# 旅客支付意愿（WTP）评估

# 算法模型调研报告



**北京理工大学**

**2020年3月**

目录

[摘要 3](#_Toc35134592)

[第1章 概述 4](#_Toc35134593)

[1.1 项目背景 4](#_Toc35134594)

[1.2 项目目的 4](#_Toc35134595)

[1.3 具体应用场景 5](#_Toc35134596)

[1.3.1 电子商务 5](#_Toc35134597)

[1.3.2 电影和视频推荐 5](#_Toc35134598)

[1.3.3 音乐推荐 6](#_Toc35134599)

[1.3.4 社交网络 6](#_Toc35134600)

[1.3.5 个性化阅读 6](#_Toc35134601)

[1.3.6 基于位置的服务 7](#_Toc35134602)

[1.4 组织结构 7](#_Toc35134603)

[第2章 需求分析 8](#_Toc35134604)

[2.1 功能性需求 8](#_Toc35134605)

[2.2 非功能性需求 8](#_Toc35134606)

[2.2.1 数据规模约束 8](#_Toc35134607)

[2.2.2 耗时约束 8](#_Toc35134608)

[2.2.3 准确率约束 9](#_Toc35134609)

[第3章 推荐系统的发展历史调研 10](#_Toc35134610)

[3.1 基于内容的过滤推荐算法 10](#_Toc35134611)

[3.2 协同过滤推荐算法 10](#_Toc35134612)

[3.2.1 基于记忆的协同过滤 10](#_Toc35134613)

[3.2.2 基于模型的协同过滤 11](#_Toc35134614)

[3.3 混合推荐算法 12](#_Toc35134615)

[第4章 推荐系统的最新进展调研 13](#_Toc35134616)

[4.1 基于图神经网络的推荐算法 13](#_Toc35134617)

[4.1.1 基于会话的图神经网络推荐算法 14](#_Toc35134618)

[4.1.2 基于堆叠和重构图卷积网络的推荐算法 15](#_Toc35134619)

[4.1.3 基于图卷积神经网络的web领域推荐算法 16](#_Toc35134620)

[4.1.4 基于图嵌入的电子商务推荐召回算法 17](#_Toc35134621)

[4.1.5 基于记忆增强图神经网络的序列推荐算法 19](#_Toc35134622)

[4.1.6 算法优缺点对比 20](#_Toc35134623)

[4.2 基于协同过滤的推荐系统算法 21](#_Toc35134624)

[4.2.1 神经图协同过滤推荐算法 21](#_Toc35134625)

[4.2.2 基于信任信息的协同过滤推荐算法 23](#_Toc35134626)

[4.2.3 基于变体自动编码器的协同过滤推荐算法 24](#_Toc35134627)

[4.2.4 基于协同记忆网络的推荐算法 25](#_Toc35134628)

[4.2.5 算法优缺点对比 26](#_Toc35134629)

[4.3 基于知识图谱的推荐系统算法 27](#_Toc35134630)

[4.3.1 基于知识的路径循环网络算法 27](#_Toc35134631)

[4.3.2 基于知识图谱增强推荐的多任务特征学习算法 28](#_Toc35134632)

[4.3.3 基于知识图谱的用户喜好推荐算法 29](#_Toc35134633)

[4.3.4 基于知识增强记忆网络的推荐顺序算法 31](#_Toc35134634)

[4.3.5 算法优缺点对比 32](#_Toc35134635)

[4.4 其他算法 33](#_Toc35134636)

[4.4.1 基于深度学习的内容协同算法 33](#_Toc35134637)

[4.4.2 用户超长行为历史下的CTR预测 34](#_Toc35134638)

[4.4.3 考虑时空域影响的CTR预测 35](#_Toc35134639)

[4.4.4 基于局部潜在模型的Top-N推荐算法 36](#_Toc35134640)

[4.4.5 基于神经网络的上下文感知算法 37](#_Toc35134641)

[参考文献 39](#_Toc35134642)

# 摘要

本文对推荐系统的基本概念及应用场景进行了介绍，对本项目的需求进行了分析。同时，本文还对推荐系统的历史发展及最新进展进行了调研，主要对近三年CCF-A类的会议和期刊中推荐系统领域的最新文章进行调研，从基于图神经网络、基于知识图谱和基于协同过滤这三类推荐系统领域的多种算法出发，并对其他算法进行分析，调研推荐系统领域近几年的最新进展和研究方向。最后对深度学习和深度强化学习相关算法进行调研，并给出最终的解决方案。

# 第1章 概述

## 项目背景

互联网的广泛应用和移动用户端的普及给用户带来了大量的信息，满足了用户在信息时代对信息的需求，但随着网络的迅速发展，网络上信息量的大幅增长，使得用户在面对大量信息时无法从中获得对自己真正有用的那部分信息，对信息的使用效率反而降低了，这就是所谓的信息超载（Information Overload）问题。 解决信息超载问题一个非常有潜力的办法是推荐系统，它是根据用户的信息需求、兴趣等，将用户感兴趣的信息、产品等推荐给用户的个性化信息推荐平台。和搜索引擎相比推荐系统通过研究用户的兴趣偏好，进行个性化计算，由系统发现用户的兴趣点，从而引导用户发现自己的信息需求。一个好的推荐系统不仅能为用户提供个性化的服务，还能和用户之间建立密切关系，让用户对推荐产生依赖。

推荐系统显式或隐式的收集有关用户对特定事物的偏好的信息，来对用户进行预测或推荐相关的事物。它尽可能利用所有信息来源以确保新颖性、多样性、稳定性等因素，主要是通过协同过滤方法（Collaborative Filtering，简称为CF）结合基于内容的过滤方法、基于知识的过滤及社会化过滤方法来实现。协同过滤是像人类真正决策一样，不仅基于自己的经历，还吸取来自身边其他人的经验和知识，结合二者进行决策。随着推荐系统的不断发展，它应用的领域从最初的新闻推广到音乐、电视、书籍、文献、电子学习、电子商务等市场应用和网络搜索推荐等领域。

推荐系统领域最核心的两个问题是数据稀疏和冷启动问题。对于个性化机票定价而言，与其他个性化商品推荐并无本质不同，且数据稀疏的问题尤为严重。因此旅客支付意愿评估的核心问题在于解决数据稀疏的问题，通过对多维度数据的提取特征，建立用户画像，实现针对具体用户的个性化定价建议。

## 项目目的

本课题的总体目标是根据用户浏览记录和购买记录、航空公司的机票历史定价及销量、社交网络信息，以及天气等实时信息，针对当下航空公司的定价来预测每个用户的未来行为，从而可以实现基于用户画像的旅客支付意愿评估，建立出估算旅客支付意愿的算法模型。

本课题主要使用的技术内容包括：深度学习，深度增强学习，知识图谱等。总体而言，本课题会设计两个模块，一个模块是常规预测算法模块，该模块主要是用来处理航空公司日常对用户行为的建模，利用分布式增强学习框架进行训练，利用好奇心机制增加训练时的模型探索，利用神经内存读写机制来对长期用户行为进行建模，利用LSTM对用户短期行为进行建模；另一个模块是突发事件应对模块，该模块利用灵活的规则推理引擎来应对一些突发状况。

该课题可以很好地对用户行为进行建模，这样就可以预测用户的支付意愿，从而可以根据用户的支付意愿预测调整航空公司的票价，为航空公司争取较大的收益。同时，该课题所设计的算法完全兼顾了大数据量时所需要的分布式训练，保障了算法的训练效率，此外，该算法使用的好奇心机制能够保障算法的鲁棒性，算法针对长序列数据所添加的神经内存读写机制可以保障对用户长期行为特征的有效提取和利用。

## 具体应用场景

在当前的商业场景下，推荐系统有着广泛的应用，在各个领域和场景下都有一席之地。经过调研我们发现，推荐系统在以下的几个应用场景中都能产生巨大的影响。

### 1.3.1 电子商务

随着网上购物行为的推广，越来越多的电商出现，如京东、淘宝、苏宁等。这些购物网站已经使用推荐系统来帮助他们的用户找到要购买的产品。在这些B2C电子商务网站，可以基于畅销商品，用户人口统计，或分析用户过去的购买行为作为对用户未来购买行为的预测。

### **1.3.2 电影和视频推荐**

在各个视频和电影租赁网站上，推荐系统被用来为用户推荐可能会感兴趣的视频或电影。平台通过这种方式，来增加用户在平台上的逗留时间，从而间接的收获利益。以美国的一家在线影片租赁提供商Netflix来说，Netflix它们的算法是基于商品的协同过滤推荐算法，即给用户推荐和他们曾经喜欢的电影相似的电影。标签是除了传统的评分外连接用户-商品数据的另一种方式。例如，MovieLens电影评论网站的用户能够使用简单的单词为电影自由分配标签。因而协同过滤和社会标签分析是这类系统中两种提高推荐性能的有效技术。除了能够提高性能，标签也有助于推荐的可解释性，使推荐结果简明直接。

### **1.3.3 音乐推荐**

推荐系统在各个音乐软件上也起到了重要的作用，用户可以使用推荐系统来对自己喜爱的音乐类型进行分析，使得用户可以从软件的推荐中发现潜在的喜爱音乐。 网易云音乐记录了所有用户的听歌记录以及用户对歌曲的反馈，在这一基础上计算出不同用户在歌曲上的喜好相似度，从而给用户推荐和他有相似听歌爱好的其他用户喜欢的歌曲。

### 1.3.4 社交网络

社交网络服务类型的产品是当今社会中使用的最为频繁的应用产品了，在各个社交网络应用中，推荐系统被用来满足用户多方面的要求，社交网络中的个性化推荐技术主要应用在以下三个方面：

（1）利用用户的社交网络信息对用户进行个性化的商品推荐。在各大社交平台上，经常会为用户推荐一些相关的商品，这些商品的推荐基于对用户的分析建模，从而保证用户与商品具有一定的关联性。（2）信息流的会话推荐。社交平台也会给用户推荐一些可能会感兴趣的信息，如新闻类的信息，一般是根据用户的浏览习惯来完成推荐。（3）给用户推荐好友。好友推荐要按照用户的交友习惯，针对这种情况，社交网络一般会计算用户的潜在好友和真实好友之间的相关性，如果相关性达到一定的阈值，则进行推荐。

### 1.3.5 个性化阅读

在一些广泛使用的阅读软件中，经常有一个“个性化阅读”的概念被提及。个性化阅读即为为每一位用户设计定制其需要的、感兴趣的、关注的阅读内容，这些内容是要以人为本的，每一位用户的所看到的都会与其他人不同。在现如今的各大阅读软件，这一理念得到广泛运用。 Google Reader是一款流行的社会化阅读工具。它允许用户关注自己感兴趣的人，然后阅读所关注用户分享的文章。个性化阅读工具Zite则是收集用户对文章的偏好信息，给用户推荐之前看过的、喜欢的类别的文章。

### 1.3.6 基于位置的服务

       在很多应用中，都会基于地理位置来为用户提供推荐服务，这一点在现如今的外卖软件中尤其关键，如美团和饿了么都会基于用户的地理位置和喜好偏向等各个方面，来为用户推荐适合的店家和商品。

## 1.4 组织结构

本文的结构安排如下：首先在第1章中，我们对项目的背景，项目目的以及具体的应用场景进行了简要介绍；然后在第2章中，从功能性与非功能性两个角度进行本项目的需求分析；在第3章中，对推荐系统的历史发展进行了调研；在第四章中，从基于图神经网络、基于知识图谱和基于协同过滤这三类推荐系统领域的多种算法出发，并对其他算法进行分析，调研推荐系统领域近几年的最新进展和研究方向；最后在第五章中，对方案相关的深度学习及深度强化学习领域的方法进行介绍，对WTP场景和数据进行分析，最终给出合理的解决方案。

# 第2章 需求分析

## 2.1 功能性需求

本项目的功能性需求是根据航空公司的用户的历史购买记录、航空公司的机票历史定价及销量、各大主流平台的社交网络信息，以及天气等实时信息，针对当下航空公司的定价来预测每个用户的未来行为，从而可以实现基于用户画像的旅客支付意愿评估，建立出估算旅客支付意愿推断的算法模型，并为旅客的购买意愿进行分级。

## 2.2 非功能性需求

### 2.2.1 数据规模约束

大部分的居民一年的出行次数是有限的，对于航空公司的旅客而言更是如此，因而平台所提供的航空公司旅客的历史购买信息数据可能会非常稀疏。对于具体用户，其所产生的出行序列信息可能不足，这就使得本项目在进行的时候，要考虑如何用较小规模的数据去训练并拟合模型，同时要考虑最终的预测效果等性能因素。在一定的数据规模下，运用一定的模型构建，是否可以完成满足准确率要求的预测，这是一个在实际研发中需要考虑的问题。如果因为数据量过小，模型不能够达到拟合，最终模型的准确率不能达到保证，那么项目的存在便会失去意义。那么在项目的开发阶段需要考虑如何解决数据规模较小的问题，并能够给出一个较为准确的最终结果。

### 2.2.2 耗时约束

对于一个功能完善的分类模型，时间消耗也是一个需要重视的问题。对于最终旅客支付意愿结果的分类，需要的各类资源和计算都是巨大的，在一定数量机器、一定的时间基础上，能否达到预先设定的准确率合格的目标，这是另一个在实际研发中必须考虑的问题。如果当算法模型的分类时间长度过长，那么项目的存在便会失去意义，与其节约时间成本的需求相矛盾。那么在项目的开发阶段需要考虑到优化算法的执行效率，使得最终的分类时长在一个可接受的范围内，并能够给出一个较为准确的最终结果。

### 2.2.3 准确率约束

对于一个合格的模型，准确率要有一定的保障。本项目是要对航空公司的旅客的支付意愿进行评级和分类，按照合同要求，本项目对旅客支付意愿的分类准确率要达到60%以上，要能给出一个较为准确的最终结果。

# 第3章 推荐系统的发展历史调研

自从20世纪90年代中期关于协同过滤的第一篇论文出现以来[1]，[2]，[3]，推荐系统已经成为一个重要的研究领域。1992年以前应用于大规模信息系统的是基于内容的过滤算法，1992年开始，协同过滤作为一种新的信息过滤方法出现在研究者面前。1994年协同过滤方法被首次应用于推荐系统中，这是推荐系统领域的开端。2003年，亚马逊提出的基于商品的协同过滤是推荐系统领域的又一里程碑事件。此后随着深度学习的发展，深度学习与推荐系统的结合逐渐成为了新的研究方向。[12]

## 3.1 基于内容的过滤推荐算法

基于内容的过滤的推荐方法起源于信息检索[4]，[5]和信息过滤研究[6]，对传统信息检索方法的改进在于对用户画像（包含用户的品味、偏好和需求等）的利用。由于一些基于文本的应用的重要性，当时的基于内容的过滤主要是推荐包含文本信息的对象，如文件、网址和Usernet新闻消息等。

文本信息的内容一般被表示为关键词的形式。信息检索中最著名的关键词度量是TF-IDF[5]，通常利用TF-IDF来计算文本信息内各个词的权重以获取关键词。通过使用关键词技术对用户历史信息的内容进行分析，来获得基于内容的用户画像。将用户画像和文章内容都使用TF-IDF向量来表示后，通过信息检索方法中的相似性度量（如余弦相似性度量等）进行匹配推荐。

除此之外，在基于内容的过滤中统计和机器学习方法（如贝叶斯分类器[7]、[8]）和各种机器学习技术（包括聚类、决策树和人工神经网络[8]）也是可行的。此方法从原始数据中学习模型，从而生成推荐结果。

## 3.2 协同过滤推荐算法

与基于内容的推荐方法不同，协同过滤方法是根据之前其他用户对商品的评分来预测该商品对目标用户的价值，分为基于记忆的协同过滤和基于模型的协同过滤。[9]

### 3.2.1 基于记忆的协同过滤

基于记忆的协同过滤是基于所有被评分过的商品集合来进行推荐。按照计算相似度的主体，又分为基于用户的协同过滤和基于商品的协同过滤。

基于用户的协同过滤的核心思想是，每个用户都会有对于一系列商品持相似态度的邻居用户，首先为活跃用户计算用户间的相似度，并为其选出一个邻居子集，确定这些相似的用户喜欢的商品集，对此商品集中的每个未评分商品利用子集中邻居们的评分预测目标用户的评分，最后基于这些预测的评分给出一个top-N的推荐商品集合。基于用户的协同过滤一般采用加权和的方式来对邻居评分进行聚合。其中用户间的相似性度量一般采用Pearson 关联相似性度量或余弦相似性度量。

基于商品的协同过滤[10]则是基于其他用户的历史评分，计算用户已经评分过的商品与目标商品的相似度，选出k个最相似的商品，同样采用加权平均的方法对目标用户对这k个商品的历史评分进行计算，从而得出预测评分。其中商品间的相似性度量一般采用余弦相似性度量、关联相似性度量和修正余弦相似性度量。

### 3.2.2 基于模型的协同过滤

与基于记忆的协同过滤使用整个用户-商品数据库来生成预测、使用如最近邻等统计学方法寻找邻居不同，基于模型的方法则是先开发一种基于用户的模型，从而生成预测。模型的种类包括概率模型（如贝叶斯网络等）、聚类模型、基于规则的方法（如关联规则）、回归模型、分类模型和LDA模型等。

基于模型的协同过滤的发展离不开Netflix 2006年举办的大奖赛。2006年，Netflix宣布了一项竞赛，任何人只要能够将它现有电影推荐算法Cinematch的预测准确度提高10%，就能获得100万美元的奖金，而Cinematch的初始均方根误差（Root Mean Squared Error，简称RMSE）得分为0.9525，需要将其降低到0.8572或更低。比赛开始一年之后，队伍Korbell融合了107种算法，以8.43%的提升赢得了第一个半程奖。他们这107中算法中的2种最有效的方法是奇异值分解（Singular Value Decomposition，简称SVD）和受限波尔兹曼机（Restricted Boltzmann Machines，简称RBM）。SVD能取得0.8914的RMSE得分，RBM则是0.8990，将这两种方法线性融合后RMSE得分能达到0.88。尽管有着数据集相较真实数据太小、没有考虑新评分的产生等缺点，但最终这两种方法成功地应用到实际产品中，而且直到现在还作为Netflix系推荐统的一部分在运行。两年后最终的获胜队伍BellKor's Pragmatic Chaos将数百种预测模型融合到一起，最终突破了0.8572的临界线。但是这些在比赛数据集上优胜的算法，到了线上系统却表现的不够出色，最终并没有被Netflix所使用。

## 3.3 混合推荐算法

混合推荐算法就是将多种推荐算法通过加权融合、切换、混合、特征合并、级联、特征增强和元级混合等方法共同使用，从而提高推荐性能，减少冷启动、流行推荐等问题。[11]，[13]

加权融合就是基于对场景的判断对各个推荐算法的结果进行加权地合并，如当场景中评分者少时，协同过滤的推荐效果就会很低，权重应该相应减小，避免对最终推荐性能的较大影响。

切换就是根据问题背景和实际情况采用不同的推荐技术，如系统首先使用基于内容的推荐方法，当基于内容的方法无法给出有效地推荐时，那么就会尝试协同过滤的方法。但切换并不能完全避免冷启动问题，因为基于内容的方法和协同过滤都有新用户的问题。切换的关键在于设置一个合适的标准，从而提升整体性能。

混合则是简单地将多种不同的推荐算法推荐出来的结果混合在一起，其难点是如何重排序。混合也不能解决新用户带来的冷启动问题。

特征合并是将来自不同推荐数据源的特征组合起来，由另一种推荐技术采用。如将来自协同过滤的产生信息（用户评分）作为额外特征信息，然后在这个增强数据集上采用基于内容的过滤来产生推荐；也可以将内容信息作为协同过滤的特殊维度（即作为来自虚拟用户的评分）。

级联是先试用一种方法对候选对象进行粗排序，然后用第二种方法对推荐结果进行细化。级联的关键在于确保第一种方法优先级低于第二种，否则反而降低性能，同时需要对结果进行有意义的排序方式。

特征增强是先利用第一种方法产生商品的评分或分类，然后再将这一信息整合进下一种方法的输入中。Libra系统对亚马逊上的图书使用朴素贝叶斯文本分类器来进行基于内容的推荐，然而这个文本中本身包含着亚马逊内部协同过滤系统产生的相关作者和相关标题信息，这就是特征增强的混合模型。特征增强与特征合并很类似，区别在于特征增强的输出是用作第二个推荐方法的输入，而特征合并将两个方法的特征表示合并。

# 第4章 推荐系统的最新进展调研

## 4.1 基于图神经网络的推荐算法

近些年来，简单的基于神经网络的推荐系统已经不能满足用户的诸多需求了，例如，对于现在很多的实际应用场景，数据都是从图等非欧式空间上（此处的图指数据结构中的图，一个图指一些顶点的集合，这些顶点通过一些边连接）生成和获取的，因此难以用普通的神经网络去处理这些数据。图数据的复杂性对现有的机器学习算法提出了重大挑战，这是因为图数据是不规则的。每个图都有一个大小可变的无序节点，图中的每个节点都有不同数量的相邻节点，导致一些重要的操作（如卷积等）在图像上很容易计算，但不再适合直接用于图域。此外，现有机器学习算法的一个核心假设是实例彼此独立。然而，对于图数据来说，情况并非如此，图中的每个实例（节点）通过一些复杂的链接信息与其他实例（邻居）相关，这些信息可用于捕获实例之间的相互依赖关系。

因此，鉴于存在着一些传统的神经网络无法处理的数据，在深度学习的成功推动下，研究人员借鉴了卷积网络、循环网络和深度自动编码器的思想，定义和设计了用于处理图数据的神经网络结构，由此一个新的研究热点——“图神经网络（Graph Neural Networks，简称GNN）”应运而生。

最早的图神经网络起源于Franco博士的论文[14], 它的理论基础是不动点理论。给定一张图，每个结点都有其自己的特征，文中用表示结点的特征；连接两个结点的边也有自己的特征，文中用表示结点与结点之间边的特征；GNN的学习目标是获得每个结点的图感知的隐藏状态，这就意味着：对于每个节点，它的隐藏状态包含了来自邻居节点的信息。GNN通过迭代式更新所有结点的隐藏状态来实现，在时刻，结点的隐藏状态按照如下方式更新：

上面这个公式中的就是隐藏状态的状态更新函数，也被称为局部转移函数。公式中的指的是与结点相邻的边的特征，指的是结点的邻居结点的特征，则指邻居结点在时刻的隐藏状态。并且函数是对所有结点都成立的，是一个全局共享的函数。图神经网络的任务就是拟合这个函数，从而训练出模型，完成接下来的任务。

### 4.1.1 基于会话的图神经网络推荐算法

基于会话的图神经网络推荐算法（SR-GNN），是由来自中科院的吴书等人于2019年在AAAI会议上发表的“Session-based Recommendation with Graph Neural Networks”一文中提出的。[15]

**1.算法介绍**

大多数现有的推荐系统都假定用户画像和历史行为被全部记录，但在很多情况下，用户的身份是未知的，且只能获取正在进行的会话中的用户的历史行为。​会话大多是匿名的且数量众多，而会话中点击所涉及的用户行为往往是有限的，因此很难预估用户表示。

现有的基于会话的推荐模型一般是对连续商品的单项转换进行建模并用作局部因素，忽略了上下文之间的转换，致使距离较远的商品的复杂转换被忽略。为了克服这些限制，文中提出了基于会话的图神经网络模型（SR-GNN），能够探索商品间丰富的转换并生成商品的潜在特征向量。

SR-GNN将会话机制与图神经网络结合，将分离的会话序列建模为图形结构的数据，并使用图神经网络来捕获商品间的复杂转换。为了生成基于会话的推荐，SR-GNN不依赖用户表示，只使用从潜在特征中获取的会话向量。

**2.算法分析**

每个会话序列都被建模为一个定向图，图中每个节点代表一个商品，图中的边代表用户点击头结点对应的商品后点击了尾结点对应的商品。根据节点被访问的次数来给边一个正则化的权重，计算方式为这条边出现的次数除以头结点的出度，然后将所有商品节点整合成一张会话图。

从这张会话图中提取节点邻居的潜在特征向量并作为门控GNN的输入。GNN中的更新门和重置门决定保留特征向量的哪些信息，然后在重置门的控制下，根据前一状态和当前状态来构建候补状态。在更新门的控制下，以前的隐藏状态和候选状态的结合产生节点的最终状态。然后更新会话图的所有节点直到收敛，获得会话图中所有节点的最终的潜在特征向量。

SR-GNN将节点表示嵌入进会话序列，将会话序列的最后一次点击作为局部特征，表示当前兴趣；通过聚合所有会话序列中的节点向量获得会话图的全局特征，考虑到不同节点的信息有不同的优先级，采用软注意力机制来更好的表示长期偏好。然后通过线性转换将局部特征和全局特征整合为混合特征，即同时考虑当前兴趣和长期偏好。最终对所有候补节点，将节点向量和会话表示的乘积作为分数，应用softmax函数获得模型的输出向量。

### 4.1.2 基于堆叠和重构图卷积网络的推荐算法

基于堆叠和重构的图卷积网络的推荐系统算法（STAR-GCN），是由来自香港中文大学的张佳妮等人于2019年在IJCAI会议上所发出的论文“STAR-GCN: Stacked and Reconstructed Graph Convolutional Networks for Recommender Systems”一文中提出的。[16]

**1.算法介绍**

推荐系统中核心的数学问题就是矩阵分解，假设有个用户，个商品，推荐算法的目的是根据这个矩阵已有的值去填充未知的值。现在常用的经典解决方法是基于矩阵分解（Matrix Factorization，简称MF），每一个已知的得分是通过用户和商品的隐因子表征相乘得到的。近年来，深度学习的发展尤其是图卷积网络（Graph Convolutional Network，简称GCN）给这个问题带来了一些新的思路。GCN将卷积的定义从规则网格推广到不规则网格（如图结构）。GCN通过一个局部共享参数算子来生成节点表示。一个图卷积算子通过转换或者聚合某个节点的局部领域节点的特征来生成该节点的表征。通过堆砌多个图卷积算子和非线性函数，就可以构建了一个深度神经网络，从而抽取这个图远处的特征。因为局部邻域集可以看作卷积核的接收场，这类领域聚合方法称为图卷积。

Monti等人针对推荐系统首次提出了基于GCN的方法，在他们的方法里，GCN用于将用户-用户图和商品-商品的两个图信息聚合，用户和商品的隐因子在每次聚合的时候更新，GCN和矩阵分解的目标函数通过组合用在了这个模型里。此后，Berg等人提出了图卷积矩阵完成模型（GC-MC），GC-MC模型直接将用户和商品的关系描述为一个二部图，模型中将两个多连接卷积层用于聚合用户和商品的特征，并通过预测边的标签来估计得分。GC-MC模型虽然取得了比较好的成绩，但GC-MC模型有两个明显的限制：为了区分每一个节点，GC-MC模型使用独热向量作为节点输入，这使得对大规模的图的使用受到了限制；另外，模型不能预测训练过程中没有出现过的商品和用户，因为不能用独热向量表示没出现过的节点，这就会导致我们熟知的冷启动问题。

针对这些问题，堆叠重构图卷积网络算法（STAR-GCN）在一些方面上做出了改进，从而可以解决这些问题，不同于GC-MC算法，STAR-GCN算法通过端到端的模式直接学习用户和商品的低维表征，为了提升模型能力，同时针对冷启动问题，STAR-GCN对部分或者整个用户和商品的表征进行了掩盖，在训练阶段通过一小块图的编码解码器来重构盖住的部分。

**2.算法分析**

STAR-GCN通过堆叠或循环支持两个连续块之间的两种不同类型的组合，选取的方式主要取决于是否在块之间共享参数。通过堆叠，可以连续放置带有单独参数集的多个编码器-解码器块。通过循环，可以展开单个编码器-解码器块，从而实现在所有块之间共享同一组参数，减少了总的内存使用量。

为了让网络扩展到大型图，算法通过查表寻找将每个节点映射到一个低维的向量，。在网络中通过端到端训练得到。但是简单的替换独热类型向量还是没法解决冷启动问题，因为新节点的嵌入不在训练数据里。于是在算法中采取了一种方法，即随机掩盖一定百分比的输入节点，然后重建新的节点的表征。在每一个簇训练中，会隐藏住20%的节点，然后重建这部分节点的表征，对于这部分盖住的节点，做以下操作：(1) 以概率将节点表征初始化为0，(2) 以的概率保持不变。

训练这个被标记的表征机制有两个优势。第一，在训练过程中，它可以学习到没被观测到的节点，在冷启动场景，可以初始化新节点为0，通过多个GCN编码器/解码器块逐步完善估计的表征。例如，第一个块通过利用邻域数据（或节点属性，如果可用）来预测新节点的表征。然后，将预测的表征内容馈送到第二个块，以预测评分和精确的表征内容。在这一过程中，评级和表征预测损失是共同优化的。

因此，STAR-GCN可以通过迭代优化嵌入来解决冷启动问题，并且与GC-MC根本不同。 其次，STAR-GCN导致转导环境的改善。 在训练阶段，部分节点嵌入被掩盖，并要求网络重构这些掩盖的嵌入，这要求网络有效地编码用户和项目之间的关系。因此，重建损失充当了多任务调整器，可提高主要评级预测任务的性能。这些就是该算法STAR-GCN较之前算法的进步和提升。

### 4.1.3 基于图卷积神经网络的web领域推荐算法

基于图卷积神经网络的web领域推荐系统算法，是由来自斯坦福大学的Rex Ying等人于2018年在SIGKDD会议上发表了“Graph convolutional neural networks for web-scale recommender systems”一文中提出的。[17]

**1.算法介绍**

图卷积神经网络算法大多数都是处理端到端的计算，需要在整个图上训练。但是在工业运用中，有很多的图的数据量非常之大，如果在整个图上进行训练的话，就会需求大量的计算能力，从而很难将算法扩展应用到实际的大规模工业应用上。算法中模型设计的目的是为了得到节点的嵌入表示，优化目标是正例内积最大化和负例内积小于给定阈值。因此，关键是如何求得图嵌入。

常见的图嵌入的方法，有node2vec、embedding等。这两种都是无监督方法，且仅根据网络结构，不考虑节点特征，因此，也正是这两种方式的普适性就导致对于特定的分类预测问题，效果表现并不好。除此之外，作为一种端到端的半监督学习方式，GCN在分类问题中，不仅考虑了节点特征，并且将特征和网络结构整合，通过将每个节点的邻居节点信息作用到自身，相当于一个消息传播的过程，同时也存在着难以调参、很难收敛的问题，而且传统的GCN在全图上做操作，对于数据量较大的图而言，有很大的计算限制。因此该算法提出了一个局部卷积的概念，不在全局的图上优化模型，而是给特定的节点形成一个包含有限领域节点的子图，在子图上构造局部卷积，然后不同节点共享同样的局部卷积参数，从而解决这一存在的问题。

**2.算法分析**

算法中的核心部分是卷积模块，采用逐点优化的算法，所以输入是当前计算的节点的嵌入，以及它所对应的领域节点的嵌入。而具体的卷积操作其实就是一些全联接构造成的映射。算法的操作先是对节点的领域节点经过映射后，再利用权重池化函数让输出的维度和输入保持一致，生成所有领域节点统一的嵌入向量。

然后需要进行的操作是对节点嵌入的更新，直接把上一层或者初始的嵌入和领域节点嵌入一起连接起来，再加上一层全联接就可以生成新的节点嵌入。整个网络结构是属于一个多层卷积网络，输入是节点和节点领域嵌入，输出是新的节点嵌入形式。

经过算法的操作，可以将特定的节点形成一个包含有限领域节点的子图，在子图上构造局部卷积，然后不同节点共享同样的局部卷积参数，从而解决问题。

### 4.1.4 基于图嵌入的电子商务推荐召回算法

基于图嵌入的电子商务推荐召回算法，是由来自阿里巴巴的Wang J等人于2018年在SIGKDD会议上发表的“Billion-scale commodity embedding for e-commerce recommendation in alibaba”一文中提出的。[18]

1.**算法介绍**

在淘宝平台的推荐部分，有三个方面存在着问题：第一方面是可扩展性，已有的推荐算法可以在小数据集上有不错效果，但是对于淘宝平台的数据量，10亿用户、20亿商品这样海量的数据集上效果差；第二方面是稀疏性，由于商品的数量级较大，且大多数的用户只会购买极少种类的商品，即仅与小部分商品交互，因此难以训练准确的推荐模型；第三方面是冷启动问题，商品上新频繁，并且这些商品在刚开始售卖的时候并没有用户行为，因此预测用户对这些商品的偏好是十分具有挑战性的。

现在业界针对海量数据的推荐问题通用框架是分成两个阶段，即召回和排序。在召回阶段，会针对不同的用户生成一个商品候选集，它的商品信息会与用户接触过的每个商品具有相似性；接着在排序阶段，通常会训练一个深度神经网络模型，它会为每个用户根据他的偏好对候选商品进行排序。基于图嵌入的电子商务推荐召回算法主要通过用户点击的行为历史构建商品图，学习图上所有商品的嵌入。同时为了解决可扩展性、稀疏性和冷启动问题，在商品图的基础上，增加商品的额外信息（如种类、品牌和价格等）丰富商品表征力度。文中提出了基于图的嵌入（BGE）、具有边信息的图嵌入（GES）和具有边信息的增强图嵌入（EGES）三个模型，核心的任务是计算所有商品之间的相似度。

**2.算法分析**

    在淘宝的实际数据中，用户的购买记录是一串包含商品信息和时间戳的序列，因此首先要考虑，如何将这样的一个序列，构建成一个关于用户的商品图。因为不可能使用一个用户的整个历史购买记录，那样的话计算开销和存储开销会非常大。又因为一个用户的兴趣趋向于随时间漂移，因此，可以通过设置了一个时间窗口，只选择用户在该窗口内的行为，来构建成用户的商品图。这被称为是基于短序列的用户行为。在论文中将该时间窗口的设置为一个小时，通过设置一个时间窗口，选取短序列用户的行为信息，从而完成商品图的构建。

论文的作者认为在实际使用中，数据肯定存在噪声，因此想出了三个方面，来对数据做一些处理，来消除噪声：首先用户在点击后停留的时间少于1秒，可以认为是误点，需要移除；其次还有一些过度活跃用户，三月内购买商品数超过1000，或者点击数超过3500，就可以认为是一个无效用户，也需要去除；还有一些商品频繁的修改商品内容的，造成前后同一id的商品信息不对应，也需要移除。

通过前面定义用户，构建出商品的加权有向图，使用Deep Walk算法，经过随机游走得到序列，输入skip-gram模型中进行训练，学习图中的节点的嵌入。利用随机游走在图中生成序列的方法是，假设是的邻接矩阵，表示从节点指向节点的加权边，那么随机游走的转移概率定义为：

通过这个概率值就可以从图中获取到大量的序列。再接着就利用skip-gram模型来学习嵌入的值,把每个商品的id看成是词，整个序列看成句子。优化目标如下：

通过使用论文中提出的BGE模型，能够将商品映射到高维向量空间，并考虑协同过滤没有考虑的用户序列关系。但是依然没有解决冷启动的问题。为了解决冷启动问题，作者提出使用边界信息（品牌, 价格等）赋值给不同的商品。因为边界信息相同的两个商品，理论而言会更接近。通过DeepWalk算法得到商品的游走序列，同时得到对应信息(品牌, 价格)序列。然后将所有序列放到word2vec模型中进行训练。针对每个商品，将得到：商品嵌入向量，种类嵌入向量，品牌嵌入向量，价格嵌入向量等。其中，表示整体的嵌入向量，表示以上每一组的嵌入向量。

计算公式如下：

该算法在离线和在线上都进行过测试，可以观察到合并边信息对于图嵌入非常有效，准确率较之前有着巨大的提升，且该算法可以解决扩展性不足，和冷启动等以前存在的问题，可以说是效果较好。

### 4.1.5 基于记忆增强图神经网络的序列推荐算法

基于记忆增强图神经网络的序列推荐算法（MA-GNN），是由马辰等人于2020年在AAAI会议上发表的“Memory Augmented Graph Neural Networks for Sequential Recommendation”[19]一文中提出的。

**1.算法介绍**

在序列推荐中，常常会出现因为信息不完全而导致最终构建的模型不健全，对结果的预测和分类存在着些许的偏差。并且在序列推荐中，除了所有常规推荐模型都捕获了的一些较为常规的用户兴趣外，还有三个额外的重要因素可以用来进行建模：用户短期兴趣，用户长期兴趣和项目共现模式。这三个因素之所以具有各自的重要性，主要还是以为它们分别代表的意义：用户短期兴趣描述了在给定短期内几个最近访问的项目和用户的偏好之间关系；用户的长期兴趣捕获了较早访问的项目与用户将来将访问的项目之间的长期依赖关系；商品共现模式说明了例如手机和屏幕保护器等共同相关商品的共同出现的现象模式。但是对于现有的大多数算法，都未将这三种因素结合起来构建模型，或是只取其中之一之二的因素来构建模型，因此对于最终构建的模型可能不足够优秀。

MA-GNN模型将这三种因素集合到一个模型中，并阐述了算法如何将这三种因素结合到一起的，论文实验发现，该算法所提出的模型在五个真实数据集中的表现大大优于最新的顺序推荐算法。

**2.算法分析**

为了整合上述因素，该算法提出了一种记忆增强图神经网络（MA-GNN）以解决顺序推荐任务。它由一般兴趣模块，短期兴趣模块，长期兴趣模块和项目共现模块组成。在一般兴趣模块中，算法采用矩阵分解项对一般用户兴趣进行建模。在短期兴趣模块中，算法使用图神经网络汇总商品的邻居，以在短期内形成用户意图，这样可以在短期内捕获本地上下文信息和结构。为了模拟用户的长期兴趣，算法使用共享的键值存储网络来根据用户的长期项目序列生成兴趣表示。这样，在推荐商品时将考虑其他具有相似偏好的用户。为了结合短期和长期利益，算法中的图神经网络框架中还引入了门控机制，该机制类似于长短期记忆网络 （Long Short-Term Memory，简称LSTM）。在项目共现模块中，算法应用双线性函数来捕获在项目序列中一个接一个出现的紧密相关的项目，从而结合多种因素，完成模型的构建。

### 4.1.6 算法优缺点对比

基于图卷积网络的推荐系统我们一共调研了四种算法，对于这四种算法，我们做了一系列的横向对比，分析了每种算法的优势和劣势。

SR-GNN模型是将会话机制和图神经网络结合的方法，相较于其他基于图神经网络的算法可以有效地保持会话数据的自然特性（即会话结构），并最大程度地避免本地信息丢失，因此更关注会话数据的动态和局部方面，从而提供更可靠的建议。但文中主要侧重于将长期偏好和当前目标相结合，对于本项目而言，机票购买数据的高稀疏性可能导致会话数据本身的不连续，无法实现项目目标。

基于堆叠和重构图卷积神经网络的推荐系统算法提出了一个新模型，即STAR-GCN，这个推荐模型是基于之前的一个名为GC-MC的模型的提升和改进。STAR-GCN 可以通过迭代优化嵌入来解决GC-MC模型存在的冷启动问题，其次，STAR-GCN 导致转导环境的改善，从而可提高主要评级预测任务的性能。

基于图卷积神经网络的 web 领域推荐系统算法提出了一种新的卷积思路，即局部卷积。这种方法，不在全局的图上优化模型，而是给特定的节点形成一个包含有限领域节点的子图，在子图上构造局部卷积，然后不同节点共享同样的局部卷积参数，从而解决了一些数据图较大的情况，大大减少了训练时间消耗。但是这种算法存在的缺点是对于一些遍及整个图的数据问题，该算法不能很好的进行解决，因此该算法有一定的局限性。

基于图嵌入的电子商务推荐召回算法解决了海量数据的电子商务召回问题，该算法通过用户点击的行为历史构建商品图，学习图上所有商品的嵌入，从而解决了一些推荐系统存在的难扩展、稀疏性、冷启动问题。但是，由于算法基于商品图的基础上，增加了商品的额外信息（例如价格，品牌等）来丰富商品图表征力度，使得存储花销得到了增大，因此该算法较同种类型的算法存在着中间数据量大、难以训练的问题。

基于记忆增强图神经网络将用户短期兴趣，用户长期兴趣和项目共现模式三种特殊因素和常规用户兴趣因素结合在一起，用来构建模型，并完成Top-K的预测，算法所提出的记忆增强图神经网络在五个真实数据集效果都优于对比的算法，提高了模型的预测准确率。但是由于引入多种因素，使得模型框架比较复杂，难以构建完成。

## 4.2 基于协同过滤的推荐系统算法

### 4.2.1 神经图协同过滤推荐算法

​ 神经图协同过滤模型推荐算法（NGCF），是由王翔等人于2019年在SIGIR会议上发表的“Neural Graph Collaborative Filtering”一文中提出。NGCF模型将用户-商品交互以二部图结构集成进向量表示过程，并利用用户-商品图来传播向量表示，这使得在用户-商品图上的建模具有高阶连通性，有效地将协同信号显式加进嵌入过程。[20]

**1.算法介绍**

​ 个性化推荐的核心就是根据用户历史行为（购买、点击等）来预估用户购买一个商品的可能性。协同过滤（CF）是假定行为模式相近的人会呈现相似的商品偏好，为了实现协同过滤方法，一般将用户和商品参数化并重建历史交互，然后基于这些参数预测用户偏好。从早期的矩阵分解到目前流行的基于深度学习的方法，都是从用户或商品的已有特征（ID、属性等）映射成向量表示。这些方法的固有缺点是没有将用户-商品交互过程中潜在的协同信号加入向量表示中去，由此产生的向量表示不能充分实现协同过滤方法的效果。

​ 与现有的协同过滤方法不同，NGCF模型在参数化过程中利用了至关重要的协同信号，也就是用户-商品交互中揭示用户（或商品）间相似性的潜在因素。但由于真实应用中用户-商品交互可以轻易地到达百万级规模，很难从交互中提取想要的协同信号。而NGCF模型利用用户-商品交互中的高阶连通性，即在用户-商品交互图中连通的节点间与目标用户连通距离大于1的节点集，将协同信号嵌入交互图结构中解决了这个问题。

​ 为了对高阶连通性信息建模，NGCF模型构造了一种在图上递归传播向量表示的神经网络方法。在算法的嵌入传播层上，通过聚合与之产生交互的商品（或用户）的向量表示来更新该目标用户（或商品）的向量表示。通过多个嵌入传播层的叠加，使用这些向量表示来捕捉交互间的协同信号。

**2.算法分析**

NGCF模型一共由三个部分组成：用来提供并初始化用户和商品向量表示的嵌入层，用来向用户和商品向量表示中添加高阶连通性关系的多个嵌入传播层，和将不同嵌入传播层的向量表示聚合并输出用户向量对间的亲和度分数的预测层。

为了沿图结构捕捉协同过滤信号并更新节点向量表示，NGCF模型基于CNN的消息传递结构构建嵌入传播层。嵌入传播过程主要包括以下两个操作：信息构建和信息聚合。信息构建使用可训练权重对商品向量表示和用户-商品交互进行加权，并乘以信息在传播过程中的衰变系数，信息聚合则是聚合目标用户的邻节点传播的信息以更新目标用户的向量表示。

嵌入传播层分为一阶传播和高阶传播。一阶传播中，与用户交互的商品可以为用户的偏好提供直观证据，相应地，消费该商品的用户也可以视为商品的特征，并用来计算商品间的协同相似度。作为邻节点传播信息的补充，NGCF模型考虑了自传播的情况，将自传播的信息与所有邻节点的信息加和并经过激励函数LeakyReLU，得出一阶传播层输出的用户节点向量表示。商品的节点表示采用完全一致的方法。高阶传播与一阶传播方法基本类似，高阶传播无缝地将协同信号融入了表示学习过程中。

最终预测层将各个层的向量表示串联，作为最终的向量表示。使用目标用户和目标商品的向量表示的内积来评估用户对商品的偏好。

### 4.2.2 基于信任信息的协同过滤推荐算法

基于信任信息的协同过滤推荐算法（TrustSVD++），是由郭贵冰等人于2015年在AAAI会议上发表的“Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings”一文中提出的模型。[21]

**1.算法介绍**

协同过滤算法一直存在两个问题：数据稀疏和冷启动。信任信息+推荐系统已经证明了有潜力提升当前推荐系统的性能并解决这两个基本问题，但事实上，最新相关研究的最优表现仍然低于仅用用户-商品评分的最先进模型。经过对数据集分析发现，信任信息本身也很稀疏，但与评分信息是互补的。在信任互补的关系下，用户的评分与社交邻居的评分呈弱正相关；在信任关系下，呈强正相关。

​ 因此作者提出了一种由SVD++改进的基于信任信息的推荐系统算法模型，该模型既考虑了用户-商品评分的显式影响（真实评分值）和隐式影响（哪个用户给哪个商品评的分），也考虑了信任信息的显式影响（信任信息）和隐式影响（谁信任谁）。

**2.算法分析**

TrustSVD模型是基于SVD++改进得来。SVD++不是直接利用用户/商品的潜在特征来进行评分预测，它的计算方法是：用户对商品的评分的预测=数据集中所有打分平均值+商品的偏差+用户的偏差+商品的d维潜在特征\*用户的特征向量。其中用户的特征向量是将用户的d维潜在特征与被评分商品集的影响相加。SVD++的结果表明，将被该用户评分的商品集合对该用户的隐式影响纳入考虑可以提升预测准确率。

TrustSVD模型先向SVD++模型中整合信任信息，将被信任的用户采用和被评分的商品一样的方式来获取隐式影响。为了降低模型复杂度，从评分矩阵和信任矩阵中计算用户特征向量时采用同一参数空间方便计算。文中改进了一种权重参数正则化的方法防止过拟合，原本该方法是对评分多的用户和被评分多的商品施加惩罚项，作者提出的方法改进是对于这些影响大的用户和商品惩罚减少，同时对冷启动的用户和商品更多正则化，目的是防止前者过拟合和后者欠拟合。

### 4.2.3 基于变体自动编码器的协同过滤推荐算法

基于变体自动编码器的协同过滤推荐算法，是由梁达文等人于2018在WWW会议上发表了一篇名为“Variational autoencoders for collaborative filtering”的论文中提出的。[22]

**1.算法介绍**

在深度学习中，线性的概率模型是最常见的，但是线性的概率模型也存在着一些问题，就是线性概率模型在有的场景下难以建模，或者是不能建模，从而造成最终的模型准确率出现问题。为了解决这种情况，作者提出了基于变体自动编码器的协同过滤算法。这种非线性概率模型能够超越线性因素模型的有限建模能力，虽然线性因素模型在很大程度上仍然主导着目前的协同过滤研究。算法引入一个具有多项式似然性的生成模型，并利用贝叶斯推理进行参数估计。尽管在语言建模和经济学中得到了广泛的应用，但在推荐系统中，多项式的可能性却很少受到关注。因此作者引入了一个不同的正则化参数，这被证明对实现有竞争力的表现至关重要。值得注意的是，存在一种使用模拟退火来调整参数的有效方法。所得到的模型和学习算法与最大熵鉴别和信息瓶颈原理具有信息理论联系。

**2.算法分析**

协同过滤是推荐系统中应用最为广泛的方法。协同过滤通过发掘用户和商品之间的相似模式来预测用户接下来将会买购买什么商品。潜在因素模型凭借其简洁性和有效性仍然是协同滤波研究的主流。然而，这些模型基本都是内部线性的，这限制了模型的表征能力。以往的研究工作证明向线性的潜在因素模型中添加精心设计的非线性特征能够显著提升推荐表现。最近，大量研究工作将神经网络引入协同滤波。

变分自编码推广了线性潜在因子模型，使我们能够在大规模推荐数据集上探索由神经网络驱动的非线性概率潜变量模型。虽然推荐通常被认为是基于大数据的问题，相反地，在该算法的思想中，认为推荐是一个十分具有挑战性的“小数据”问题：大多数用户只与很小一部分商品有联系，因此该算法的目标是整体上对每个用户的偏好作出明智的推断。为了能够充分地利用来自用户的稀疏信号并避免出现过拟合，算法中构建了一个能够在用户和商品之间共享统计强度的概率潜变量模型。并且根据实验表明，无论数据的稀缺性如何，采用原则贝叶斯方法都更加具有鲁棒性。

给定用户的点击历史，算法基于未归一化的预测多项式概率，对于x的隐式表达，其构成方式为，对于Mult-VAE,简单选取变分分布的均值，对于Mult-DAE。

引入自编码器的优点显而易见，对于大多数的隐式因数协同滤波模型，矩阵因子分解，当给定非训练集中的用户的点击历史记录时，通常需要执行某种形式的优化，以获得该用户的潜在因素。这使得自动编码器的使用在工业应用程序中特别具有吸引力，在工业应用程序中，以低成本和低延迟进行预测非常重要。

### 基于协同记忆网络的推荐算法

基于协同记忆网络的推荐算法（CMN），是由Ebesu等人于2018年在SIGIR会议上发表的“Collaborative memory network for recommendation systems”一文中提出的。[23]

**1.算法介绍**

在当今的运用环境下，协同过滤算法大致分为以下三类：第一种是基于邻域的方法，包含基于商品的和基于用户的，该方法根据用户的历史行为，计算商品与用户之间的相似度来进行推荐，只使用了局部的结构，而忽略了很多全局的信息；第二种是基于隐向量的方法，常用的就是矩阵分解及其各种变体。这种方法将用户与商品表示成一个大小为的评分矩阵，由于评分矩阵是稀疏的，可以将其转换成的用户矩阵与的商品矩阵相乘。这种做法虽然考虑了全局的信息，但是忽略了一些比较强的局部联系；第三种是混合模型方法，这种方法经常是对上述两个模型进行一系列的改进，常用的有因子分解机和SVD++方法。

在使用协同过滤模型推荐时，我们通常需要关注两点：第一点是需要考虑全局的信息，充分利用整个评分矩阵；第二点是需要考虑局部的信息，考虑用户或者商品之间的相似性，相似性高的用户或者商品给予更高的权重。其中提出了一种混合模型，该模型利用记忆网络和注意力机制用于具有隐式反馈的协同过滤。算法中提出了协同存储网络，这是一种深层架构，用于统一基于邻域和基于隐向量这两个模型，以非线性方式利用潜在因子模型的全局结构和和邻域模型局部结构的优势。

**2.算法分析**

该算法模型利用记忆网络和注意力机制用于具有隐式反馈的协同过滤，记忆组件允许读写操作来编码内部存储中复杂的用户和商品关系。关联寻址方案充当最近邻域模型，基于自适应用户商品状态来发现语义上类似的用户。神经注意机制对具有相似偏好的用户的特定子集施加更高权重，最后，根据局部信息与全局信息之间的非线性变换计算排名分数，以非线性方式利用潜在因子模型的全局结构和和邻域模型局部结构的优势。

而且，论文中表明CMN组件，记忆网络和三类协同过滤模型之间的紧密联系。CMN增强了外部记忆和神经注意力机制。记忆模块的关联寻址方案作为识别相似用户的最近邻模型。该注意力机制根据特定的用户和商品学习用户邻域的自适应非线性加权。输出模块利用自适应邻域状态与用户、商品记忆之间的非线性相互作用，推导出推荐值。综合实验结果也证明了协同存储网络在三个公开的有效数据集的表现优于竞争基准。

### 算法优缺点对比

对于协同过滤在推荐系统上的运用，我们调研了多篇论文，从中选出了四篇，对此四篇论文所提出的算法做了一系列的横向对比，分析了它们的优缺点。

NGCF模型将图神经网络和协同过滤结合，相较于其他协同过滤的算法可以从数据中提取用户-商品间的高阶协同关系，并加入用户或商品的向量表示中，更充分地实现了协同过滤方法的效果到，与目前最先进的协同过滤算法HOPRec、协同记忆网络（CMN）相比，有明显的推荐质量的提升。此模型基本满足本项目的需求，可以作为解决方案之一，但仍有冷启动和解决数据稀疏性等方面可以改进。

TrustSVD模型将信任信息与SVD++结合，信任信息作为评分数据的补充，考虑商品与用户间相互的隐式影响，有效地提升了SVD++的推荐性能，并在一定程度上解决了冷启动和数据稀疏问题。对于本项目而言，同时考虑评分的显式影响和隐式影响可以提升性能，但信任信息可能获取的难度非常大，无法获取充分数据的情况下实用性不足。

基于变体自动编码器的协同过滤算法提出了一个新颖的想法，设计了非线性的概率模型，该模型超越了线性模型所具有的建模能力，使得基于协同过滤推荐系统可以在更广的领域上得到运用，适合处理低成本和低延迟的一些问题。但是，基于非线性模型的协同过滤算法较线性模型相比来说，解释性较差，也难以进行移植运用。

基于协同记忆网络的推荐系统引入了协同存储网络，这是一种深层架构，用于统一基于邻域和基于隐向量这两个协同过滤模型，以非线性方式利用潜在因子模型的全局结构和和邻域模型局部结构的优势，并且在实验中可以发现，该模型的准确率较高。但是引入了协同存储网络也带来了一系列的问题，例如存储花销过大，计算量大，这使得模型训练较慢。

## 4.3 基于知识图谱的推荐系统算法

### 4.3.1 基于知识的路径循环网络算法

基于知识的路径循环网络算法（KPRN），是由王翔等人于2019年在AAAI会议上发表的“Explainable Reasoning over Knowledge Graphs for Recommendation”一文中提出的。[24]

**1.算法介绍**

将辅助信息加入推荐系统是长期以来的研究方向，如用户画像、商品属性等。在知识图谱（Knowledge Graph，简称KG）中，用户和商品的连接可以看作路径，这不仅可以解释解释为什么用户喜欢该商品，还可以揭示用户的喜好。因而将知识图谱纳入推荐系统具有重要意义。

因而作者提出了一种基于知识的路径循环网络算法模型（KPRN），该模型可以将主体（用户、商品）与关系的语义组合来获得路径表示，并利用路径的顺序依赖给出用户与商品交互的背后的原因。

**2.算法分析**

由于知识图谱中的主体包括用户和商品，将这两者按照商品的字符串匹配合并，整合进一个浓缩的知识图谱，并作为模型的输入。通过路径分析用户喜好时，可以将知识图谱中的路径有权重地组合来预测特定用户对特定商品的交互。

KPRN模型的输入为每个用户商品对的所有路径集合，输出为该用户和该商品产生交互的可能性的分数。整个模型有三个关键组成部分：嵌入层中每个都包含三个独立层：主体、主体类型、关系类型；LSTM层将这些元素顺序编码，目的是为了捕捉到这个路径上的主体们的组合语义；权重池化层将多条路径结合，并输出给定用户对目标商品的最终分数

KPRN模型将知识图谱与循环网络相结合，利用由主体与关系的语义组合来获取路径表示，能够在循环遍历过程中更全面地发现潜在的用户-商品关系，同时利用路径的顺序以来使得推荐有更高的可解释性。

### 4.3.2 基于知识图谱增强推荐的多任务特征学习算法

基于知识图谱增强推荐的多任务特征学习算法（MKR），是由Hongwei Wang等人于2019在WWW会议上发表的“Multi-Task Feature Learning for Knowledge Graph Enhanced Recommendation”一文中提出的。[25]

**1.算法介绍**

协同过滤通常在实际推荐场景中存在稀疏性和冷启动问题，因此，在使用一些协同过滤算法的时候通常使用辅助信息来解决问题并改善推荐系统的性能。知识图谱具有的特性可以在模型训练和预测的时候视为辅助信息的来源。MKR算法是一个深入的端到端框架，利用知识图谱嵌入任务来协助推荐任务，使得知识图谱嵌入任务和推荐任务这两个任务由交叉和压缩单元相关联，交叉和压缩单元自动共享潜在特征并学习推荐系统中的项目与知识图中的实体之间的高阶交互。在这个算法中，证明了交叉和压缩单元具有足够的多项式逼近能力，并表明MKR是推荐系统和多任务学习的几种代表性方法的通用框架。

但有一点要注意的是注意，这两个任务不是相互独立的，而是高度相关的，因为推荐系统中的商品可以与知识图谱中的一个或多个实体相关联。因此，项目及其对应的实体可能在推荐系统和知识图谱中具有类似的邻近结构，并且在低级别和非任务特定的潜在特征空间中共享相似的特征。为了模拟项目和实体之间的共享功能，算法在MKR中设计了一个交叉和压缩单元。交叉和压缩单元明确地模拟项目和实体特征之间的高阶交互，并自动控制两个任务的交叉知识转移。通过交叉和压缩单元，项目和实体的表示可以相互补充，协助两个任务避免拟合噪声和改进泛化。整个框架可以通过交替优化不同频率的两个任务来训练，这使得MKR在实际推荐场景中具有高度的灵活性和适应性，这也是MKR算法的最大的特点。

**2.算法分析**

MKR的算法框架由三个主要组件组成：推荐模块，知识图谱嵌入模块和交叉和压缩单元。

（1）推荐模块将用户和商品输入，并使用多层感知器（Multi layer Perceptron，简称MLP）和交叉和压缩单元分别为用户和项目提取短和密集的特征，然后将提取的特征一起送到另一个多层感知机中以输出预测的概率。

（2）与推荐模块类似，知识图谱嵌入模块也使用多个层感知器和交叉和压缩单元学习头部实体向量和关系向量，并在分数函数的监督下输出预测尾部实体向量，分数函数的作用是为预测结果打分，做出评价。

（3）推荐模块和知识图谱嵌入模块由专门设计的交叉和压缩单元桥接。交叉单元可以自动学习知识图谱中推荐系统和实体中商品的高阶特征交互，这样可以帮助推荐系统获取知识图谱模块的信息，更好地对商品和用户进行刻画，完成预测。

该算法通过对现实世界数据集的大量实验，证明了MKR模型在电影，书籍，音乐和新闻等几个应用方面的推荐取得了实质性的进步，超过了比较新的模型。即使用户项目交互稀疏，MKR也能够保持良好的性能。在现在的科研中，交替学习是一种较为创新和前沿的思路，其中如何设计两个相关的任务以及两个任务如何关联起来都是值得研究的、并且充满难点的方向。从实际运用和时间开销上来说，交替学习是介于依次学习和联合学习中间的：训练好的知识图谱特征学习模块可以在下一次训练的时候继续使用（不像联合学习需要从零开始），但是依然要参与到训练过程中来（不像依次学习中可以直接使用实体向量），因此交替学习也有着自己独特的特点。并且在该算法中，运用了知识图谱的嵌入来辅助训练和预测，知识图谱作为推荐系统的一种新兴的辅助信息，在各个模型的运用上都体现了非凡的价值，在本项目的实施中，也可能会有所运用。

### 基于知识图谱的用户喜好推荐算法

基于知识图谱的用户喜好推荐算法，是由Yixin Cao等人于2019在WWW会议上发表了一篇名为“Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences”的论文中提出的。[26]

**1.算法介绍**

知识图谱作为最近深度学习类中一种新兴的辅助信息框架，经常可以发现在模型的运用中有着正向的作用，作者因此提出了一种基于知识图谱的用户喜好推荐，目的是通过引入知识图谱这一辅助信息，来解决一些用户喜好的推荐问题。将知识图引入推荐系统，可以提高推荐的准确性和可解释性。然而，现有的方法在很大程度上假定知识图是完整的，并且只是在实体原始数据或嵌入的浅层次上传输知识图中的“知识”。这可能导致性能不够理想，因为实际情况下，想要完全地构建一个知识图是几乎不可能完成的，而且通常知识图缺少事实、关系和实体。因此，作者认为在将知识图引入推荐系统时，考虑其不完全性是非常重要的，因此与以往基于KG的推荐方法不同，他们将关系信息以KG的形式传递，从而了解用户喜欢某件商品的原因。例如，如果用户观看了由(关系)对于同一个人(实体)，那么推荐系统理应可以推断出导演关系在用户决策时起着至关重要的作用，从而有助于在较好的粒度上理解用户的偏好。

从技术上讲，该算法提出了一种新的基于翻译的推荐模型，该模型特别考虑了用户对某项翻译的各种偏好，然后结合多种转换方案，用知识图完成模型对其进行联合训练。该算法解决了用户的商品推荐和知识图的完整性不足两个问题。

**2.算法分析**

由于知识图存在的难以完全构建，现实生产中使用的知识图常常是不完全的，该算法对这种问题的解决方法是基于对用户-商品图的模拟，填补了知识图中事实的缺少。运用了联合嵌入的方法，利用知识图中的事实作为辅助信息，并且提出了两个模型，分别是基于翻译的用户偏好模型（Translation-based User Preference，简称TUP）和知识增强的基于翻译的用户偏好模型（Knowledge-enhanced TUP，简称KTUP）。

文中提出了这两个模型，对于这两个模型进行分别训练，并且对比训练后的成果。对于TUP模型，论文中采用了联合知识图来进行学习，并利用用户-商品（user-item）之间的多重隐性关系，揭示用户的偏好。为了使偏好具有明确的语义，将它们与知识图中的关系对齐，从而赋予它们一定的语义，增强了可解释性，使得项目属性的类型在用户决策过程中扮演着至关重要的角色。对于KTUP模型，论文将学习到的关系嵌入和实体嵌入进行了一系列的操作，具体是从知识图中迁移到了TUP中，同时训练知识图的完整性和推荐任务。联合学习用户、商品、实体和关系的表示，并将将TUP和TransH算法进行联合学习。在最终的实验中，可以发现这种算法具有很大的优势，既能提高知识图谱在推荐系统中使用的正确率，又可以解决知识图不完全这一存在的问题。

### 基于知识增强记忆网络的推荐顺序算法

基于知识增强记忆网络的推荐顺序算法（KSR），是由Wayne Zhao等人于2018年在SIGIR会议上发表了一篇名为“Improving sequential recommendation with knowledge-enhanced memory networks”的论文中提出的。[27]

**1.算法介绍**

当今在用户上购物网站时，首页上常常会出现针对用户的喜好所做出的推荐，这使得用户更有可能地在购物网站上消费，从而增加网站的收益。众多互联网公司也花了大功夫在构建推荐系统上。为了给出精准的推荐，首先需要把握用户在想什么，对什么感兴趣，而这往往不是一成不变的，会随着用户在平台上的活动而日益变化。鉴于此，相较于以往的协同过滤等推荐方式，学术界提出基于时序神经模型的序列化推荐系统，利用循环神经网络来捕捉用户兴趣点随着时间的动态变化。在序列化推荐建模中，通常将用户过去的交互记录作为输入，利用隐状态向量来编码各个交互记录，以此来表示用户在序列中体现的偏好。但是，这种方法有两个不足之处：第一点是只考虑了用户序列偏好，却忽视了细致的用户偏好，如用户具体喜欢某个商品的哪个属性等；第二点是推荐过程中的隐含特征向量表示过于抽象，难以解释其推荐结果。KSR算法利用了知识图谱与记忆网络，在提高推荐结果准确性的同时，还能捕捉更为细致的用户偏好，提高推荐系统的可解释性。

**2.算法分析**

因为从用户与商品的交互记录中只能体现出商品的隐语义模型，还不够更细粒度地刻画用户对商品的偏好，于是，该算法借助知识库中的实体信息来表示推荐商品在不同属性的特征，这样既可以更为细致地刻画商品，又可以增强可解释性。如一首歌曲拥有歌手、专辑等属性，通过知识库我们可以得知这些属性对应的具体实体，来利用嵌入的方式可以将这个实体在知识库中的上下文信息转化成一个嵌入向量，也就是特征表示向量。那么，在解决了细粒度表示与解释性问题之后，就要开始考虑如何将其与序列化推荐模型结合在一起。算法中利用门控循环神经网络（Gated Recurrent Unit，简称GRU）来构建序列化推荐模型，虽然GRU可以记忆相邻几个节点的信息，但对于长期记忆存储知识库实体信息来说，GRU的记忆还是过于短暂。因此，算法又引入了记忆网络（Memory Networks）来存储不同属性的信息，与知识库交互。

首先，GRU网络根据短程记忆得出用户序列兴趣表示向量；接着，键值对记忆网络根据GRU的表示向量与商品本身的特征表示向量得到细致的用户偏好特征表示向量，并于序列兴趣向量连接；最后，根据用户对各个兴趣的权重，即可解释推荐系统产生推荐的原因。

虽然针对序列化推荐系统不具有解释性，并且无法获取用户细粒度特征的特点，但是该算法提出了利用结合知识库的记忆网络来增强推荐系统的特征捕获能力与解释性，不仅获得了更高的准确率，还使得推荐系统具有了较强的解释性。并且经过实验结果表明，该模型比以往的模型在准确率与解释性方面均有显著突破。

### 算法优缺点对比

对于知识图谱在推荐系统上的运用，我们调研了多篇论文，从中选出了四篇，对此四篇论文所提出的算法做了一系列的横向对比，分析了它们的优缺点。

KPRN模型将知识图谱与循环网络相结合，利用由主体与关系的语义组合来获取路径表示，相较于其他利用知识图谱的方法，此方法能够在循环遍历过程中更全面地发现潜在的用户-商品关系，同时利用路径的顺序以来使得推荐有更高的可解释性。对于本项目而言，此方法有重要意义，因为机票购买背后的原因非常复杂，需要考虑天气、节假日、机票论坛等错综复杂的因素，知识图谱能够有效地利用好这些因素并从中提取潜在的协同关系，但文中方法利用的数据集都是单一维度的，如音乐、电影等，本项目如果采用知识图谱可能面临数据稀疏的问题，即图中初始边的数目过少，这是一个需要解决的问题。

基于知识图谱增强推荐的多任务特征学习算法解决了协同过滤问题中常常存在的稀疏性和冷启动问题，该算法将知识图谱作为辅助信息，知识图谱嵌入任务和推荐任务这两个任务由交叉和压缩单元相关联，在运用场景中具有高度的灵活性和适应性。

基于知识图谱的用户喜好推荐模型也是将知识图谱引入新的模型中，作为辅助信息去解决一些问题，便于训练模型，细化粒度。这些算法也体现了知识图谱在当今的推荐系统算法中的一系列作用。

基于知识增强记忆网络的推荐顺序算法提出了将知识图谱和记忆网络进行交互，在提高推荐结果准确性的同时，还能捕捉更为细致的用户偏好，提高了推荐系统的可解释性。

这几种基于知识图谱的推荐系统大多具有一定的相似性，即知识图谱在这些场景下起到的作用大多是丰富数据层次和内容，更好地完成建模，并使得模型具有层次。但是引入知识图谱也会引入大量数据，增加存储消耗，产生一些因为数据量过大而导致数据繁杂，冗乱的问题。

## 4.4 其他算法

### 4.4.1 基于深度学习的内容协同算法

基于深度学习的内容协同算法（SFNET），是由Sheikh等人于2019年在Recsys会议上发表的“A deep learning system for predicting size and fit in fashion e-commerce”一文中提出的。[28]

**1.算法介绍**

个性化的尺寸及适合度推荐对于任何时尚电商平台都有着至关重要的意义。预测正确的适合度可以提高用户满意度，减少尺寸问题带来的成本，从而使企业受益。传统的协同过滤算法试图根据用户之前的订单来建立用户偏好模型。这种方法的一个经典挑战来自于用户-商品订单的极度稀疏。在这项工作中，我们提出了一种基于深度学习的内容协作方法，用于个性化的尺寸及适合度推荐。标准的协同过滤方法完全依赖于交互数据来对用户行为进行建模，但是对于绝大多数用户来说，这种交互数据是稀疏的，因而导致了提取出来的用户-商品交互矩阵极其稀疏，这使得在在个性化的层次上对每个用户的偏好进行建模变得很困难。但以用户和商品属性的形式提供的附加信息有助于解决冷启动和数据稀疏问题。因此文中提出的方法既使用交互数据，也使用任意用户和商品特征来进行个性化的尺寸/适合度预测。该方法采用具有全局参数和实体参数的分裂输入神经网络结构。全局参数集允许模型捕获与跨用户预测尺寸和适合度相关的信息，而实体级嵌入变量使模型能够发现单个用户和商品的隐属性。因此该方法可以单个账户背后的多用户或多意图进行建模。

**2.算法分析**

SFNET模型的架构通过不同的前馈输入路径接收用户和商品信息。用户和商品路径的输入层都嵌入了类别特征(如用户ID、商品ID、品牌等)，以便将特征值映射到可训练的特征向量。通过在映射过程中嵌入用户/商品的独特标识，SFNET模型具备了用端到端的方式从单一用户或商品中学习个性化潜在特征的能力。用户路径和商品路径分别将对应的连续或非连续的特征串联起来，将串联后的结果输入到使用残差连接的前馈全连接层级联中去，来获得用户和商品的潜在特征向量。其中初始的连续特征和从类别特征中学习到的特征的结合使得模型能够在个性化的层面表示用户和商品。通过这样的嵌入模式，SFNET可以做到理清信息，识别单个账户背后的多个具有不同尺寸、偏好的角色，并发现某些品牌或商品固有的属性。在获得了用户和商品潜在特征向量后，将它们作为残差连接的非线性层的输入，获得对大小和适合度的预测。

此算法是在电商领域尤其是服饰领域的新突破，以往的分层贝叶斯模型只考虑尺寸的因素，SFNET将尺寸与适合度结合，是该领域的一个新见解。此算法的表现在多个数据集上优于基于尺寸的分层贝叶斯算法，但是这种领先随着Top-K的增加会被缩小。

### 4.4.2 用户超长行为历史下的CTR预测

用户超长行为历史下的CTR预测，是由阿里在2019年的KDD会议上发表了一篇名为“Practice on Long Sequential User Behavior Modeling for Click-Through Rate Prediction”的论文中提出的。[29]

**1. 算法介绍**

对于一些大型企业来说，大多具有丰富的用户历史行为信息，从用户的行为信息中可以发掘出用户的行为偏好。对于一些网站来说，广告收入也是诸多收入之一，而点击通过率（Click-Through-Rate，简称CTR）是衡量互联网广告效果的一项重要指标，指网络广告（图片广告/文字广告/关键词广告/排名广告/视频广告等）的点击到达率，即该广告的实际点击次数（严格的来说，可以是到达目标页面的数量）除以广告的展现量。用户超长行为历史下的CTR预测方法解决了对于超长的用户行为历史下，如何对CTR进行预测的问题。

**2. 算法分析**

因为数据量过大，在存储和使用延迟的时候存在限制，论文设计了一个名为用户兴趣中心模块(User Interest Center，简称UIC)的模块，UIC保持最新兴趣状态对于每个用户，其更新取决于实时用户的行为事件，而不是根据流量请求。因此，UIC没有延迟实时点击率预测。从机器学习算法观点，论文中也提出了一种新颖的基于内存的架构，名为多渠道用户兴趣存储网络（简称MIMN），从用户长期的顺序行为数据中获得兴趣，从而超越最新模型。

UIC和MIMN的协同设计解决方案可以实现用户以无限的长度来处理用户兴趣建模顺序行为数据。该算法专注于利用长序列用户行为数据与机器学习算法和在线协同设计服务系统。在计算方面，存储是主要的。从相当长的时间内捕获长期用户兴趣的瓶颈顺序用户行为数据。该算法提出的解耦的UIC服务器进行实时推断-用户兴趣建模和基于内存的MIMN模型的作用可以逐步实现，并且胜过其他最先进的模型。

### 4.4.3 考虑时空域影响的CTR预测

考虑时空域影响的CTR预测（DSTN），是由阿里于2019年在KDD会议上发表的 “Deep spatio-temporal neural networks for click-through rate prediction”的一文中提出的。[30]

**1. 算法介绍**

时空域在一定程度上可以分解为空间域和时间域。空间域的意思即在一屏的推荐中，内容是相互关联的，当推荐了第一条广告之后，第一条广告会对第二条广告的点击率产生影响，从而影响第二条推荐的广告。时间域的意思即用户之前的点击或未点击的广告会影响当次的推荐。该方法同时考虑空间域信息和时间域信息，来进行广告的点击率预估。该论文在计算广告点击率，不但考虑了用户的偏好和广告的内容，还考虑了其他广告对该广告的点击率造成的影响。

**2. 算法分析**

在该论文介绍的算法之前的模型，大多是一次计算所有广告的点击率，然后按点击率进行排序，取Top-K进行展示。但这里算法把一次推荐个广告过程看作个单次推荐的过程集合。先推荐第一个位置的广告，再推荐第二个位置的广告，依次类推。这样的考虑引入了上下文信息的影响，为了将这些信息加入到模型中，模型必须要注意以下几点：

（1）每种类型的辅助广告数量可能相差很多，模型必须适应这些所有可能的情况。

（2）辅助的广告信息可能与目标广告是不相关的，因此，模型需要具备提取有效信息，而过滤无用信息的能力。举例来说，用户点击过的广告可能有咖啡广告、服装广告和汽车广告，当目标广告是咖啡相关的广告时，过往点击中咖啡相关的广告可能是起比较大作用的信息。

（3）不同类型的辅助广告信息，有时候起到的作用可能是不同的，模型需要能够有能力对此进行判别。

总的来说，就是模型需要有能力有效处理和融合各方面的信息。因此论文中提出了考虑时空域的CTR预测模型来处理和融合各种辅助广告信息。

### 4.4.4 基于局部潜在模型的Top-N推荐算法

基于局部潜在模型的Top-N推荐算法，是由Karypis等人在2018年的KDD会议上发表了一篇名为“Local latent space models for top-n recommendation”的论文中提出的。[31]

**1.算法介绍**

用户的行为是由他们对商品的各个方面的偏好驱动的，这些偏好决定了用户会购买或只是查看浏览商品。潜在空间方法以潜在因素的形式对这些方面进行建模。尽管这些方法已被证明能够产生良好的效果，但对不同用户来说重要的方面可能会有所不同。在许多领域中，可能存在一组所有用户都关心的方面，以及一组特定于不同用户子集的方面。为了明确地捕获这一点，文中提出的模型中的一些潜在因素专门用来捕获共享的方面，而另一些用户子集的特定潜在因素用来捕获不同的用户子集关心的特定的方面。

在论文中，提出了两个潜在空间模型：RGLSVD和SGLSVD，它们结合了这样一个全局和用户子集特定的潜在因素集。RGLSVD模型根据用户的评级模式将用户分配到不同的子集，然后估计一个全局和一组特定于用户子集的本地模型，这些模型的潜在维度数量可能会有所不同。SGLSVD模型通过保持这些模型中潜在维度的数量相同来估计全局和用户子集特定的局部模型，但优化用户分组以实现最佳近似。

**2.算法分析**

Top-N推荐系统通过利用来自该用户和其他用户的历史信息，在这些项目的大量集合中识别出少量的个项目，用户会发现这些项目很容易引起他们发生购买、查看、喜欢、单击等行为。它们广受欢迎，从Netflix电影推荐，到亚马逊产品推荐，再到Facebook好友推荐等等。

该算法建议通过估计全局低阶模型和多用户子集特定的低阶模型来显式编码这种潜在模型。算法中提出了两种方法：即通过变秩的全局和局部奇异值分解（RGLSVD）和变子集的全局和局部奇异值分解（SGLSVD）。RGLSVD方法考虑混合的用户子集，但允许不同的本地模型具有不同的列组。SGLSVD方法解决了一个联合优化问题，即估计固定列的局部模型，同时自动阻止挖掘用户的各种子集。这两种方法探索了不同的方法来学习本地的低级表示，从而为用户获得最佳的Top-N推荐质量。

该算法较之前的算法GLSLIM相比，虽然GLSLIM方法也估计了一个全局和多个局部模型，但是结果表明，尽管GLSLIM获得了更好的Top-N推荐结果，但SGLSVD比基线全局方法快一个数量级，并且SGLSVD方法其相对于基线全局方法的改进百分比高于GLSLIM对slim的改进百分比。这个数据可以看出，该算法对于Top-N推荐优于其他类型的算法，在推荐结果上做了较大的提升和改善。

### 4.4.5 基于神经网络的上下文感知算法

神经上下文感知方法（CoNCARS），是Felipe Soares da costa等人于2019年在Recsys会议上发表的“Collective Embedding for Neural Context-Aware Recommender Systems”一文中提出的模型。[32]

**1.算法介绍**

上下文感知推荐系统将上下文特征视为预测用户偏好的额外信息。时间是上下文信息中一项重要特征，因为用户偏好往往会随时间而变化或在不远的将来保持下去。为了解决现有模型无法捕捉特定时间模式的问题，文中提出了神经上下文感知推荐系统。

CoNCARS模型对用户、商品和时间向量共同建模以捕捉时态范式，然后使用外积在向量空间维度内对用户-商品-时间相关性进行建模，并利用这些隐藏特征作为卷积神经网络（Convolutional Neural Networks，简称CNN）的输入来学习不同特征间的非线性关系，最终在融合层中将CNN的输出结合起来预测用户的偏好分数。

**2.算法分析**

CoNCARS系统一共分为六层，包括输入层、嵌入层、交互层、卷积层、融合层和预测层。在输入层中，用户和商品分别被预处理为独热表示向量和，在时间时用户和商品的可见交互被分别标识为用户-时间二进制交互向量和商品-时间二进制交互向量。

共同嵌入层的目标是将有关用户-商品交互的有意义的信息嵌入进输入层产生的四个向量（用户-商品、商品-用户、用户-时间和商品-时间）中，将和分别与目标用户和目标商品求外积作为特征向量，将和分别利用注意力机制按评分重要程度进行加权求和作为基于时间的用户/商品特征向量。

在交互层中利用和、和这两组特征向量对用户在时间对商品的交互进行重建。在卷积层中，CNN负责从交互层的输出中学习非线性交互。第一层的特征图被表示为一个三维张量，使得CoNCARS能够在公共潜在空间中捕捉特定时间模式，最终四个卷积层分别输出一个一维的特征向量。

在融合层中将卷积层输出的四个潜在特征向量串联成一个最终的潜在特征向量。预测层通过对权重和潜在特征向量求外积来进行评分，输出结果为在时间时用户对商品的评分。

# 参考文献

[1] Hill W, Stead L, Rosenstein M, et al. Recommending and evaluating choices in a virtual community of use[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. 1995: 194-201.

[2] Shardanand U, Maes P. Social information filtering: algorithms for automating “word of mouth”[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. 1995: 210-217.

[3] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM conference on Computer supported cooperative work. 1994: 175-186.

[4] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval[M]. New York: ACM press, 1999.

[5] Salton G. Automatic text processing: The transformation, analysis, and retrieval of[J]. Reading: Addison-Wesley, 1989, 169.

[6] Belkin N J, Croft W B. Information filtering and information retrieval: Two sides of the same coin?[J]. Communications of the ACM, 1992, 35(12): 29-38.

[7] Mooney R J, Bennett P N, Roy L. Book recommending using text categorization with extracted information[C]//Proc. Recommender Systems Papers from 1998 Workshop, Technical Report WS-98-08. 1998, 1188.

[8] Pazzani M, Billsus D. Learning and revising user profiles: The identification of interesting web sites[J]. Machine learning, 1997, 27(3): 313-331.

[9] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[J]. arXiv preprint arXiv:1301.7363, 2013.

[10] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet computing, 2003, 7(1): 76-80.

[11] Basu C, Hirsh H, Cohen W. Recommendation as classification: Using social and content-based information in recommendation[C]//Aaai/iaai. 1998: 714-720.

[12] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. IEEE transactions on knowledge and data engineering, 2005, 17(6): 734-749.

[13] Claypool M, Gokhale A, Miranda T, et al. Combing content-based and collaborative filters in an online newspaper[J]. 1999.

[14] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.

[15] Wu S, Tang Y, Zhu Y, et al. Session-based recommendation with graph neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 346-353.

[16] Monti F, Bronstein M, Bresson X. Geometric matrix completion with recurrent multi-graph neural networks[C]//Advances in Neural Information Processing Systems. 2017: 3697-3707.

[17] Berg R, Kipf T N, Welling M. Graph convolutional matrix completion[J]. arXiv preprint arXiv:1706.02263, 2017.

[18] Zhang J, Shi X, Zhao S, et al. Star-gcn: Stacked and reconstructed graph convolutional networks for recommender systems[J]. arXiv preprint arXiv:1905.13129, 2019.

[19] Ma C, Ma L, Zhang Y, et al. Memory Augmented Graph Neural Networks for Sequential Recommendation[J]. arXiv preprint arXiv:1912.11730, 2019.

[20] Ying R, He R, Chen K, et al. Graph convolutional neural networks for web-scale recommender systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 974-983.

[21] Wang X, He X, Wang M, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. 2019: 165-174.

[22] Liang D, Krishnan R G, Hoffman M D, et al. Variational autoencoders for collaborative filtering[C]//Proceedings of the 2018 World Wide Web Conference. 2018: 689-698.

[23] Ebesu T, Shen B, Fang Y. Collaborative memory network for recommendation systems[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018: 515-524.

[24] Wang X, Wang D, Xu C, et al. Explainable reasoning over knowledge graphs for recommendation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 5329-5336.

[25] Wang H, Zhang F, Zhao M, et al. Multi-task feature learning for knowledge graph enhanced recommendation[C]//The World Wide Web Conference. 2019: 2000-2010.

[26] Cao Y, Wang X, He X, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences[C]//The world wide web conference. 2019: 151-161.

[27] Huang J, Zhao W X, Dou H, et al. Improving sequential recommendation with knowledge-enhanced memory networks[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018: 505-514.

[28] Sheikh A S, Guigourès R, Koriagin E, et al. A deep learning system for predicting size and fit in fashion e-commerce[C]//Proceedings of the 13th ACM Conference on Recommender Systems. 2019: 110-118.

[29] Pi Q, Bian W, Zhou G, et al. Practice on long sequential user behavior modeling for click-through rate prediction[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 2671-2679.

[30] Ouyang W, Zhang X, Li L, et al. Deep spatio-temporal neural networks for click-through rate prediction[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 2078-2086.

[31]Christakopoulou E, Karypis G. Local latent space models for top-n recommendation[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018: 1235-1243.

[32] Costa F S, Dolog P. Collective embedding for neural context-aware recommender systems[C]//Proceedings of the 13th ACM Conference on Recommender Systems. 2019: 201-209.