推荐系统中的托攻击检测技术。本书分别从以下几个方面展开研究。

（1）构建面向社会化推荐系统的托攻击模型，并从攻击成本与攻击效果角度对所提模型进行分析。托攻击模型是托攻击者向系统注入虚假用户概貌的手段。通过分析现有社会化推荐技术的工作原理，归纳出托攻击者可能的攻击形式，从而提出托攻击模型。然后分析攻击模型对推荐结果的影响，得到所提托攻击模型对社会化推荐系统的攻击效果。

（2）针对评分驱动的推荐系统中的托攻击问题，提出一种基于流行度分类特征的托攻击检测方法。推荐系统中托攻击者通过注入虚假评分影响推荐结果，传统方法大多从托攻击者的评分方式入手，此类方法难以对新形式攻击进行检测。为了解决这个问题，从托攻击者与正常用户不同的项目选择行为入手，分析用户概貌中项目流行度分布存在的差异，得到用于检测推荐系统托攻击的特征提取方法，最后结合分类器对推荐系统中的托攻击进行检测。

（3）针对关系驱动的社交网络中的托攻击问题，提出一种基于拉普拉斯得分的托攻击检测方法。社交网络中托攻击者通过注入虚假关系提升自己的影响

力，从而达到传播虚假信息的目的。现有方法在训练模型时使用的特征维度较高，造成检测准确性不足。为了解决这个问题，提出无监督的特征选择方法，该方法通过拉普拉斯得分衡量特征的局部信息保持能力，以进行特征选择。在此基础上，结合半监督学习方法对社交网络中的托攻击进行检测。

（4）面向社会化推荐系统中的托攻击检测问题，提出一种基于半监督协同训练的社会化推荐系统托攻击检测方法。社会化推荐系统中的用户包括评分特征与关系特征，因此可以利用推荐系统与社交网络中检测托攻击的特征提取方法，得到用户评分视图与关系视图的特征。同时考虑到系统中标签不足的问题，将半监督协同训练算法用于模型构建，在两个独立的特征子图上分别训练分类器，从而对社会化推荐系统中的托攻击进行检测。

本书受国家自然科学基金“基于用户可信度的抗托攻击协同过滤推荐机理研究”（项目编号：71102065）、重庆市前沿与应用基础研究计划“基于多维社交关系挖掘的抗干扰社会化推荐研究”（项目编号：CSTS2015JCYJA40049）、中国博士后基金“基于虚假用户群体特征的抗托攻击协同过滤关键技术研究”（项目编号：2012M521680）、中央高校基金“多视图协同训练的托攻击检测研究”（项目编号：106112014CDJZR095502）等项目的资助，在此表示感谢。

限于本书作者的学识水平，书中不足之处在所难免，恳请读者批评指正。

作　者

2016 年 6 月于重庆大学

目　录

第1章 绪 论

1.1 研究背景及意义

随着 Web 2.0 时代的到来，用户在获得海量信息的同时[1]，也面临着“信息过载”问题[2]。个性化推荐系统通过向用户推送个性化的信息，能够缓解信息过载问题，从而在各大电子商务站点中得到广泛的应用[3]，如亚马逊 2006 年的销售报表显示其 35%的销售额来自于推荐系统[4]，京东 2015 年的总订单中推荐系统贡献占比约为 13%[5]。推荐系统根据用户的购买记录向用户推送感兴趣的信息。推荐系统在帮助用户获得良好购物体验的同时，也能够提高企业的销售额，实现用户与企业的共赢[6]。

近年来，在线社交网络得到飞速发展，社交网络站点如 Twitter 和 Facebook 等已经成为通信与信息传播的重要平台[7]。这些网站具有的开放性、及时性等特点使其拥有大量的用户[8]。据报道 2015 年 Facebook 的月平均活跃用户人数为 15.9 亿[9]，Twitter 的月平均活跃用户人数为 3.05 亿[10]，截至 2015 年 9 月，新浪微博的月活跃人数为 2.22 亿，且用户群体逐渐稳定并保持持续增长[11]。

电子商务零售业的迅猛发展和结合社交网络的社会营销的兴起，使得社会化推荐成为研究热点[12]。社会化推荐系统在电子商务领域特指以社交关系作为额外输入的推荐技术[13]。社会化推荐基于有社交关系的用户具有相似性这个基本假设[14]，利用与人们息息相关的社交关系将信息推送至兴趣群体，以提高商品的销售和提高用户的购物满意度。近年来，Facebook 和 Twitter 等社交网站尝试在站内进行购物尝试，在特定场景为用户推荐商品[15]；亚马逊、eBay、淘宝等也在与社交网络等社会媒体合作或收购社交电商以开展商品社交推荐[16]。

社会化推荐利用社交关系揭示用户的购物喜好，对解决推荐系统中的冷启动问题和提高推荐的准确性具有重要作用[17]。为了以示区分，本书把未使用社交关系产生推荐的传统推荐系统称为评分驱动的推荐系统或简记为推荐系统，而同时利用评分与关系的推荐系统称为社会化推荐系统。

由于社会化推荐系统天然开放性的特点并且同时受到评分与关系驱动，所以社会化推荐系统容易遭受两方面的攻击：①虚假评分攻击，由于推荐结果的产生依赖于用户对项目的评分，所以托攻击者注入虚假评分可以使得目标项目的推荐可能性提高或者降低[18]；②虚假关系攻击，由于社会化关系的建立成本较低，托攻击者可以注入虚假用户并建立用户关系，对推荐进行干扰，从而提高商品被推荐的概率[19]。

这些攻击统称为“托攻击”，即托攻击者注入虚假欺骗信息达到影响推荐结果，获取不良经济利益的目的。托攻击问题严重影响了推荐的公正性和真实性，并降低了用户对系统的信任度。

现有托攻击研究大多分别检测虚假评分与虚假关系，如评分驱动的推荐系统中攻击概貌的检测[18]或社交网络中垃圾用户的检测[19]等。然而，社会化推荐系统同时受到评分与关系驱动，所以虚假评分与虚假关系对社会化推荐系统的正常运行均可能带来潜在的安全隐患。托攻击的存在可能造成推荐结果被人为操纵，进而导致用户的购物满意度降低，商家的经济利益受到侵害，因此亟须开展社会化推荐系统托攻击检测研究。

针对现有研究的不足，本书着重讨论社会化推荐托攻击问题，并主要回答如下两个问题：①社会化推荐系统是否容易受到托攻击的影响；②社会化推荐系统中的托攻击是否能够检测。为此，本书通过社会化推荐系统中托攻击用户行为建模、用于检测虚假信息的特征提取方法以及面向社会化推荐的托攻击检测算法三个方面进行研究。

本书的研究在理论上能够为相关的托攻击检测研究提供基础，使得更多研究者关注此类托攻击问题；在实践中能够作用于现有的社会化推荐系统，用于维护社会化推荐系统的推荐结果公平性。

1.2　研 究 现 状

1.2.1　社会化推荐系统研究现状

个性化推荐系统的概念在 20 世纪 90 年代提出[20]，其中代表性的工作是 GroupLens 项目组于 1994 年建立的文章推荐系统[21]以及于 1997 年建立的 MovieLens 电影推荐系统[22]。自此，越来越多的研究者关注个性化推荐系统。

根据推荐系统使用的算法不同，可分为三类[23]。

（1）基于内容的推荐（content-based recommendations）[24]。基于内容的推荐技术根据用户已购项目的内容信息与目标项目的内容信息之间的相似度关系产生推荐。该方法的思想源于信息检索中的关键词搜索技术，能够利用该用户的历史记录产生推荐结果，但是对电影等难以进行特征表达的商品难以产生推荐结果。

（2）协同过滤推荐（collaborative filter recommendations）[25]。协同过滤是实际中广泛应用的个性化推荐技术。协同过滤技术首先通过计算评分相似度查找目标用户的近邻用户，然后综合近邻用户对商品的评分信息得到目标用户对商品的预测评分。根据推荐机制的不同，协同过滤包括基于存储的（memory-based）的和基于模型的（model-based）两类推荐算法。基于存储的协同过滤预先计算所有用户（项目）之间的相似度然后产生推荐，而基于模型的算法则训练一个模型对未知评分进行预测。协同过滤推荐算法的主要缺陷在于容易面临稀疏性、冷启动与托攻击等问题。

（3）混合推荐（hybrid approaches）[26]。

以上各类算法在实际中会有自己的缺陷，因此有研究者提出在实际中把不同种类的算法进行结合，从而利用各种算法的优势，以提高推荐的准确性。

随着在线社交网络的兴起，以社交关系作为额外输入的社会化推荐系统成为研究的热点。社会化推荐系统利用社交关系改善推荐结果，能够解决评分驱动的推荐系统面临的问题，同时能够提高推荐的准确性。

主流的社会化推荐大多以协同过滤算法作为基本模型，按照协同过滤算法的分类方式，社会化推荐也可分为两类。

（1）基于存储的社会化推荐。这种方法的核心在于近邻用户的计算，如Golbeck[27]根据目标用户到所有用户的最短路径长度逐级进行信任度计算，并将信任度大于阈值的用户作为目标用户的社交近邻；周超等[28]在对信任关系强度和用户购物兴趣进行建模的基础上，识别出与目标用户有共同爱好的朋友作为近邻用户以产生推荐。

（2）基于模型的社会化推荐。该方法以基于模型的协同过滤为基础，并将社交关系融入其中，如 Ma 等[29]提出概率矩阵分解模型，从而将用户之间的关系体现到用户购物喜好上，得到社会信任集成的方法；胡祥等[30]采用流形排序方法度量用户间的社会相似度，然后利用正则化技术将社交关系融入到矩阵分解模型中；此外还有一些基于张量分解模型和多维信任关系模型的社会化推荐技术等[31]。

1.2.2　评分驱动的推荐系统中托攻击检测研究现状

在评分驱动的推荐系统中，托攻击者通过注入虚假评分构建攻击概貌，以便伪装成正常用户。现有研究者对常见的托攻击形式进行归纳，得到推荐系统中的托攻击模型，并分析了这些模型对推荐系统的影响。在面向评分的推荐系统中，随机攻击、均值攻击[32]、流行攻击和段攻击[33]等是几种基础的攻击模型，后续又有研究者[34,35]提出爱憎攻击、逆流行攻击、探测攻击和混淆攻击，用于构造攻击概貌。

托攻击者采用托攻击模型构建攻击概貌，托攻击检测的目的就是检测出托攻击者注入的攻击概貌。根据使用标签信息的不同，评分驱动的推荐系统中托攻击检测可以分为有监督、无监督和半监督等检测方法。

（1）有监督的方法主要通过探寻用户概貌的特征以进行托攻击检测，如 Chirita 等[36]最早提出平均误差值（Rating Deviation from Mean Agreement，RDMA）和平均相似度（Degree of Similarity，DegSim）用于区分正常用户概貌与攻击概貌。后来 Mobasher 等[37]在 RMDA 特征的基础上提出加权评分偏离度（Weighted Degree of Agreement，WDA）和过滤平均目标偏差（Filter Mean Target Difference，FMTD），以解决小规模段攻击难以检测的问题。

（2）无监督的托攻击检测方法主要通过对两类用户的概貌特征进行聚类实现托攻击检测，典型的方法包括基于概率潜在语义分析（Probabilistic Latent Semantic Analysis，PLSA）和主成分分析变量选择（Variable-Selection Using Principal Component Analysis，VarSelect-PCA）的方法[38,39]、利用协同谱聚类在用户和项目两个视图上同时聚类的方法[40]等。

（3）基于半监督的托攻击检测方法同时利用有标签样本与无标签样本训练分类器，如 Cao 等[41,42]提出首先训练一个初始朴素贝叶斯分类器，使用最大期望（Expectation Maximization，EM）算法对参数进行求解，然后利用初始分类器预测样本的标签，将最可能标记正确的样本加入有标签的样本训练集中训练新的分类器，迭代多次直到达到一定的停止条件。

1.2.3　关系驱动的社交网络中托攻击检测研究现状

近年来，社交网络中的托攻击检测问题也获得了广泛关注。在社交网络中

“托”又称为“水军”[19]，即托攻击者注入虚假用户并与正常用户建立社交关系以达到发布不良信息、传播广告信息等目的。

根据使用标签数据的多少，社交网络托攻击检测也可以分为有监督、无监督和半监督的检测方法。

（1）有监督的检测方法通过提取社交网络中用户的各种特征训练分类模型，例如，从用户注册信息、用户发布内容等信息中抽取特征构建分类器[43]，针对基于诱捕系统抓取行为特征的检测方法[44]等。

（2）无监督的检测方法主要利用社交网络的拓扑关系结合聚类方法，从而识别网络中的异常点，如利用文本和统一资源定位符（Uniform Resource Locator，URL）相似度对帖子进行聚类的检测方法[45]。

（3）半监督的检测方法同时利用有标签数据与无标签数据对用户数据进行分类，如 Li 等[46]提出一种结合信任传播的半监督检测框架，以对社交网络中的虚假用户进行检测。

1.3 研究内容和目的

1.3.1 研究内容

本书面向社会化推荐系统的托攻击模型与检测技术展开研究，并从社会化推荐系统中的托攻击建模、基于流行度的推荐系统托攻击检测、基于拉普拉斯得分的社交网络托攻击检测和基于半监督协同训练的社会化推荐系统托攻击检测技术四个方面进行深入探讨。

本书的研究路线如图 1.1 所示，主要研究内容包括以下几方面。

（1）面向社会化推荐的托攻击模型研究。社会化推荐系统同时受到评分与关系的同时驱动，所以容易受到托攻击者注入虚假评分与关系的影响。本书从社会化推荐系统的工作机制入手，归纳总结了社会化推荐系统中托攻击者可能的攻击形式，提出相应的托攻击模型，然后在社会化推荐算法上评估攻击模型的效果。

（2）用于检测托攻击的特征提取方法。社会化推荐系统容易受到虚假评分与虚假关系的影响，所以分别讨论评分驱动的推荐系统中和关系驱动的社交网

络中托攻击的检测策略，重点研究用于检测虚假评分与虚假关系的特征提取方法，以用于社会化推荐系统托攻击的检测。

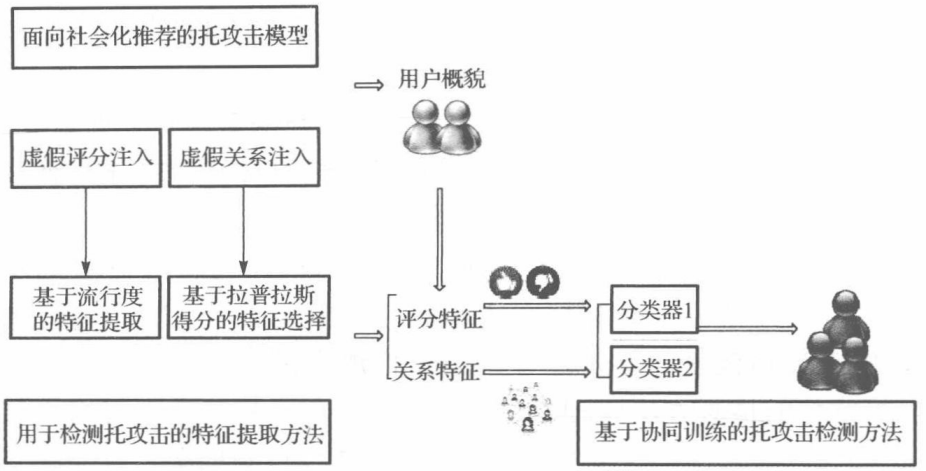


图 1.1　研究路线图

在检测注入虚假评分的托攻击者时，提出一种基于流行度的分类特征提取方法，该方法从用户的选择行为入手，分析得到正常用户与虚假用户在选择项目时由于偏好不同，导致用户概貌中项目的流行度分布的不同。并从流行度分布中抽取特征用于特征提取，从而对推荐系统中的虚假用户进行检测。

在检测注入虚假关系的托攻击者时，提出一种基于拉普拉斯的特征提取方法，该方法能够根据特征保持原有空间几何信息的能力对特征进行选择。该方法是一种无监督的判别分析方法，可以与半监督学习结合，从而对社交网络中的虚假用户进行检测。

（3）基于协同训练的托攻击检测方法。由于社会化推荐系统中的托攻击者可以注入虚假评分与虚假关系，所以可以利用上述研究的特征提取方法从这两个信息中分别提取特征。同时由于系统中有大量的无标签用户数据和少量的有标签用户数据，所以可以利用半监督协同训练在两个特征子图上分别训练分类器，从而提高托攻击检测的准确性，并更加适合在现实中对社会化推荐系统中的托攻击进行检测。

1.3.2　创新点

本书按照研究路线图展开面向社会化推荐系统的托攻击模型与检测研究，具有以下几个创新点。

（1）提出面向社会化推荐系统的托攻击模型。现有社会化推荐系统的研究大多集中在开发推荐准确性高的社会化推荐算法，但是很少有工作对社会化推荐系统的托攻击问题进行研究。社会化推荐系统受评分与关系驱动，所以托攻击者注入的虚假评分与虚假关系可能对推荐结果产生影响。本书对社会化推荐系统中托攻击者可能的攻击形式进行概括，并提出相应的托攻击模型，然后通过实验分析了所提托攻击模型的攻击效果。

（2）提出用于检测虚假评分的基于流行度的特征提取方法。评分驱动的推荐系统中用户可以注入虚假评分操纵推荐结果，传统的检测方法大多从用户概貌中项目的评分分布入手，本书分析了用户的选择行为，提出从用户概貌中项目的流行度分布入手对用户的特征进行提取，从而对托攻击进行检测。

（3）提出用于检测虚假关系的基于拉普拉斯得分的特征选择方法。社交网络中垃圾用户通过注入虚假关系提升自身影响力以达到传播虚假信息，获取经济利益的目的，本书将此类攻击也归于托攻击的范畴。传统的社交网络托攻击检测方法大多利用网络结构对托攻击进行检测，但是这类方法在训练模型时面临特征空间维数较高的问题，导致检测精度不高。基于此，本书提出基于拉普拉斯得分的特征选择方法，从原有特征中选择价值较大的特征，从而在降维的同时提高分类的准确性。该特征选择方法没有利用标签信息，所以可以与半监督学习进行结合。

（4）提出用于检测社会化推荐系统托攻击的基于协同训练的检测方法。本书的主要目的是对社会化推荐系统中的托攻击进行检测，由于社会化推荐系统中用户具有评分信息与关系信息，所以可以对两类特征进行特征提取。在得到这两类特征之后，就可以使用协同训练的方式对用户进行分类，以找到虚假用户。该方法同时利用少部分的有标签数据与大量的无标签数据，且使用协同训练以保持对无标签数据利用的准确性，从而能够提高托攻击检测准确性。

1.4 本书的组织结构

本书共七章，其中第 1 章和第 2 章阐述研究的背景及提供相关的背景知识。第 3 章提出面向社会化推荐系统的托攻击模型。由于社会化推荐系统中的托攻击者可以注入虚假评分与虚假关系，所以第 4 章与第 5 章分别对评分驱动的推荐系统中和关系驱动的社交网络中的托攻击问题进行研究，并提出用于检测托

攻击的特征提取方法，为社会化推荐系统中的托攻击检测做准备。第6章在前面工作的基础上，利用协同训练得到一个分类器以对托攻击进行检测。第7章对本书进行总结，并对以后的工作进行展望。每章的具体内容如下。

第1章，绪论。简要介绍社会化推荐系统的原理和托攻击问题以及本书的主要研究内容、研究目的及创新点等。

第2章，社会化推荐系统与托攻击检测相关技术。首先对常见的社会化推荐系统技术进行总结，然后讨论评分驱动的推荐系统中的托攻击模型与检测技术，其次讨论关系驱动的社交网络中的托攻击模型与检测技术，最后，对本书使用的一些基本技术进行概述。

第3章，面向社会化推荐系统的托攻击模型。本章归纳社会化推荐系统中托攻击者的可能攻击形式，具体来说，从注入虚假评分策略与注入虚假关系策略入手，提出相应的模型，并在典型的社会化推荐算法上测试攻击的效果。

第4章，基于流行度分类特征的推荐系统托攻击检测方法。本章对评分驱动的推荐系统中的注入虚假评分的托攻击进行检测，主要探究基于流行度的托攻击检测算法。首先对系统中的项目的流行度进行统计，然后分析每一个用户已评分项目的流行度分布，从而提取特征，最后结合分类器对虚假用户进行检测。实验表明该方法对混合攻击与新型攻击有较好的检测效果。

第5章，基于拉普拉斯得分的社交网络托攻击检测方法。本章对关系驱动的社交网络中的托攻击进行检测，主要探究如何利用拉普拉斯得分对用户进行特征提取。首先利用特征的几何分布信息保持能力进行特征提取，然后利用半监督随机森林方法对托攻击进行检测。实验表明该方法对社交网络中的托攻击有较好的检测效果。

第6章，基于协同训练的社会化推荐系统托攻击检测方法。在前面工作的基础上，可以得到社会化推荐系统中托攻击者注入的虚假用户的评分特征与关系特征视图。考虑到实际环境中有标签数据量有限，所以利用协同训练建立模型以同时利用有标签和无标签数据，进而达到更加好的检测效果。仿真实验表明该方法对提出的社会化推荐系统托攻击模型有较好的检测效果，从而能够维护社会化推荐系统的正常运行。

第7章，总结与展望。对本书中提出的工作进行总结，并对未来需要从事的工作进行展望。

第 2 章　社会化推荐系统与托攻击检测相关技术

由于社会化推荐系统开放性的特点[13]，托攻击者可能注入虚假评分和虚假关系以操纵推荐结果，破坏推荐的公平性。社会化推荐系统可以看成现有评分驱动的推荐系统与关系驱动的社交网络两者结合的产物，所以推荐系统与社交网络中存在的攻击方式可能也会对社会化推荐系统的推荐结果产生影响。因此，本章首先介绍传统的推荐算法和常见的社会化推荐算法；然后对评分驱动的推荐系统中和关系驱动的社交网络中的托攻击模型与检测手段进行概述；最后对用于托攻击检测的半监督学习算法进行简要的介绍。

2.1　评分驱动的推荐算法

传统推荐系统依据用户对项目的评分记录产生预测评分，进而形成推荐。本书将传统推荐系统称为评分驱动的推荐系统或简记为推荐系统，以与社会化推荐系统进行区分。假定系统中有 M 个用户 $\boldsymbol{U}=\{u_1, u_2, \cdots, u_M\}$，$N$ 个项目 $\boldsymbol{I}=\{i_1, i_2, \cdots, i_N\}$。定义 $\boldsymbol{R}$（$M\times N$ 维）为用户-项目评分矩阵，其中元素 $R_{i,j}$ 表示第 i 个用户对第 j 个项目的评分，如果没有评分，则 $R_{i,j}$ 为 0。评分驱动的推荐系统如图 2.1 所示，推荐的基本原理是依据用户的评分记录，对评分矩阵 $\boldsymbol{R}$ 中缺失的评分进行预测，并选择预测评分较高的项目推荐给用户。

推荐系统有不同的分类方式，根据推荐系统使用的算法不同，可分为三类[20]。

1）基于内容的推荐

基于内容的推荐利用用户已经选择的项目的属性信息，将未评分项目的属性信息与之进行匹配进而产生推荐[24]。这种方法首先根据用户的购买记录，形成用户的内容特征向量，然后将项目也表达为特征向量的形式，最后计算用户特征向量与项目特征向量之间的相似度，并将相似度较高的项目推荐给用户。