国家计量技术规范

《基于用户黏着度的大数据推荐算法测评方法》

编制说明

《基于用户黏着度的大数据推荐算法测评方法》编制组

2024年6月

**《基于用户黏着度的大数据推荐算法测评方法》编制说明**

**一、任务来源及计划要求**

根据国家市场监督管理总局计量技术法规制修订任务的安排，（监计量发[2022]70号市场监管总局办公厅关于印发2022年国家计量技术规范项目制定、修订及宣贯计划的通知），中国计量科学研究院、福建省计量科学研究院受全国数字计量技术委员会委托，作为主要起草单位负责牵头制订《基于用户黏着度的大数据推荐算法测评方法》，参加起草单位包括：华侨大学，之江实验室等。

**二、项目意义**

近年来，算法应用在给政治、经济、社会发展注入新动能的同时，算法歧视、“大数据杀熟”、诱导沉迷等算法不合理应用导致的问题也深刻影响着正常的传播秩序、市场秩序和社会秩序，给维护意识形态安全、社会公平公正和网民合法权益带来挑战。在互联网信息服务领域出台具有针对性的算法推荐规章制度，是防范化解安全风险的需要，也是促进算法推荐服务健康发展、提升监管能力水平的需要。

国家市场监管总局发布的《关于平台经济领域的反垄断指南》规定，基于大数据和算法，根据交易相对人的支付能力、消费偏好、使用习惯等，实行差异性交易价格或者其他交易条件；对新老交易相对人实行差异性交易价格或者其他交易条件；实行差异性标准、规则、算法;实行差异性付款条件和交易方式等，都可能被认定为“大数据杀熟”等不正当竞争行为而面临更严格监管。

本规范是对大数据推荐算法合理性进行评估，辅助市场监管机构对电商领域实行差异性交易和服务等“大数据杀熟”行为进行监管，有利维护消费者权益，促进社会公平。主要意义如下：

1、服务市场监管：辅助对开发大数据推荐算法的电商企业进行监管，防止其利用算法对消费者进行差异性交易和服务。对于提高监管机构的履职能力和政府公信力的提升具有一定作用。

2、服务电商企业：通过对算法的测评，提高企业的知名度和品牌影响力，同时赢得消费者的信任。

3、服务消费者：目前大数据杀熟已经层出不穷，如不同的手机或者不同的运营公司打车出行的价格不同、电商平台根据个人特征进行差别化的交易等，根据此标准对相关企业进行第三方算法测评，并通过相关平台进行社会公布，将起到增强消费者信心，维护社会公平正义的作用。

**三、编制依据**

本规范依据和参考JJF 1182—2021《计量器具软件测评指南》、JJF 1071-2010《国家计量校准规范》、GB/T 35295 -2017《信息技术 大数据 术语 第2部分：大数据 Big Data》、GB/T 8567《计算机软件文档编制规范》、GB/T 9385《计算机软件规格需求说明规范》等基础性系列规范进行制定。

本规范规定了基于内容的大数据推荐相关软件的算法标识、用户满意度、预测准确度、覆盖率、置信度、多样性等技术指标及测评方法，适用于对大数据推荐算法进行测评。

本规范为首次发布。

**四、编制过程**

为规范互联网信息服务算法推荐活动，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益，促进互联网信息服务健康发展，国家互联网信息办公室、工业和信息化部、公安部、国家市场监督管理总局联合发布《互联网信息服务算法推荐管理规定》，自2022年3月1日起施行。规定明确了算法推荐服务提供者的信息服务规范，要求算法推荐服务提供者应当坚持主流价值导向，积极传播正能量，不得利用算法推荐服务从事违法活动或者传播违法信息，应当采取措施防范和抵制传播不良信息。

本规范的编制过程中，严格按照《互联网信息服务算法推荐管理规定》相关规定，并与国内从事相关专业工作的专家学者们进行广泛的交流和探讨，得到了他们的大力支持和帮助，为起草技术规范的制定奠定了良好的基础。主要起草与编制过程如下：

1）2022年12月初组成技术规范起草组，并就规范所包含的内容、技术指标、参考数据集、推荐算法测评方法等问题进行了初步讨论；

2）2023年2月底，起草组初步征求并听取了从事大数据推荐相关算法的专家意见，对前期开展的起草工作进行相关的内容补充，在此基础上完成初稿的总体框架；

3）2023年3月至2023年10月，调研收集相关技术资料并参考相关国家、行业标准撰写本技术规范初稿；

4）2023年10月，数字计量工作组西安召开会议讨论本技术规范文本，对本技术规范文本进行了讨论，并对技术规范中存在的问题提出相应的修改意见；

5）2023年11月，起草组召依据西安召开会议讨论的修改意见，对本技术规范面向的框架、内容进行了讨论与重新修改；

6）2023年12月，起草组召开会议讨论本技术规范文本，对本技术规范面向使用对象、黏着度定义、测评算法的选择进行了讨论，并对技术规范中存在的问题提出相应的修改意见；

7）2024年1月，数字计量工作组在厦门召开学术会议，对本技术规范文本进行了满意度等技术指标进行讨论，补充并完善了技术规范相应技术指标内容；

8）2024年1月，秘书长崔伟群、起草组召开线上会议，对本技术规范文本中的测评指标、黏着度定义、面向使用对象以及进行二次讨论，补充并完善了技术规范相应技术指标内容；

9）2024年2月，秘书长崔伟群、起草组召开线上会议，对本技术规范文本进行了三次讨论，会上就技术规范中的“用户黏着度、推荐算法”的定义进行完善，并补充“用户满意度、预测准确度、覆盖率、信任度、多样性”等关键参数的技术指标内容。

**五、主要技术内容的说明**

本规范引用了下列文件：

JJF 1182—2021 计量器具软件测评指南

JJF 1071-2010 国家计量校准规范

GB/T 35295 -2017 信息技术 大数据 术语 第2部分：大数据 Big Data

GB/T 8567 计算机软件文档编制规范

GB/T 9385 计算机软件规格需求说明规范

本规范是根据大数据推荐算法特性评定方法，本着科学、合理、规范的原则，结合大数据推荐算法的工作原理、国家标准、行业标准、专家意见和建议，以大数据推荐内容和黏着度为前提，明确大数据推荐算法测评要求，解决其推荐可信度测评方法，制订本计量技术规范。

本规范规定了基于内容的大数据推荐相关软件的算法标识、用户满意度、预测准确度、覆盖率、置信度、多样性等技术指标及测评方法。其中，使用的“用户满意度、预测准确度、覆盖率、置信度、多样性”五个评价指标对推荐算法的推荐合理性进行测评，评价指标的结果数值越高，说明其算法推荐的越合理。其中，对于参考数据集Ⅰ，本规范选取了CSD、RMSE、MAE、Coverage、Confidence、HD作为评价指标，但是对于参考数据集Ⅱ，RMSE和MAE需要计算误差，不再适用于参考数据集Ⅱ，本规范采取了HR来替代RMSE和MAE，相同的置信度Confidence也是参考数据集Ⅰ，也不再适用于参考数据集Ⅱ。在参考数据集Ⅱ中由于计算HD时间复杂度过高，因此考虑选择香农熵Entropy代替。

本规范依据大数据推荐相关内容将推荐信息的数据集划分成参考数据集Ⅰ和参考数据集Ⅱ，这两类数据集应分别包含至少数十万条由用户行为，且所有数据在确保真实性的前提下，必须遵守隐私保护和数据保护法规，对敏感数据进行恰当处理。本规范的测评以符合参考数据集Ⅰ和参考数据集Ⅱ要求的MovieLens-1M数据集、UserBehavior数据集为训练数据集，在AutoRec算法、DGRec算法模型下训练得到参考结果。

表1 MovieLens-1M数据集在AutoRec算法下的的参考结果

|  |  |  |
| --- | --- | --- |
| 评价指标/算法 | AutoRec：top10 | 建议结果（≥） |
| 准确度（RMSE） | 0.8129±0.0036 | 0.8093 |
| 准确度（MAE） | 0.6520±0.0036 | 0.6484 |
| 满意度（CSD） | （36.81±0.49）% | 36.32% |
| 覆盖率（Coverage） | （22.53±0.26）% | 22.27% |
| 置信度（Confidence） | 0.3445±0.0018 | 0.3427 |
| 多样性  （Hamming Distance） | 0.8672±0.0105 | 0.8567 |

表2 UserBehavior数据集在DGRec算法下的参考结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评价指标/算法 | DGRec：top100 | 建议结果(≥） | DGRec：top300 | 建议结果(≥) |
| 准确度（HR） | 0.2967 ± 0.0001 | 0.2966 | 0.4727 ± 0.0004 | 0.4723 |
| 满意度（CSD） | 0.0043±0.0001 | 0.0042 | 0.0029±0.0001 | 0.0028 |
| 覆盖率（Coverage） | （4.58±0.01）% | 4.57% | （9.22±0.02）% | 9.20% |
| 多样性  （Entropy） | 0.6409 ± 0.0019 | 0.60390 | 0.6079 ± 0.0050 | 0.6049 |

其中，表1显示了基于参考数据集Ⅰ要求的MovieLens-1M数据集在AutoRec算法下运行5次的测评结果，以此分别得到5次运行的评价指标结果。再将5次指标结果的平均值，作为结果的中心点，将结果的标准差，作为结果的正负误差以了解结果分布的泛化情况，最终得出基于MovieLens-1M数据集的AutoRec算法测评参考结果，评价指标的结果数值越高，说明其算法推荐的越合理；表2显示了基于参考数据集Ⅱ要求的UserBehavior数据集在DGRec算法下的参考结果运行了3次的测评结果，分别得到3次运行的评价指标结果。将3次指标结果的平均值，作为结果的中心点，将结果的标准差，作为结果的正负误差以了解结果分布的泛化情况，最终得出基于UserBehavior数据集的DGRec算法测评参考结果。将得到的参考结果再与测评对象的结果进行对比测评，最终得出大数据算法推荐结果的合理性。评价指标的结果数值越高，说明其算法推荐的越合理。此外，另需说明的是，为了避免读者后期在使用本规范中出现对测评结果值读数上的疑义，本规范在发布的规范文本中去掉了表1和表2中的测评结果的正负误差值，被测评方的测评指标结果值只需大于标准结果，便可认定为其推荐算法是合理。

**六、工作小结**

此次《基于用户黏着度的大数据推荐算法测评方法》国家计量技术规范的编写，起草组不仅对大数据推荐算法的工作原理有了更进一步的深入研究，而且在基于用户黏着度的大数据推荐算法评价应用中也做了大量的测评工作，提升了起草组技术人员的算法测评能力和技术水平。由于我们的水平和能力有限，难免有些差错和考虑不周之处，还请各位专家和委员多提宝贵意见，以便我们逐步完善本规范，在此，向为我们提出意见并付出辛勤劳动的专家委员和有关单位致以真诚感谢。

《基于用户黏着度的大数据推荐算法测评方法》编制组

2024年6月