推荐系统面临的挑战及解决方案

马尔瓦·侯塞因·穆罕默德 信息系统部 法尤姆大学 埃及开罗

[Eng\_maroo1@yahoo.com](mailto:Eng_maroo1@yahoo.com)

Mohamed Helmy Khafagy 计算机科学系 法尤姆大学 埃及开罗

Mhk00@fayoum.edu.eg

穆罕默德·哈桑·易卜拉欣 信息系统部 法尤姆大学 埃及开罗

mhi11@fayoum.edu.eg

**摘要**

今天的推荐系统是机器学习中一个相对较新的研究领域。推荐系统的主要思想是建立产品、用户之间的关系，并为特定用户选择最合适的产品。推荐系统产生推荐列表的方法有基于用户内容的、协作的、人口统计的和混合过滤等四种主要方式。在基于内容的过滤中，该模型使用项目的规格作为参考，以便推荐具有相似属性的附加项目。协同过滤使用用户过去的行为，如用户以前查看或购买的项目。总之，用户对这些项目的评分和其他用户的项目列表得出的类似结论，来预测用户可能感兴趣的项目。人口统计过滤是查看用户个人资料数据，如年龄类别、性别、教育和居住区域，以找到与其他个人资料的相似之处，从而获得一个新的推荐者列表。混合过滤结合了所有三种过滤技术。本文介绍了推荐系统的概况、技术以及面临的挑战，并列举了一些解决这些挑战的研究论文。

关键词：推荐系统，大数据，基于内容的过滤，协同过滤，混合过滤，机器学习。

**第一章 引言**

推荐系统在帮助电子商务在万维网的许多应用中发展方面发挥着重要作用。此外，推荐系统是当今最重要的研究领域之一，因为它帮助用户找到他们对互联网的兴趣点所在[1]。

许多产品都是在网上购买的，而且越来越多的顾客要求对网站上的大量商品进行过滤，以便根据他们的实际兴趣更容易地找到特定的商品[2]。

推荐系统，用于通过使用来自网站上的用户评级列表历史的信息过滤系统，或者通过查找项目的相似性规范，或者通过使用来自他们的简档[1]的人口统计信息来查找共同兴趣，如果这是第一次在网上看到，会预测用户将给予项目的评级，以下是使用推荐系统的网站列表:

1. 电影推荐（网动画）
2. 图书推荐（亚马逊）
3. 音乐推荐（彭杜拉）

4、新闻推荐（雅虎）

数据挖掘是遥感中重要的分析技术之一。遥感中最常用的数据挖掘方法:分类、聚类和关联规则发现[3]。

推荐系统有不同类型的过滤，通过协作的、基于内容的和混合的过滤算法来制作有效的推荐引擎。如果你从网站上购买一个物品，你将被推荐更多的物品，基于你购买的内容项目规格[4]。例如，系统会向您推荐您购买的同一本书的主题或同一作者出版的多本书。

使用推荐系统对企业的好处如下所示:

收入——在过去的几年里，许多研究人员已经研究并生成了许多算法来学习像亚马逊网站这样的在线客户的增长率。此外，这些算法还研究了在线购物网站与其他使用商品推荐系统的网站之间的差异，通过增加销售额来增加收入。

客户满意度——很多时候，客户倾向于从他们在网站上的最后一次浏览搜索中看到相似的产品推荐。主要是因为他们相信他们会有更多机会获得更好的产品。当他们离开现场，然后回来；如果他们从以前的购物或查看产品列表中浏览数据，将会有所帮助。这可以进一步方便和指导他们的电子商务活动，类似两个有经验的助手。这种顾客满意的情况有助于留住顾客。

个性化——我们经常从朋友那里得到推荐。他们比任何人都更清楚我们喜欢什么。这是他们擅长推荐事物的唯一原因，也是推荐系统试图建模的原因。您可以利用间接收集的数据来改善您的网站的整体服务，并确保它们适合用户的偏好。

发现——当人们找到购物或电影、歌曲等网页时，他们需要被推荐他们喜欢或喜欢的项目。满足他们的希望，他们一定会再次访问这个网站。

提供报告—是个性化方案的一个组成部分。让客户准确及时，报告让他对自己的网站和运动管理做出强有力的结论。基于这些报告，客户可以获得缓慢移动产品的报价，以推动销售。

如今，全世界都在努力构建使用大数据来处理快速增长的数据的技术和软件[5]。大数据需要高性能地匹配和连接数据库中表之间的输出结果。一些技术像地图缩小[6]、蜂巢[7]，spark，等等。以及如何在不同的查询之间连接结果。所有这些技术都将影响推荐系统的未来。

在本文中，我们重点讨论了目前在该领域广泛使用的推荐系统技术:基于内容的推荐、协作推荐、人口统计推荐和混合方法以及推荐系统中使用的数据挖掘方法。列出推荐系统面临的问题和挑战，以及如何克服这些挑战，最后我们提出了我的研究论文，解决了过去推荐系统面临的许多挑战。

本文的其余部分组织如下:第二节介绍推荐系统技术，第三节介绍推荐系统的研究挑战，第四节详细介绍了推荐系统技术的文献综述。最后，在第五节当中给出了结论。

**第二章 推荐系统技术**

推荐系统使用不同的方法和来源获得关于用户的信息，以预测哪个用户需要并根据该分析过程结果推荐项目[3]。图1表示推荐系统的一般概念，同时将用户简档与其他用户简档进行匹配，以获得它们之间的公共项目，并推荐它们之间的差异[8]。推荐引擎软件[4]基于三种技术对我们可能喜欢购买、收听或观看的内容进行预测：基于内容的建议、合作建议、人口统计建议以及它们的混合方法。

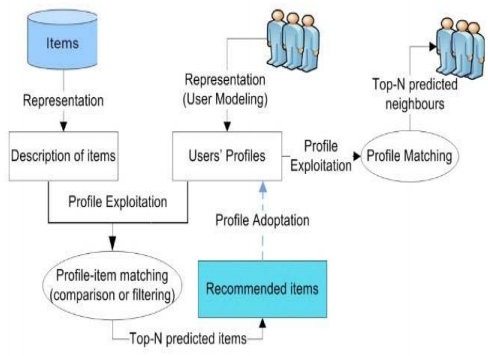


图1 推荐系统的一般概念[8]。

2.1基于内容的建议

基于内容的过滤方法基于对项目的描述和用户喜好的描述[9]。它还推荐类似于给定用户喜欢的项目费率列表中的项目。事实上，由基于内容的系统执行的基本操作包括将用户基本数据(如年龄、性别、位置和存储在他的账户中的网站上的分级项目列表)与具有共同规范的类似项目进行匹配，以便推荐满足他/她的兴趣的新项目。基于内容的推荐流程步骤如下所示:

1、内容分析器

当信息没有像文本那样的结构时，需要某种预处理步骤来提取结构化的相关信息。组件的主要职责是呈现项目的内容，如文档、网页、新闻、产品描述等。以适合以下处理步骤的形式来自信息源。通过特征提取技术对数据项进行分析，以便将原始信息空间的项表示转换到目标信息空间。该代理是简档学习和过滤组件的输入。

2、简介学习者

该模型通过收集用户偏好数据来构建用户简档，并试图概括这些数据。通常，泛化策略是通过机器学习技术实现的，机器学习技术能够从过去喜欢或不喜欢的项目开始理解用户兴趣的好例子。举例来说，网页推荐器的简档学习器可以实现相关反馈方法，其中学习技术将正面和负面示例的向量组合成表示用户简档的原型向量。培训示例是用户提供正面或负面反馈的网页。

3、过滤组件

该模型使用用户数据配置文件来查找与项目列表相关的匹配项目，但是是新产品，并将这个新项目呈现给用户。潜在有趣项目的推荐列表。通过计算原型向量和项目向量之间的余弦相似度来实现匹配。图2讨论了基于内容的过滤步骤[9]。

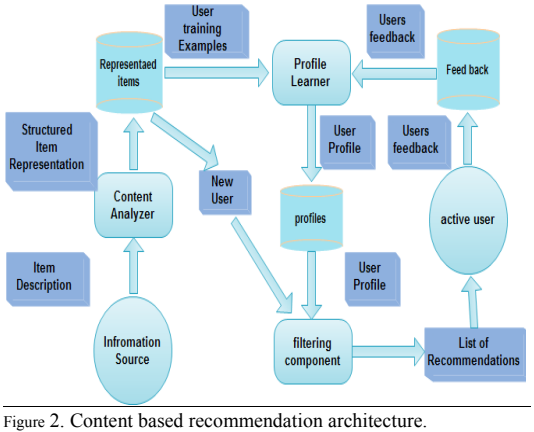


图2 基于内容的推荐体系结构

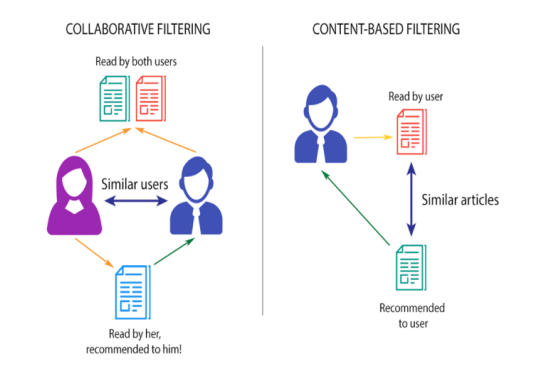


图3 基于内容的过滤和协作过滤推荐

2.2协作建议

基于协同过滤[9]的系统以用户反馈、评级、偏好和活动的形式收集和分析用户的行为信息。基于这些信息，然后利用几个用户或项目之间的相似性来预测缺失的评级，从而提出合适的建议。协作过滤(CF)方法基于评级或使用模式(如购买)生成用户特定的项目推荐，而不需要关于项目或用户的数据。

有以下三个陈述：如果人P1喜欢第1项，第2项，第3项；如果人P2喜欢第1项，第3项，第4项；如果人P3喜欢第1项。由上可知，P3很有可能喜欢第3项，因为从前两个陈述中，我们知道P1、P2、P3喜欢第1项。

图3显示了基于内容的协同过滤推荐[9]。合并这些不同的方法可以实现高性能，并通过仅使用基于内容的过滤或协作过滤来缓解出现的问题和挑战。

2.3人口统计建议

这种推荐系统的技术使用用户简档信息，如年龄、性别、人口统计区域、教育、兴趣以及他们对评级项目的看法，并且发现普通用户具有相似的评级项目和兴趣，将用户按年龄组和居住区域划分[10]。

2.4用于推荐系统和大数据的数据挖掘方法

随着数据规模的快速增长，许多用于从大数据等大型数据集检索有价值的业务信息的数据挖掘方法在今天非常重要。在推荐系统[3]中，当用户数量和产品数量增加时，我们获得了高性能。以下是讨论数据挖掘和推荐系统中使用的一些方法:

1. 异常检测:我们选择的值不同于任何被称为异常的数据样本。
2. 聚类分析:我们用数据样本在某些方面与同一组中的其他样本相似。该算法需要在大数据结构中找到规格相似的项目或购买相似项目的用户。然后匹配这些结果并向用户建议新的推荐项目列表。

3、分类:用聚类方法将数据分组后。我们使用分类从不同的聚类中进行匹配。这可以通过决策树来实现。我们在项目或用户树中定义主节点和以下对象。

4、关联分析:该方法旨在通过定义数据集之间的关系来制定推理规则。我们需要找出哪些产品已经一起购买，以确定相关性。

5、回归分析:是一种分析方法，用于创建模型，通过多个独立变量找到因变量之间的关系。

2.5混合方法

在混合方法中，我们合并了基于内容和协作过滤的两个推荐技术，以获得最佳优势和获得更好的结果，并减少这些应用的问题和挑战[8]。混合方法有多种方法[11]:

1. 加权:我们将系统给定不同分数的每个推荐组件进行数字组合。
2. 切换:系统向用户提供不同推荐项目的多种选择，并根据用户偏好选择首选项目。
3. 混合:系统同时向用户推荐多个不同的项目。
4. 特征组合:将多个知识源合并，构建推荐系统特征。
5. 特征增强:下一个技术的重要部分之一是用于计算推荐系统的一组特征的特征增强。
6. 级联:推荐者列表中有一个加权优先级的项目，高的优先，低的递减。
7. 元级:是一种输入技术，用于为推荐系统算法的下一步生成某种模型。通过仅使用基于内容的过滤或协作过滤，合并这些不同的方法可以实现高性能，并缓解出现的问题和挑战。

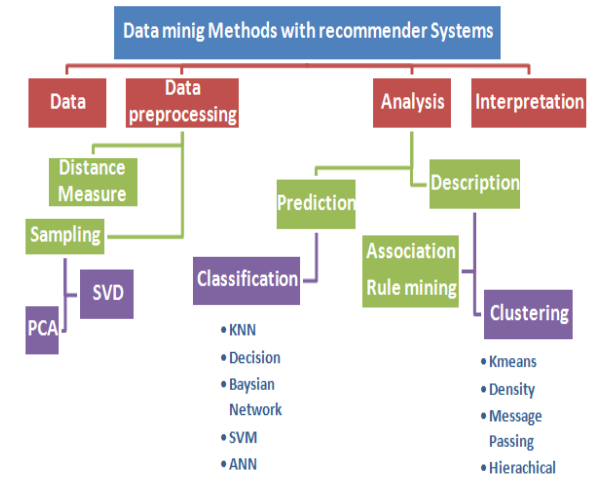


图4 基于推荐系统的数据挖掘方法

2.6 推荐系统技术的比较

在这一部分，我们将讨论推荐系统的优点和缺点，如表1所示。

表1 推荐系统的优缺点

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 技术 | 优点 | 缺点 |
| 1 | 基于内容的推荐过滤 | 1.系统没有使用用户的数据来推荐商品。  2.该系统能够基于项目规格之间的相似性向用户推荐新项目 | 1.我们需要分析和检测所有项目特征来创建推荐列表  2.该系统不依赖于用户对该项目的评价，因此不包括对产品质量的评价。 |
| 2 | 协同推荐过滤 | 1.系统没有使用人口统计信息来推荐商品  2.系统在用户之间匹配相似的项目。  3.系统能够向用户推荐他们偏好之外的项目，并且可能喜欢这个项目。 | 1.系统的质量取决于最高评级项目列表。  2.如何向新用户推荐商品存在问题(冷启动问题)。 |
| 3 | 人口统计推荐过滤 | 它不是基于用户项目评分，而是在用户对任何项目评分之前给出建议。 | 1.收集人口数据会导致隐私问题。  2.稳定性与塑性问题。 |
| 4 | 混合方法 | 1.它结合了基于内容和协作过滤的所有优势。  2.它基于内容描述和用户评估。  3.解决过度专业化  4.提高客户满意度。 | 1.遭受冷启动问题。  2.产品的早期评级问题。  3、稀疏问题。 |

**第三章 推荐系统研究挑战**

推荐系统面临许多需要研究和解决的挑战，在这一部分我们将介绍其中的一些。

冷启动问题:冷启动问题主要出现在我们有一个新用户访问网站或向系统添加新项目时。首先，我们如何向新用户推荐商品，我们不知道他的兴趣，他也没有给任何商品打分。第二，我们可以推荐这个新项目给其他人，甚至没有人评价这个项目对用户来说是好是坏[4]。

稀疏性问题:稀疏性问题是推荐系统中的一个重要问题，当用户拥有包含购买项目或观看电影或音乐列表的大矩阵时，就会出现稀疏性问题。当用户不对这些项目进行评级时，稀疏性就产生了。而推荐系统依赖于用户评级矩阵用户向其他人推荐[4]。

可伸缩性:可伸缩性衡量系统在信息增长的同时高效工作的能力。当用户数量增加或项目数量增加时，推荐系统需要向用户推荐项目而不改变。为了实现这一点，我们需要更多的计算和昂贵的[1]。

过度特殊化问题:向用户推荐的项目是基于那些已经知道的或由用户配置文件定义的项目，而没有发现新的项目和其他可用的选项[4]。

多样性:确保推荐结果尽可能跨越你的物品空间，而不是来自同一个聚类[12]。

新奇:推荐的项目必须包含新的意外发现。除了新奇之外，一些被推荐的项目不仅闻所未闻，而且令人惊讶的是用户在[12]之前不会想到，这也是一个客观事实。

隐私:隐私是推荐系统面临的重要挑战之一。推荐系统要推荐符合自己兴趣的项目，就必须知道一些关于用户数据的信息。用户需要知道哪些信息需要推荐给他更好的项目，以及如何应用[13]。

先令攻击:如果一个恶意用户或竞争对手进入一个系统，并开始对某些项目给出错误的评级，以增加该项目的受欢迎程度或降低其受欢迎程度，就会发生这种情况[8]。

灰色绵羊:灰色绵羊出现在协作过滤系统中，在该系统中，用户的意见不等同于任何群体，因此，不能获得推荐的好处[8]。

**第四章 文献评论**

在这一部分，我们将讨论研究推荐系统技术的研究论文。表2讨论了其他论文的作者姓名和他们用来克服推荐系统挑战的解决方案，以及每篇论文的优势。

表2 论文解决方案及优势比较

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 作者姓名 | 结论 | 优点 |
| 1 | Ghazanfar and Prugel-Bennett [14] | 他们结合了所有的推荐系统技术 | 它用于消除推荐系统中的冗余记录问题。 |
| 2 | Qian Wang, Xianhu Yuan, et al. [15] | 他们将项目人口统计信息与搜索一组具有相同兴趣的相邻用户相结合，并使用遗传算法 | 它提高了系统的可扩展性 |
| 3 | Kumar, Vikas, et al. [16] | 他们正在构造二元矩阵分解的两类结构 | 生成比序数矩阵更精确的等级矩阵 |
| 4 | Xuelin Zeng et al. [17] | 他们使用spark来实现ALS矩阵分解，通过使用MLib来生成SGD矩阵分解来进行比较 | 提升了效率和速度 |
| 5 | Dianping. Lakshmi et al. [18] | 他们发现不同项目之间的关系，并与用户项目评分矩阵进行比较，以发现相似之处。 | 为该用户生成新的推荐项目列表，并解决新颖性和多样性问题 |
| 6 | Sajal Halder et al. [19] | 他们提出了电影群挖掘的概念，该算法使用了两个剪枝规则和频繁项挖掘。 | 解决了冷启动问题 |
| 7 | Arpan V Dev et al. [20] | 他们使用映射简化技术来删除重复的计算开销。 | 与其他常规算法相比，获得更高的性能 |
| 8 | Rahul Katarya, Om Prakash Verma [23] | 他们使用模糊，一种为用户寻找邻居的c均值技术 | 解决电影推荐系统的可扩展性和准确性 |
| 9 | Panigrahi, et al. [24] | 他们提出了一种新的混合算法，使用了K均值和降维技术，如ALS | 解决了稀疏性和可伸缩性问题 |
| 10 | M.R. Lee et al. [25] | 他们提出了一个使用机器学习和脸书粉丝页面数据的混合过滤推荐系统 | 解决冷启动和精度问题。  提高客户满意度。 |

本论文的主要思想是综合运用各种推荐系统技术，提出一种新的推荐系统——级联混合推荐系统。该技术具有三种技术的所有优点:基于内容的过滤、协作过滤和人口统计过滤。它用于消除推荐系统中的冗余记录问题。

Qian Wang, Xianhu Yuan等[15]使用人口统计学过滤技术来提高系统的可扩展性和遗传算法来提高推荐系统的准确性。他们提出了一种混合用户模型，将项目人口统计信息与搜索一组具有相同兴趣的相邻用户相结合，并使用遗传算法来确定用户模型中的权重特征。

Kumar，Vikas，等[16]提出了一种新的方法，通过构造二元矩阵分解的两类结构来生成比序数矩阵更精确的等级矩阵完备性。他们使用K均值和降维技术，如ALS。

李明哲等人[25]提出了一种混合滤波推荐系统，利用机器学习和脸书粉丝页面数据来实现对用户推荐项目的高满意度和准确性，解决冷启动问题。在这个算法中，从雅虎、电影和脸书的粉丝页面中提取内容，使用基于内容的过滤。然后将该算法的输出结果与其他推荐系统进行比较，如网飞、YouTube和亚马逊。

解砾等[26]本文利用spark的大数据分布式计算平台的力量，解决了在推荐系统合并中使用ALS算法所遇到的问题。通过设计一个新的损失函数来确定用户和项目之间的相似性连接，解决了这个问题。算法输出通过比较不同迭代中的均方根与其他现有的协同过滤技术计算的结果。

林雪曾等[17]论文需要解决他们的模型的效率和准确性。他们提出了一种被称为平行潜群模型(PLGM)的技术。他们用火花实现ALS矩阵分解，用MLib生成SGD矩阵分解进行比较。

Dianping、Lakshmi等人，[18]在文章中发现不同项目与用户项目评分矩阵之间的关系，以查看相似性，并为该用户生成新的推荐项目列表。由基于项目的协作过滤技术命名。

萨加尔·哈尔德等人[19]提出了电影群挖掘概念。这用于解决冷启动问题，向新用户推荐项目，以及如何管理和增加最新和著名电影的观看时间。该算法使用了两个剪枝规则和频繁项挖掘。

Arpan V Dev等人[20]提出了一种基于项目和用户兴趣之间的相似性连接的算法，称为扩展前缀过滤。该算法使用映射约简技术来删除重复的计算开销。与其他常规算法相比，大数据应用程序使用这种技术获得了更高的性能[21，22]。

Rahul Katarya，Om PrakashVerma[23]提出了一个基于协同过滤技术的混合电影推荐系统。在本文中，为了解决电影推荐系统的可伸缩性和准确性，他们使用了模糊，一种为用户寻找邻域的c-means技术。

Panigrahi等人，[24]文章通过提出一种新的混合算法来消除面向用户的协同过滤技术的挑战，如可伸缩性和稀疏性。

**第五章 结论与展望**

推荐系统是当今一个重要的研究领域。疯狂地，数据量的增加，如网站上的项目和用户数量的增加，引发了大数据分析技术，如Spark、Map-Reduce、Apache Hadoop等[27]。推荐系统用于根据用户的兴趣和以前的项目评分表向用户推荐项目。在本文中，我们讨论了推荐系统中的四种技术，并列出了各自的优缺点[11]。讨论推荐系统面临的挑战:如冷启动、可扩展性、隐私、灰色羊、先令攻击、新颖性、稀疏性、多样性、过度专业化等问题。此外，还讨论了一些用于克服挑战的研究课题解决方案及其优势。在未来的工作中，如何在推荐系统中使用大数据算法如地图缩减来提高性能[28]。

**参考文献**

[1] Soanpet .Sree Lakshmi and Dr.T.Adi Lakshmi. "Recommendation Systems:Issues and challenges".(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5(4) ,2014, pp.5771-5772.

[2] Mohammed FadhelAljunid and Manjaiah D.H."A SURVEY ON RECOMMENDATION SYSTEMS FOR SOCIAL MEDIA USING BIG DATA ANALYTICS". International Journal of Latest Trends in Engineering and Technology, Special Issue (SACAIM 2017), pp.48-58.

[3] P.Priyanga , Dr.A.R.Nadira Banu Kamal. "Methods of Mining the Data from Big Data and Social Networks Based on Recommender System". International Journal of Advanced Networking & Applications (IJANA), vol 8(5) 2017, pp.55-60.

[4] Lalita Sharma and Anju Gera. "A Survey of Recommendation System: Research Challenges". International Journal of Engineering Trends and Technology (IJETT), vol 4(5) 2013, pp.1989-1992.

[5] Radhya Sahal, Mohamed helmy Khafagy, Fatma A. Omara,"Comparative Study of Multi-query Optimization Techniques using Shared Predicate-based for Big Data", International Journal of

Grid and Distributed Computing Vol. 9, No. 5 (2016), pp.229-240

[6] Mina samir shnoda,mohamed helmy khafagy,samah ahmed senbel,"JOMR: Multi-join Optimizer Technique to Enhance MapReduce Job",The 9th International Conference on INFOrmatics and Systems (INFOS2014),pp:80-86

[7] Hussien SH. Abdel Azez, Mohamed H. Khafagy & Fatma A. Omara ,"Optimizing Join in HIVE Star Schema Using Key/Facts Indexing", IETE Technical Review, 2017.DOI:10.1080/02564602.2016.1260498

[8] Shah Khusro, Zafar Ali and Irfan Ullah. "Recommender Systems: Issues, Challenges, and Research Opportunities". Information Science and Applications (ICISA) 2016, pp.1179-1189.

[9] Debashis Das, Laxman Sahoo,Sujoy Datta." A Survey on System". International Journal of Computer Applications ,vol 160(7), February 2017 , pp.6-10.

[10] Dr. Sarika Jain, Anjali Grover, Praveen Singh Thakur, Sourabh Kumar Choudhary." Trends, Problems And Solutions of Recommender System". International Conference on Computing, Communication and Automation (ICCCA2015) , (2015). pp.955-958.

[11] Yagnesh G. Patel, Vishal P.Patel. "A Survey on Various Techniques of Recommendation System in Web Mining", International Journal of Engineering Development and Research, 2015 IJEDR,Vol 3(4),2015,pp.696-700.

[12] Tanya Maan, Shikha Gupta, Dr. Atul Mishra." A Survey On Recommendation System",international conference on recent innovations in management,engineering,science and technology (RIMEST 2018), (2018).pp:543-549.

[13] Hadeer Mahmoud , Abdelfatah Hegazy , Mohamed H. Khafagy," An approach for big data security based on Hadoop distributed file system", 2018 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, 2018, pp. 109-114.

[14] Mustansar Ali Ghazanfar and Adam Prugel-Bennett,” A Scalable, Accurate Hybrid Recommender System”, 2010 Third International Conference on Knowledge Discovery and Data Mining.

[15] Qian Wang, Xianhu Yuan, Min Sun “Collaborative Filtering Recommendation Algorithm based on Hybrid User Model”, FSKD, 2010.

[16] Kumar, Vikas, et al. "Collaborative filtering using multiple binary maximum margin matrix factorizations." Information Sciences 380 (2017): 1-11.

[17] Zeng, Xuelin, et al. "Parallelization of Latent Group Model for Group Recommendation Algorithm." Data Science in Cyberspace (DSC), IEEE International Conference on. IEEE, 2016.

[18] Ponnam, Lakshmi Tharun, et al. "Movie recommender system using item-based collaborative filtering technique." Emerging Trends in Engineering, Technology and Science (ICETETS), International Conference on. IEEE, 2016.

[19] Halder, Sajal, AM Jehad Sarkar, and Young-Koo Lee. "Movie recommendation system based on movie swarm." Cloud and Green omputing (CGC), 2012 Second International Conference on. IEEE, 2012.

[20] Dev, Arpan V., and Anuraj Mohan. "Recommendation system for big data applications based on set similarity of user preferences." Next Generation Intelligent Systems (ICNGIS), International Conference on. IEEE, 2016.

[21] Marwa Hussien Mohamed and Mohamed Helmy Khafagy. “Hash semi cascade join for joining multi-way map reduce.” 2015 SAI Intelligent Systems Conference (IntelliSys)(2015): pp.355-361.

[22] Marwa Hussien Mohamed, Mohamed Helmy Khafagy, Mohamed Hasan Ibrahim . “Hash Semi Join Map Reduce to Join Billion Records in a Reasonable Time.” Indian Journal of Science and Technology, vol. 11, no. 18, Jan. 2018, pp. 1–9.,doi:10.17485/ijst/2018/v11i18/119112.

[23] Katarya, Rahul, and Om Prakash Verma. "A collaborative recommender system enhanced with particle swarm optimization technique." Multimedia Tools and Applications 75.15 (2016): 9225-9239.

[24] Panigrahi, Sasmita, Rakesh Ku Lenka, and Ananya Stitipragyan. "A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark." Procedia Computer Science 83 (2016): 1000-1006.

[25] Lee, Maria R., Tsung Teng Chen, and Ying Shun Cai. "Amalgamating Social Media Data and Movie Recommendation." Pacific RimKnowledge Acquisition Workshop. Springer International Publishing, 2016.

[26] Xie, Li, Wenbo Zhou, and Yaosen Li. "Application of Improved Recommendation System Based on Spark Platform in Big Data Analysis." Cybernetics and Information Technologies 16.6 (2016): 245-255.

[27] Radhya Sahal, Mohamed helmy Khafagy, Fatma A. Omara: Exploiting coarse-grained reused-based opportunities in Big Data multiquery optimization.J. Comput. Science ,volume 26,2018,pp: 432-452

[28] Marwa Hussien Mohamed, Mohamed Helmy Khafagy, Mohamed Hasan Ibrahim . "From Two-Way to Multi-Way: A Comparative Study for Map-Reduce Join Algorithms". WSEAS Transactions on Communications, Volume 17, 2018, pp. 129-141.