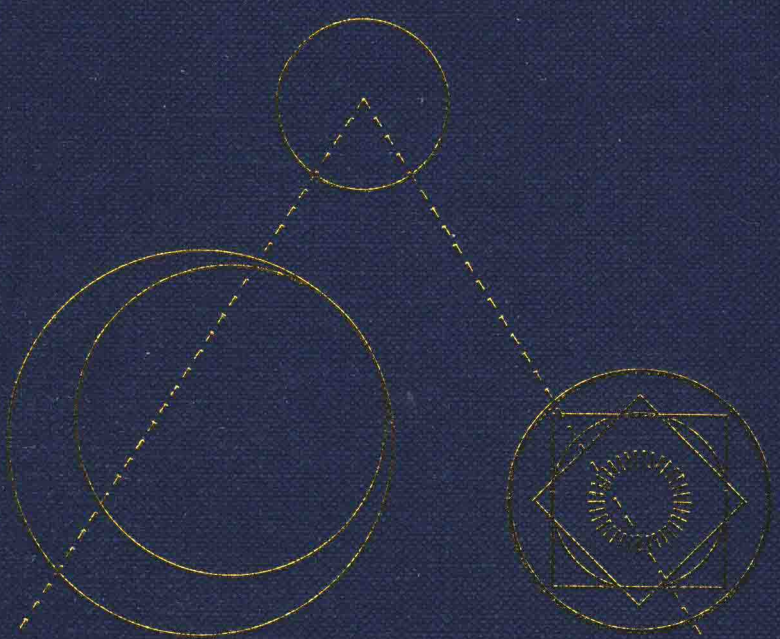
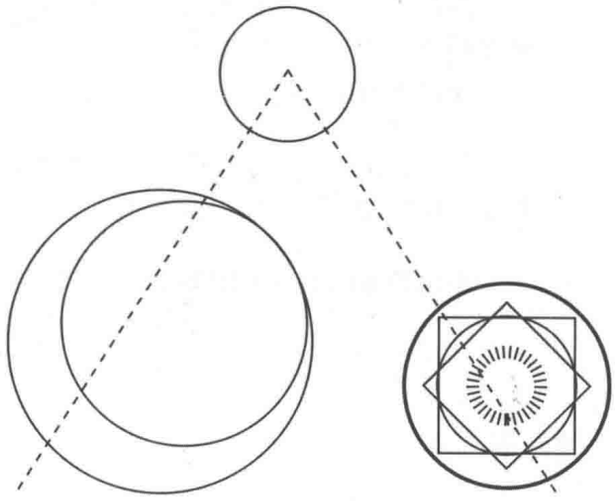
数据科学与人工智能研究

Research on Data Science and Artificial Intelligence

邓玉辉 主编





数据科学与人工智能研究

Research on Data Science and Artificial Intelligence

邱玉辉 主编

西南师范大学出版社
国家一级出版社 全国百佳图书出版单位

图书在版编目(CIP)数据

数据科学与人工智能研究 / 邱玉辉主编. —— 重庆 : 西南师范大学出版社, 2018.10

ISBN 978-7-5621-5560-7

Ⅰ. ①数… Ⅱ. ①邱… Ⅲ. ①数据处理－研究②人工智能－研究 Ⅳ. ①TP274②TP18

中国版本图书馆 CIP 数据核字(2018)第 237132 号

数据科学与人工智能研究

SHUJU KEXUE YU RENGONG ZHINENG YANJIU

邱玉辉　主编

责任编辑：罗　渝　李相勇　廖小兰

装帧设计：尚品 CASTALY 尹　恒

排　　版：重庆大雅数码印刷有限公司·王　兴

出版发行：西南师范大学出版社

地址：重庆市北碚区天生路 2 号

邮编：400715　市场营销部电话：023-68868624

经　　销：新华书店

印　　刷：重庆大雅数码印刷有限公司

幅面尺寸：290mm×215mm

印　　张：22

字　　数：601 千字

版　　次：2018 年 11 月　第 1 版

印　　次：2018 年 11 月　第 1 次印刷

书　　号：ISBN 978-7-5621-5560-7

定　　价：180.00 元

序

人工智能是计算机科学的一个重要分支，是研究模拟人类智能、智能行为及其规律的一门重要学科。人工智能自1956年提出以来，经历了几起几落的发展变化，如今随着深度学习的深入研究和有效应用，又一次展现出勃勃生机。国务院印发的《新一代人工智能发展规划》，明确指出了人工智能将成为国际科技竞争的新焦点。

当今时代是数据为王的大数据时代，并由此催生了数据科学这门新兴的学科。数据科学是一门多学科交叉的综合学科，包含数据获取、数据分析、数据管理、机器学习、统计优化和数据可视化等内容，逐渐成为探明大数据集本源，并把大数据转换成可执行智能的有效方法。随着国务院《促进大数据发展行动纲要》的实施，数据科学及其技术作为国家战略已迅速推进。

为了给广大的数据科学及人工智能研究人员提供一个交流与合作的平台，分享数据科学与人工智能领域的研究成果、创新思想以及最新进展，在我的倡导下，由重庆市人工智能学会、重庆市计算机学会和重庆计算机安全学会主办，陆军勤务学院军事物流系、重庆胜券科技有限公司协办，西南大学计算机与信息科学学院承办“数据科学及人工智能高峰论坛”。论坛中交流的论文经筛选，集结成册，《数据科学与人工智能研究》一书由此诞生。它也是我和我的学生们多年来在人工智能、数据科学、计算智能等多方面的研究和探索的成果选集。文集共征文六十余篇，经过三轮审稿，修改，定稿，最终确定了四十篇入选；同时我也推荐了我的学生们近年来在高水平杂志出版发表的论文摘要十篇，供对人工智能与数据科学感兴趣的读者参考。

由于作者水平有限，文集中不足和错误之处难免，欢迎读者批评指正。

衷心感谢对本次论坛给予大力支持的有关单位！感谢我的弟子们！

邱玉辉

2018年11月1日

目 录

第一部分 数据科学理论与方法

第二部分 数据处理技术及应用

第三部分 人工智能理论与方法

第四部分　人工智能技术及应用

第五部分　相关理论及应用

第六部分　已发表论文

第一部分

数据科学理论与方法

大数据测试面临的挑战与展望

任问筠　丁晓明　沈鹏

西南大学计算机与信息科学学院，重庆　400715

摘　要　随着云计算等技术的飞速发展，大数据时代的到来对各行各业信息处理能力提出了更高的要求，各种大数据技术应运而生，也给软件测试带来了挑战。本文分析了软件测试的ORACLE问题、杀虫剂效应问题、大数据隐私保护程度测试问题等，最后对未来软件测试技术进行了初步展望。

关键词　大数据；软件测试；挑战

1　引言

随着云计算、物联网等技术的快速发展，各行各业的数据量都出现了飞速增长。中国产业信息网报告指出，2012－2020年全球数据总量年增长率将维持在50%左右，到2020年全球数据总量将达到40ZB。40ZB相当于整个世界人口（截至2017年为76亿人）全年每天观看14.5h的高清视频流所产生的数据量。数据量的快速增长已经远远超越单个计算机的存储和处理能力[1]。为了快速地处理海量复杂的数据，尽可能地挖掘数据的潜在价值，全球掀起了大数据研究的热潮。

对于大数据，各领域从不同角度有着自己的见解，业界较为认可的是用“4V”来总结大数据的特征。第一，数据体量大（Volume）。存储单位从GB到TB，再到PB、EB，甚至全球数据量已经以ZB计算。大型社交网络、购物网站每天的数据都达到TB级别。第二，数据类型多（Variety）。广泛的数据来源导致了大数据类型的多样性。大体可分为结构化数据、非结构化数据和半结构化数据三大类，其中结构化数据中，如信息管理系统等，数据间有较强的因果关系；非结构化数据，如视频、图片等，数据间没有因果关系；半结构化数据如HTML文档等，数据间有较弱的因果关系。第三，处理速度快（Velocity）。处理大数据的响应时间以秒计算，实时分析，立竿见影。第四，价值密度低（Value）。大数据的价值就是从大量不相关的数据中，通过深度分析获得少量有价值的数据，发现新规律、新知识，对未来进行预测分析[2]。

随着大数据的发展，各种针对大数据的架构以及智能应用应运而生。大数据背景下的软件形态不断演化，软件测试也随之面临巨大的变革。由于大数据本身的特征，大数据架构的复杂性，以及传统测试方法几乎不再适用等各方面因素，国内外大数据测试的能力还较为薄弱。大数据测试在发展的同时还面临着巨大的挑战。

2 大数据测试面临的挑战

2.1 大数据背景下的 ORACLE 问题日益突出

软件测试有“证伪”和“求真”两种，但是其基本前提都是在确定的输入下，存在确定的输出。测试需要将软件的实际结果和预期的结果相比较，从而得出软件运行正确与否，这就是软件测试的 ORACLE 问题。大数据背景下，不论是趋势分析法还是图论计算法都变得复杂，导致无法直接获得预期的输出结果，从而软件正确与否无法直接判断[2]。

大数据处理有两种模式，即基于化学作用和基于物理作用的处理模式[3]。物理作用的处理是指在不改变数据属性的情况下进行数据清洗，以减小数据规模，同时不损失数据价值。如此，使数据规模由大变小的测试的 ORACLE 不存在问题。

而基于化学作用的数据处理的核心是图的快速算法和预测。很多应用的输出结果没有明确的对错之分，只能以模糊的好坏区分。这使得 ORACLE 问题变得困难。例如，在今天的电子商务领域，人工智能技术被广泛应用以达到精准营销的目的，从而为用户提供个性化推荐，或者进行潜在用户的挖掘。但是对于推荐的商品，用户可能喜欢，也可能不喜欢，这就使结果成为一个概率性问题，其结果正确与否无法像确定性问题一样得到判断，这就导致了 ORACLE 确定困难。

2.2 测试数据输入问题

大数据处理是为了得到数据的分布特性，进而达到对未来同类情况进行预测的目的[2]。而数据的分布特性就是数据间的某种规律，由数据间的相互关系体现。而这种关系则需要数据量达到一定程度时才能反映出来，少量数据是无法感知数据间隐含的逻辑关系的。

如果要采用大量的数据输入，这个合适的数据量的大小又成了亟待解决的问题。如果采用少于全部数据的数据集，很有可能得不到同等的分布特性甚至无法分析出结果；如果采用与全部数据等价的数据集，又必将消耗巨大的成本，而且这种做法的必要性也值得思考。

2.3 数据类型繁多导致的判断问题

大数据的基本特征之一就是数据类型多。而且就目前社会发展现状而言，主要是非结构化和半结构化数据迅速增长，相对而言，结构化数据的增长则较为缓慢。其中，半结构化数据具有一定的结构性，但不方便模式化。结构化数据能够用数据或统一的结构加以表示，传统的关系数据模型存储于数据库，可用二维表结构表示，可以在软件测试时验证其正确性。非结构化数据一般指无法结构化的数据，没有格式。尽管一些自动化工具可以抽取非结构化数据的内容，但是仍存在两方面问题。一是由于数据本身存在异常，无法检测数据的正确性。二是在测试过程中，数据格式发生一定变化。不论是在数据输入还是数据输出时，都会由于繁多的数据类型导致类似的问题。比如当数据输出端输出的是海量的非结构化数据时，由于输出数据的分布特性，判断会变得更加困难。

2.4 传统测试平台无法满足大数据处理的需求

传统的测试平台通常应对较少数据量的测试。如性能测试时，对于并发用户数在百、千量级的应用服务，其应用系统由少量服务器构成，基于局部的物理机进行压力测试的方式能够满足应用的需

求[3]。在大数据背景下，数据处理平台均架构于可动态扩展的 Paas 平台，以应对数据爆炸性增长。早期具有代表性的就是 Hadoop 平台，其数据处理的软件可以架构于千万级服务器的资源上。由此，给测试带来了前所未有的困难。由于规模巨大，测试客户端的能力可能无法达到服务器端的测试需求[4]。

2.5 大数据框架的复杂性带来的测试挑战

大数据框架用于收集、整理、处理大容量数据集。随着时代的发展，各种大数据处理框架应运而生。其中仅批处理框架有 Apache Hadoop 等，仅流处理框架有 Apache Storm、Apache Samza 等，混合框架有 Apache Spark、Apache Flink 等，框架的功能越来越丰富。比如较为主流的大数据处理系统 Hadoop 框架的 MapReduce，该系统的数据处理过程主要分为 Map 和 Reduce 两个阶段。用户在使用时的主要工作就是实现 Map( )和 Reduce( )两个函数，而其中的细节部分都交由 MapReduce 系统进行处理[5]。在这样的情况下，框架自身所具有的功能远远大于用户操作部分的功能，也就是说框架自身具有较高的复杂性，而用户能掌握的信息又很少。因此，测试面临着巨大的困难和挑战。

2.6 杀虫剂效应日益显现

软件测试行业中的“杀虫剂效应”是用于描述测试人员对同一测试对象进行测试的次数越多，发现的缺陷就会越来越少的现象[5]。就像农药一样，如果经常使用单一的农药，害虫的抗药性会越来越强，农药的效果就会越来越弱。在构件开发的过程中，中前期会发现一些模式的缺陷，这些缺陷经过校验后集成在构件里，成为构件的一部分。这些构件就对已有的测试方法具有免疫力。复杂的大数据处理软件再由多种构件集成，就更是对测试有天然的免疫力。

面对杀虫剂效应，软件测试技术必须进行不断的升级，才能发现软件中的缺陷。原本在测试初期，少量的测试用例会发现较多的软件缺陷，后期能够发现的软件缺陷逐渐变少，甚至缺陷数量停止增长。而随着杀虫剂效应的显现，在测试初期，发现缺陷的技术难度已经增加，同样的测试用例可能发现的缺陷数量在减少，缺陷的总数量也会降低。

2.7 软件服务化引发的测试问题

从开发模式的角度来看，软件开发主要经历了 4 个阶段，包括完全编码阶段、构件化阶段、服务阶段、云计算阶段[6]。在完全编码阶段，开发人员人工编码每一行代码，可以对代码进行完全掌握。所以在该阶段，软件具有完全的可测性，即可以使用任何测试方法。在构件化阶段，为了提高开发效率，大量使用可复用的组件。由于组件是运行在本地的，开发人员还是对代码有较大程度的掌握。用户仍然可以对软件运行的结构进行追踪。在服务阶段，大量组件的提供方式变成了远程调用的方式。用户对服务的掌握减小，随之可测性逐渐减小。如今，到了云计算阶段，尤其是在 Paas 模式下，用户能了解的只剩下输入和输出接口。由此，用户基本完全丧失了对程序运行情况的了解。因此，用户测试的难度进一步加大。

2.8 大数据隐私保护程度测试问题

大数据应用在处理分析大数据获取有价值的信息的同时，数据的安全和隐私保护存在很大的问题。为了降低数据隐私泄露的风险，一般通过数据扰乱和数据加密两种主要途径对数据进行一定的隐私保护[2]。数据扰乱是对原始数据进行一些修改以隐去敏感部分。数据加密如同态加密，是将数据先加密再运算得出结果。而对于这些隐私保护效果的评估还存在很多的困难。

现有的隐私保护效果评估方法主要分为两大类。一类是将隐私保护后的数据与原始数据的信息本身进行对比，提出一些计算方法。另一类是度量相关联数据的信息的量。虽然已被提出的多种评估方法都具有一定的作用，但仍然存在一些问题。第一类隐私保护效果评估方法作为通用方法是具有一定局限性的。大数据具有价值密度低的特征，同样的，敏感数据也可能只存在于较小部分的数据中，这导致数据信息本身的变化程度与隐私保护效果可能有较大差异。当没有低密度的特点时，例如去隐私时采用了将数据打乱排序的方法，当敏感数据是从数据间的相互关联中获得而与顺序无关时，就会导致数据与原始数据差异较大，却没有起到隐私保护的效果。第二类评估方法较第一类更加有针对性，它是以相关联数据的信息的量作为依据。但是当表面看似没有关联的数据经过处理分析之后，就有可能暴露出隐私信息。

另一方面，隐私保护程度越高，就意味着数据的损失程度越大，隐私保护效果的评估结果多少为最佳并没有标准，这也是一个有待解决的问题。

3 结语

随着大数据的不断发展，大数据背景下的软件测试能力在逐步提高，同时也带来了挑战。面对日益显现的杀虫剂效应问题，可以增加测试技术的多元性，不断更新测试方法，采用传统方法与新方法结合的模式，采用不同测试人员测试同一软件的方式，进一步解决问题。面对大数据测试的输入问题，可以从大小、分布特性等方面研究合适的样本集。面对大数据隐私保护程度的测试问题，可以有针对性地采取更细化的评估方法，制定评估标准。未来，大数据测试技术还有长足的发展。

参考文献

[1] 中国产业信息中心. 2017年全球及中国数据中心发展前景分析及预测[EB/OL]. (2017－10－18)[2018－06－23]. http://www.chyxx.com/industry/201710/573390.html.

[2] 蔡立志，武星，刘振宇. 大数据测评[M]. 上海科学技术出版社，2015：137.

[3] 蔡立志，阎婷. 大数据背景下软件测试的挑战与展望[J]. 计算机应用与软件，2014(2)：5－8.

[4] 合云峰. 大数据背景下软件测试的挑战与展望[J]. 通讯世界，2016(8)：34－35.

[5] 黄瑞国. 大数据测试技术的特点及前景研究[J]. 电脑知识与技术，2016，12(27)：1－2.

[6] 李宁，庄丽华，石林，等. 大数据云计算时代软件测试所面临的挑战[J]. 教育教学论坛，2017(51)：275－276.

[7] 周晓云，覃雄派，王秋月. 大数据评测基准的研发现状与趋势[J]. 计算机应用，2015，35(4)：1137－1142.

[8] 宋杰，李甜甜，朱志良，等. 云数据管理系统能耗基准测试与分析[J]. 计算机学报，2013，36(7)：1485－1499.

[9] 代亮，陈婷，许宏科，等. 大数据测试技术研究[J]. 计算机应用研究，2014，31(6)：1606－1611.

[10] 揣立武. Hadoop平台基准性能测试工具的设计与实现[D]. 哈尔滨工业大学，2015.

[11] Garg N，Singla S，Jangra S. Challenges and Techniques for Testing of Big Data[J]. Procedia Computer Science，2016，85：940－948.

[12] S. Nachiyappan，S. Justus. Understanding and Addressing Technical Challenges for Big Data Testing[J]. International Journal of Engineering Science and Computing，2016，6(3)：2132－2135.

[13] Alexandrov A，Markl V. Issues in big data testing and benchmarking[C]// International Workshop on Testing Database Systems. ACM，2013：1－5.

[14] Abidin A，Lal D，Garg N，et al. Comparative analysis on techniques for big data testing[C]// International Conference on Information Technology. IEEE，2017：219－223.

[15] Hong M. Research On Software Testing Technology Under Big Data Background[C]// Conference on Informatization in Education，Management and Business，2015.

[16] Nagdive A S, Tugnayat R M, Tembhurkar M P. Overview on Performance Testing Approach in Big Data [J]. International Journal of Advanced Research in Computer Science,2014,5(8):165－169.

[17] Chaitanya Kadam, Yasoda Thapa. Big Data: Features, Challenges & Solutions[J]. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),2016,5(5):1649－1652.

[18] Nachiyappan S, Justus S. Getting ready for Big Data testing: A practitioner's perception[C]// Fourth International Conference on Computing, Communications and Networking Technologies. IEEE,2014:1－5.

[19] Singh K K, Dimri P, Rohatgi S. Cloud testing and authentication model in financial market Big Data analytics [C]// International Conference System Modeling & Advancement in Research Trends. 2016:242－245.

[20] Mahesh Gudipati, Shanthi Rao. Big Data:Testing Approach to Overcome Quality Chanllenges. Infosys Labs Briefing,2013,11(1):65－72.

稀疏聚类和用户信任度结合的协同过滤算法

侯宇博　唐雁

西南大学计算机与信息科学学院，重庆　400715

摘　要　针对在数据稀疏情况下聚类效果不理想、维度约简时舍去过多有用信息的问题，采用稀疏子空间聚类算法对用户聚类，保留更多有用信息；在相似度计算中结合用户信任度进行计算：首先计算用户在数据集上的有效、公正评分，为每个用户建立用户的可信任度矩阵，然后将改进的基于评分众数的用户信任度融合进去，结合传统的相似度量方法进行相似度计算。在电影数据集合上的实验结果表明，算法可以积极缓解在数据稀疏情况下查找近邻不精准的问题，提高推荐质量。

关键词　数据稀疏；稀疏子空间聚类；用户信任度；电影推荐；协同过滤

1　引言

由于互联网技术日新月异和社交网络的快速发展，人们正在进入一个信息爆炸的时代，互联网出现了信息的冗余与信息过载的问题。个性化推荐系统的出现就是为了解决信息过载的问题，它分析用户的习惯和偏好，向用户推荐其有可能感兴趣的项目，协同过滤算法是现行推荐系统中被广泛应用，也是最成功的推荐算法之一[1]。协同过滤这一概念首次于 1992 年由 Goldberg D 等人提出[2]，它主要基于如下假设[3]：具有相似兴趣的用户会喜欢相似的项目，可以根据用户项目的评分信息找到用户或项目之间的关联，根据这种关联向用户推荐项目。协同过滤推荐技术广泛应用于各大流行的推荐系统中，有着很高的实用价值。随着推荐系统应用中商业网站规模越来越大，用户通常只会对少部分物品进行评分[4]，数据稀疏性越来越大。协同过滤推荐技术依据用户历史行为进行偏好预测，用户项目评分矩阵的稀疏性会导致不正确的推荐结果，影响预测质量。针对数据稀疏性问题，现有的技术在空值填补和新的相似性计算方法方面做了很多的研究和尝试。

邓爱林等[5]填补用户评分项并集中的空值，在填补后的并集上计算相似度，提高了推荐的效率。李聪等[6]采用领域最近邻方法预测评分项目中的未评分值。张锋等[7]将 BP 神经网络引入对用户的评分预测中。冷亚军等[8]使用混合用户和项目的协同过滤方法，对评分矩阵中循环填补。石教开[9]采用基于聚类和信任度的评分算法，引入评分众数进行信任度的计算。这些都对数据稀疏性问题进行了缓解，但在聚类过程中会将一些有用的信息当作冗余信息舍弃，新的相似性方法的研究也未考虑到用户的信任度问题。

针对上述问题提出一种基于稀疏聚类和用户信任度的协同过滤算法，首先用稀疏子空间聚类算法对用户项目评分矩阵进行聚类，减少有用信息的舍弃；其次，提出了用户评分的概念，在用户电影数据集合上统计用户对于电影有效、客观的评分数目，建立用户信任评分集；然后，统计得到用户在信任评分集上的评分众数，计算用户和用户之间的信任度；最后结合信用评分、信任度和传统的相似性计算方法计算用户之间的相似度，对用户相似度进行从小到大的排序，选出相似度最大的前 N 个用户作为最近邻集合，对项目进行预测。在 MovieLens 数据集上的结果表明，该方法可以有效提高推荐系统的推荐精度。

2 相关知识

2.1 基于用户的协同过滤技术

基于用户的协同过滤技术[10]认为具有相似兴趣的用户会对相同的项目感兴趣。例如，我们发现用户 A 和 B 喜好非常相似，那么我们可以将用户 A 喜欢的项目推荐给 B。

2.2 稀疏子空间聚类

稀疏子空间聚类[11]有较快的聚类速度。它利用评分矩阵数据的稀疏表示，构造数据的相似度矩阵，对低维子空间中的数据聚类，利用谱聚类[12]获得最终结果。

2.3 传统相似度计量方法

$$\mathrm{Sim}(u,v)=\frac{\sum_{i\in iuv}(R_{u,i}-\overline{R_u})(R_{v,i}-\overline{R_v})}{\sqrt{\sum_{i\in iu}(R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{i\in iv}(R_{v,i}-\overline{R_v})^2}} \tag{1}$$

其中：$R_{u,i}$ 表示用户 u 对物品 i 的评分，$\overline{R_u}$ 表示用户 u 对物品 i 评分的平均值。$R_{v,i}$ 表示用户 v 对物品 i 的评分，$\overline{R_v}$ 表示用户 v 对物品 i 评分的平均值。

3 算法框架

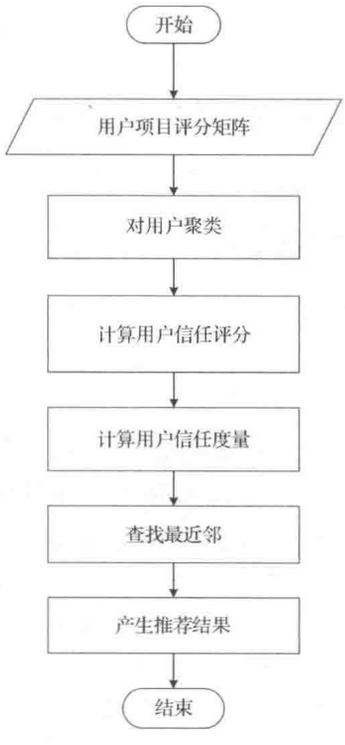


图 1　基于稀疏聚类和用户信任度的协同过滤算法流程图

3.1 用户项目评分矩阵

对数据初始化，形成用户－项目评分矩阵 R，用户－项目评分矩阵 R 中包含 m 个用户的集合 $U=\{u_1,u_2,\cdots,u_m\}$ 和 n 个项目集合 $I=\{i_1,i_2,\cdots,i_n\}$。

3.2 对用户聚类

采用稀疏子空间聚类算法对 R 进行聚类，生成 k 个用户聚类簇，找出目标用户 u 所在的类 Cluster_u，通过类 Cluster_u 中的用户评分信息对目标用户 u 中未评分的项目进行初始预测评分。具体步骤如下：

(1)把用户－项目评分矩阵 R 作为输入，得到稀疏系数矩阵 C；

(2)对矩阵 C 的每一列进行正则化，$c_i \leftarrow \frac{c_i}{\| c_i \| \infty}$；

(3)由稀疏系数矩阵 C 得到构造相似度矩阵 W；

(4)利用谱聚类算法对相似度矩阵 W 进行划分，得到 W 的 k 个子空间聚类簇，找出目标用户 u 所在的类 Cluster_u。

3.3 计算用户信任评分

本文认为在评分中有效评分项目越多，评分越公正，即越接近项目的真实评分，那么用户的可信度就越高。我们设置一个阈值 μ，根据经验值 μ 取 0.3。可信任评分集合用 S_u 来表示。计算公式如下：

$$S_u=\{i\in S\left|\frac{r_{u,i}-\overline{r_i}}{\overline{r_i}}\leqslant\mu\right\} \tag{2}$$

上式中 $r_{u,i}$ 表示用户对电影 i 的评分，$\overline{r_i}$ 表示电影 i 的真实评分。本文用 TW 表示用户可信任度，那么用户 u 的用户可信任度为 TW_u。另外借用 sigmoid 函数，我们把评分数目的影响 $\frac{1}{1+e^{-|s|}}$ 控制在 $[\frac{1}{2},1]$ 之间，用户 u 的用户可信任度 TW_u 计算公式如下：

$$TW_u=\frac{|S_u|}{|S|}\frac{1}{1+\mathrm{e}^{-|s|}} \tag{3}$$

上式中，S 代表用户对所有电影的评分数目。用户评分过的电影越多，可以信任的评分越多，用户可信任度就越高。

3.4 计算用户信任度量

用户对电影都有不同的偏好，有的会喜欢喜剧电影，有的则偏爱爱情文艺类电影。用户会对他们喜欢的电影给予高于平均值的评分，而对自己讨厌的电影类型给予低于平均分的分值。所以单纯依靠传统的相似性计算方法会产生计算偏差。用户之间有着信任度的存在，在用户可信任集合中对用户 u 和用户 v 的信任度定义如下：

$$\text{trust}_{u,v}=\frac{|S_u\cap S_v|}{\min(S_u,S_v,\eta)} \tag{4}$$

融合用户可信任度和用户之间信任度量，可以得到用户 v 对用户 u 的信任度表示关系。M_{TW}-$\text{trust}_{u,v}$ 的公式如下所示：