

基于大数据技术的专家知识库设备画像 推荐算法研究

王 焯¹, 郭玲利², 宋文超³, 杨善友¹, 程 龙¹

(1. 科学技术部科技人才交流开发服务中心, 北京 100045;

2. 北京科技大学 自动化学院, 北京 100083;

3. 广道高新(北京)科技有限公司, 北京 100081)

摘要: 为了解当前设备画像推荐算法采用的训练数据较少, 无法充分体现设备故障特征, 存在数据稀疏问题和扩展性问题的弊端, 提出一种基于大数据技术的设备画像推荐算法; 将专家知识库控制策略资源作为推荐资源, 通过开源 MapReduce 技术完成对其的处理, 把开源怀卡托智能分析环境和 Hadoop 大数据处理平台结合在一起, 为设备画像推荐算法提供依据; 通过 k-means 算法对设备故障进行聚类; 利用评分矩阵相似性和设备画像特征相似度加权的思想对相似性进行计算, 将相似度最大的前若干结果作为最近邻; 为了保证推荐准确性, 通过奇异值分解算法, 针对评分矩阵通过各行设备故障评分均值对各行缺少值进行填充, 获取填充评分矩阵, 为目标故障推荐专家知识规则和控制策略; 经验证, 所提算法推荐准确性高。

关键词: 大数据技术; 设备画像; 推荐; 工业控制

Research on Personas Recommendation Algorithm Based on Big Data Technology

Wang Ye¹, Guo lingli², Song Wenchao³, Yang Shanyou¹, Cheng Long¹

(1. Exchange, Development & Service Center for Science & Technology Talents, The Ministry of Science & Technology (MoST), Beijing 100045, China; 2. School of Automation and Electrical Engineering,

University of Science & Technology Beijing, Beijing 100083, China;

3. Suntang High-technology (Beijing) Co.Ltd2., Beijing 100081, China)

Abstract: In order to solve the problem of sparse data and scalability, a new recommendation algorithm for device based on large data technology is proposed. Large-scale multi-disciplinary devices resources are used as recommendation resources, and their processing is completed by open source MapReduce technology. The open source Waikato intelligent analysis environment and Hadoop large data processing platform are combined to provide a basis for the recommendation algorithm of device. K-means algorithm is used to device fault. The similarity is calculated by weighting the similarity of the scoring matrix and the feature of the device. The first results with the largest similarity are taken as the nearest neighbors. In order to ensure the accuracy of recommendation, singular value decomposition (SVD) algorithm is used to fill in the missing values of each row through the average of each fault score matrix, and the filled score matrix is obtained to recommend control strategies to the target devices. It is verified that the proposed algorithm has high accuracy.

Keywords: big data technology; device portrait; recommendation; industrial control

0 引言

工业大数据是新工业革命转型升级的重要战略资源。而真正为企业带来价值的是数据分析和挖掘之后产生的结果, 通过工业大数据的分析, 可以帮助企业提高运行效率, 带来新的利润和业绩。但在我国, 工业控制系统中关键系统^[1-2]设备的份额大多为国外主流厂商, 尤其是高端设备。这些设备的故障代码和处理方式大多不掌握在我们手中, 这就需要通过机器学习和日常处理故障时, 将专家的

处理方式形成专家知识库, 而后根据出现的故障信息, 从专家知识库中获取最新的预测结果和解决方案。怎样快速匹配到问题, 及时找到处理问题的规则和策略成为亟待解决的问题。

设备画像就是根据自动化设备在运行过程中留下的数据, 主动或被动采集基本信息, 再针对设备本身提取有效信息当成标签, 按照标签信息构造设备的抽象模型^[3]。通过设备画像推荐算法, 为用户推荐合适的专家知识库规则, 为当前设备的故障诊断进行及时响应并提供解决方案。基于大数据的设备画像技术已渐趋成熟, 但对于设备画像推荐算法采用的训练数据较少, 无法充分体现设备信息特征, 存在数据稀疏问题。为此, 提出一种新的基于大数据技术

收稿日期: 2018-04-17; 修回日期: 2018-05-15。

作者简介: 王 焯(1981-), 女, 工学博士, 助理研究员, 主要从事电子文件、数据挖掘、科技人才管理等方向的研究。

的设备画像专家知识库推荐算法。

1 基于大数据技术的设备画像推荐算法

1.1 大数据平台技术架构及数据挖掘工具选择

文章选择工业大数据中专家知识库资源作为推荐资源,通过 Hadoop 建立大数据信息存储和处理平台^[4],把专家知识的基本信息、规则数据和故障描述、处理方式等信息存储至 HDFS (Hadoop Distributed File System) 中,同时通过开源 MapReduce 技术完成对大规模数据的计算与处理^[5]。

MapReduce 技术的关键部分主要包括 Map 与 Reduce,两个部分的输入与输出为 <key, value> 的键值对,对键值对进行准确定义,充分利用不同部分间的 Shuffle 过程为 MapReduce 技术实现的关键。

Map 部分主要用于读入数据,得到设备故障分组与专家知识库分组,获取设备故障、专家知识库组合分组的 ID,同时将其当成键值对中的 key^[6,7]。Value 值就是故障号 ID 与专家知识 ID 以及故障设备对专家知识的相关度进行评分。Shuffle 过程即将分布于不同 Map 节点的键值对数据依据某种规则分组,划分至对应 Reduce 节点。

Reduce 部分主要用于得到优化问题的解,输出键值对专家知识 ID 或故障 ID 特征向量^[8]。

Map 端处理相对较为容易,主要是对数据进行映射划分,对不完整数据进行过滤。Reduce 端主要用于完成故障更新以及特征向量的更新。Reduce 端需采集最新故障、专家知识特征向量,可在 Reduce 端执行 reduce 函数实现装载,将输出的特征向量当成下一次迭代的输入,直至符合收敛条件,对结果进行保存。

数据挖掘工具选用怀卡托智能分析环境,其为将数据挖掘算法作为应用的开源平台,存在大规模机器学习算法,可实现对数据的处理,也可完成新交互式界面的可视化处理。

把开源怀卡托智能分析环境和建立的 Hadoop 大数据处理平台结合在一起,为设备画像推荐算法提供依据。

1.2 大数据聚类分析

聚类分析为数据挖掘的一种主要算法,聚类的目的为在无先验知识的情况下,按照数据相似性把大数据划分成不同类,令同类元素最大程度类似,不同类元素间差异尽可能大,所以又被称作非监督聚类^[10]。

当前聚类分析法主要有基于划分的聚类分析法、基于密度的聚类分析法以及基于网络的聚类分析法等。K-means 算法可简单高效地处理大数据,有效改善可扩展性问题,在工业与科学领域中有很大的影响,本节选用 K-means 算法。

在大数据环境下,将 K-means 算法和改进协同过滤算法结合在一起,实现设备画像推荐的基本思想如下:首先任意确定 k 个聚类中心,通过相似度求出某条专家知识和各聚类中心的相似性,同时把专家知识划分至和聚类中心最

相似的簇中,直至全部专家知识均被划分完成,产生 k 个簇,所有簇中专家知识间相似度均很高。完成聚类后,求出故障所需寻找的专家知识和各聚类中心间的相似程度,在相似度最高的簇中为故障找到最近邻集合,向故障做出合理推荐。

也就是针对簇集合 $A = \{A_1, A_2, \dots, A_k\}$, 对下式求最小值:

式中, μ_i 用于描述分类 A_i 的均值。

$$\operatorname{argmin} \sum_{i=1}^k \sum_{x_j \in A_i} x_j = \mu_i^2 \quad (1)$$

详细流程为:

- 1) 从 n 个对象中随机选择 k 个对象当成初始聚类中心。
- 2) 按照各聚类对象的均值,求出所有对象和中心对象的差异性,把元素划分至差异度最低的簇中。
- 3) 重新对各存在变化的聚类的均值进行计算。
- 4) 重复进行步骤 2) 与步骤 3), 直至各聚类均不再出现改变。
- 5) 输出结果。

1.3 最近邻集合选择

在对最近邻选择的过程中,本节通过多种相似性衡量因子加权的思想对相似性进行计算^[11],选用的相似性衡量因子主要有评分矩阵相似性和设备画像特征相似度,则本节相似性计算公式为:

$$\operatorname{sim}(u, v) = \alpha \operatorname{sim}_r(u, v) + \beta \operatorname{sim}_e(u, v) \quad (2)$$

式中, $\operatorname{sim}_r(u, v)$ 用于描述故障 u 与其他故障 v 间评分相似性, $\operatorname{sim}_e(u, v)$ 用于描述故障 u 与其他故障 v 间设备画像相似度, α, β 依次用于描述相应权重。

得到相似度后,将相似度最大的前若干结果作为最近邻。

下面对设备画像相似度进行计算。设备画像为对设备特征的多维描述,是通过若干表示设备特征的标签及设备故障组成的集合,例如设备的故障表现形式、故障影响程度等。

如图 1 所示,集合故障 u 与其他故障 v 中的元素依次代表故障 u 与其他故障 v 的画像特征,相交部分代表故障 u 与其他故障 v 存在的共同的画像特征。因此设备画像相似度可通过设备画像特征文本集合相似性求出。

针对设备画像信任度,假设存在 m 种设备画像标签,故障 u 与其他故障 v 画像特征文本均包含 m 个标签,依次用 Q_u 与 Q_v 进行描述。故障 u 与其他故障 v 的画像相似度也就是 Q_u 与 Q_v 的相似度,计算公式如下:

$$\operatorname{sim}_e(u, v) = \frac{|Q_u \cap Q_v|}{|Q_u \cup Q_v|} \quad (3)$$

式中, $Q_u \cap Q_v$ 用于描述故障 u 与其他故障 v 共同设备画像特征元数数量, $Q_u \cup Q_v$ 用于描述故障 u 与其他故障 v 设备画像含有的全部特征数量。

评分相似性通过皮尔逊相关系数进行描述。皮尔逊相

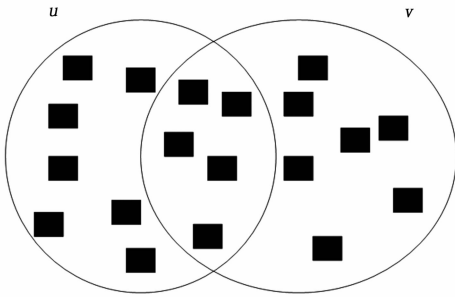


图 1 设备画像特征图

关系数在 $[-1, 1]$ 范围内取值, 皮尔逊相关系数绝对值越高, 则相关性越强。在皮尔逊相关系数为 1 的情况下, 认为相似性最高, 呈正相关关系; 在皮尔逊相关系数为 -1 的情况下, 认为呈完全负相关关系。在皮尔逊相关系数为 0 时, 认为无任何关系。

故障 u 与其他故障 v 间评分相似性 $sim_r(u, v)$ 利用皮尔逊相关系数度量如下:

$$sim_r(u, v) = \frac{\sum_{c \in C} (r_{c,u} - \bar{r}_u)(r_{c,v} - \bar{r}_v)}{\sqrt{\sum_{c \in C} (r_{c,u} - \bar{r}_u)^2} \sqrt{\sum_{c \in C} (r_{c,v} - \bar{r}_v)^2}} \quad (4)$$

式中, C 用于描述对故障 u 与其他故障 v 均有评分的故障集合, $r_{c,u}$ 用于描述对 u 的评分, $r_{c,v}$ 用于描述对 v 的评分, \bar{r}_u 用于描述对 u 的评分平均值, \bar{r}_v 用于描述对 v 的评分平均值。

本节依据协同过滤算法的特点, 对评分相似度进行优化, 防止在整个对象空间中查找最近邻。

针对皮尔逊相关系数公式, 因为 $r_{c,u} - \bar{r}_i$ 与 $r_{c,v} - \bar{r}_j$ 不一定是非负的, 所以有:

$$(r_{c,u} - \bar{r}_u)(r_{c,v} - \bar{r}_v) \leq 4[\max(\bar{r}_u, R - \bar{r}_u, \bar{r}_v, R - \bar{r}_v)] \quad (5)$$

其中, R 用于描述评分最大值。

分析上述关系可知, 在对皮尔逊相关系数进行求解时, 完成第 h 个专家知识的计算后, 相似度至少是 $sim_h(u, v)$, 则剩下的未计算相似度的部分符合下式:

$$\max sim_r(u, v) \leq sim_r(u, v) + 4(C - h)[\max(\bar{r}_u, R - \bar{r}_u, \bar{r}_v, R - \bar{r}_v)] \quad (6)$$

1.4 设备画像推荐

采用协同过滤算法向目标故障推荐相关专家知识时, 主要依赖最近邻, 选择前若干相似度最大的当成最近邻。目标故障按照最近邻用户对专家知识的兴趣度评分, 得到传统评分预测结果, 公式描述如下:

$$e_{ui} = \bar{r}_u + \frac{\sum (r_{c,u} - \bar{r}_u) sim(u, v)}{\sum sim(u, v)} \quad (7)$$

式中, e_{ui} 用于描述故障对某专家知识中描述的故障匹配程度预测评分。

在进行任务推荐时, 数据规模逐渐增加, 造成协同过滤评分矩阵规模越来越大。在大规模评分矩阵中为故障选

择最近邻需耗费大量的时间, 这主要是由于需对大规模专家知识或故障间的相似性进行计算, 有很大的计算量。同时在添加新故障或新专家知识后, 又需重新搜寻最近邻, 无法及时有效地体现目标故障的偏好, 大大降低了实时性, 也会对推荐算法的推荐准确性产生影响, 造成推荐质量大大降低。

为了避免出现上述分析问题, 首先通过奇异值分解算法把 $a \times b$ 维评分矩阵 R 通过各行故障评分均值对各行缺少值进行填充^[12], 获取填充矩阵 R_g , 对其进行分解, 获取三个低秩矩阵 W 、 D 、 Z , 同时令其相乘, 则有:

$$R_g = W_g \times D_g \times Z_g \quad (8)$$

其中: $g \ll \min\{a \times b\}$ 。

求出 $W_g D_g^+$ 与 $Z_g D_g^+$ 两个矩阵, 通过下式对故障 u 对专家知识 i 中描述的故障匹配程度预测评分进行改进, 同时对矩阵进行填充:

$$e_{ui} = \bar{R}_u + W_g D_g^+(u) + Z_g D_g^+(i) \quad (9)$$

利用聚类把相似度高的故障聚集在一起, 形成簇, 从目标故障所处簇中选择最近邻, 为目标故障推荐专家知识, 有效改善了扩展性问题。

推荐算法步骤如下:

输入: 评分矩阵, 聚类中心。

输出: 设备画像推荐结果。

1) 任意选择 k 个聚类中心, 求出故障和聚类中心的相似度, 把故障划分成 k 个簇。

2) 求出目标故障和聚类中心间的相似度, 获取和目标故障相似度最高的簇。

3) 在目标故障所属簇中, 求出目标故障和其余故障间的相似性, 选择前几个相似度最高的故障当成其最近邻。

4) 通过式 (9) 对目标故障对某专家知识中描述的故障匹配程度进行预测。

5) 向目标故障推荐前若干专家知识规则。

2 实验结果及分析

2.1 仿真实验环境设置

实验在 Apache Mahout 开源平台进行, 该平台可建立机器学习算法, 包含聚类算法、协同过滤算法的应用。通过平台提供的工具构造推荐引擎。

为了验证本文提出算法的性能, 实验环境尽可能贴近真实网络, 实验过程如下:

1) 通过随机技术产生设备属性, 得到相应取值, 实验数据源于我们历时两年从工业现场采集到的共 269 条风箱故障数据。

2) 得到训练集, 训练集中包含设备一个月内的历史数据, 测试集中设备为训练集中的活跃设备, 可保证数据的真实性。

3) 完成对故障属性与设备信息的统计, 获取设备间互交换信息的分布情况, 交换信息的统计结果为实验提供

数据。

4) 通过统计获取的数据与评价标准对本文算法的有效性进行验证。

由上述分析可知, 实验数据选用的是工业现场采集的实际数据, 最大程度地模拟了实际情况。虽然真实应用场景更加复杂, 然而上述实验设置基本体现了工业系统实体的主要状态, 基本满足实际情况。

实验通过随机函数形成的设备故障属性值对故障属性相似程度进行计算。构造故障属性集合 $\{u_{nom}, u_{nom}, user\}$, 包括数值型属性、名称型属性与故障属性。

2.2 权重值确定

针对相似度计算中的权重 α 与 β , 二者符合 $\alpha + \beta = 1$ 。按照实际数据和线性规划中的最小二乘拟合法, 对 α 与 β 值进行调整, 从而获取最优结果。

下面针对 $\alpha = 0.65, \beta = 0.35; \alpha = 0.55, \beta = 0.45; \alpha = 0.35, \beta = 0.55$ 三种不同情况进行实验, 比较三种情况下覆盖率、成功率、召回率, 确定权重取值。

覆盖率计算公式为:

$$B = \frac{n(M)}{n(N)} \tag{10}$$

式中, 集合 N 用于描述故障集合, $n(N)$ 用于描述故障量, M 用于描述 N 中接收推荐的集合, $B = \frac{n(M)}{n(N)}$ 用于描述相应数量。

成功率为故障成功匹配数量占总量的比重。

召回率即候选集中成功匹配数量占实验过程中实际接收匹配量的比重。

权重分析结果用表 1 进行描述。

表 1 权重分析结果

权重值	召回率/%	成功率/%	覆盖率/%
$\alpha=0.65, \beta=0.35$	86.21	91.26	89.89
$\alpha=0.55, \beta=0.45$	89.35	75.95	72.32
$\alpha=0.35, \beta=0.65$	61.25	53.15	49.65

分析表 1 可知, 在 α 增加的情况下, 成功率逐渐增加, 在 α 为 0.65 的情况下, 成功率达到最高, 但召回率较 α 为 0.55 时低, 覆盖率也随 α 的增加而增加, 在 α 为 0.65 时最高, 所以, 取 $\alpha = 0.65, \beta = 0.35$ 。

2.3 推荐结果主观评价

为了对本文算法推荐结果进行主观评价, 实验利用衡量相关故障和目标知识库规则对本文算法的推荐有效性进行测试。

对 80 个目标故障集合进行训练, 分别采用本文算法、主体模型算法和概率推荐算法推荐两条专家知识规则, 和直接按照设备故障号查找规则相比, 以其中一个目标故障为例, 给出其推荐结果, 比较三种算法给出的解决方案可知, 本文算法推荐结果和目标故障描述匹配结果最为相符,

与主题模型算法和概率推荐算法相比, 更加符合目标故障的研究兴趣, 验证了本文推荐算法的有效性。

为了更加准确地验证本文算法的有效性, 在上述实验的基础上, 进行数值分析验证。

推荐准确性为评价推荐算法的可靠指标, 本节将平均绝对误差 (Mean Absolute Wrror, MAE)、均方根误差 (Root Mean Squared Error, RMSE)、Perplexity 指标作为衡量指标。

平均绝对误差计算公式如下:

$$MAE = \frac{\sum_{(u,v) \in T} |r_{uv} - e_{uv}|}{|T|} \tag{11}$$

均方根误差计算公式如下:

$$RMSE = \frac{\sqrt{\sum_{(u,v) \in T} |r_{uv} - e_{uv}|^2}}{|T|} \tag{12}$$

平均绝对误差和均方根误差值越低, 认为推荐算法推荐结果越准确。

P 值属于信息理论测量指标, 是似然估计值的体现, 用于衡量推荐结果和故障所需结果的相关性, 其值越高, 认为推荐结果越可靠, 公式描述如下:

$$P(u) = \exp\left\{\frac{\sum \log p(\omega)}{\sum N_u}\right\} \tag{13}$$

式中, u 用于描述测试集故障, $p(\omega)$ 用于描述推荐给故障的专家知识规则, N_u 用于描述设备故障集合总量。

P 值越高, 认为推荐结果和设备所需结果越一致。

在上述实验的基础上, 比较本文算法、主题模型算法和概率推荐算法的平均绝对误差、均方根误差和 P 值, 比较结果用表 2 进行描述。

表 2 三种算法设备画像推荐准确性比较结果

算法	平均绝对误差	均方根误差	P 值
本文算法	0.702 5	0.852 6	92.15
主题模型算法	0.823 6	0.893 1	76.32
概率推荐算法	0.831 5	0.916 5	69.12

分析表 2 可以看出, 采用本文算法进行设备画像推荐, 平均绝对误差为 0.7025, 均方根误差为 0.8526, 低于主题模型算法的 0.8236 与 0.8931, 也低于概率推荐算法的 0.8315 和 0.9165。说明本文推荐算法推荐结果准确性明显由于主题模型算法和概率推荐算法。本文算法 Perplexity 值为 92.15, 比主题模型算法的 76.32 和概率推荐算法的 69.12 明显更高, 说明本文算法推荐的专家规则和目标故障所需要的策略更加相似, 说明本文算法推荐结果和故障需要结果更加相符, 进一步验证了本文算法的有效性。

3 算法在烧结机控制中的应用

该算法也可以用在烧结工况的控制中, 通过文中的算法, 将工况不同模式聚在一起, 按最接近的专家知识策

略推荐给当前的工况模式, 使得工况在该模式下的运行达到预期的控制效果。

不妨设烧结机的控制模型为:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.1445 & 0.1430 & 0.4231 \end{bmatrix} \begin{bmatrix} x_{1c}(k) \\ x_{2c}(k) \\ x_{3c}(k) \end{bmatrix} + \begin{bmatrix} 0.0591 \\ 0.0103 \\ -0.0041 \end{bmatrix} u(k) \quad (10)$$

首先对状态空间进行胞划分, 即对 $[-5 \ 5] \times [-5 \ 5] \times [-5 \ 5]$ 均匀划分为 $10 \times 10 \times 10 = 1000$ 个胞。每一维对应的胞中心为图 2。

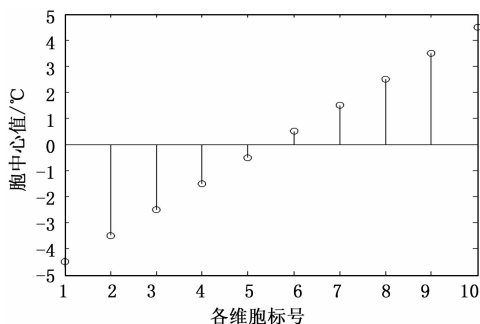


图 2 划分形成的聚类中心

利用前面给出的有关故障的专家知识, 即这 10 个专家知识规则给出的控制策略, 最终得到系统的控制效果如图 3 所示。图 4 展示了三种推荐算法的比较结果。

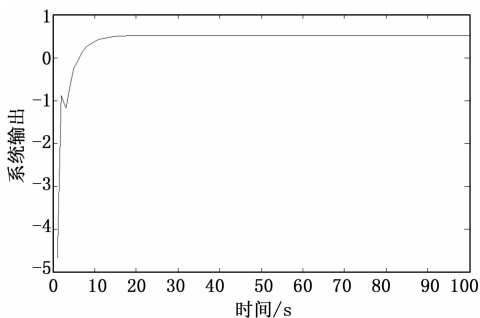


图 3 在专家给出的控制策略下得到的系统输出

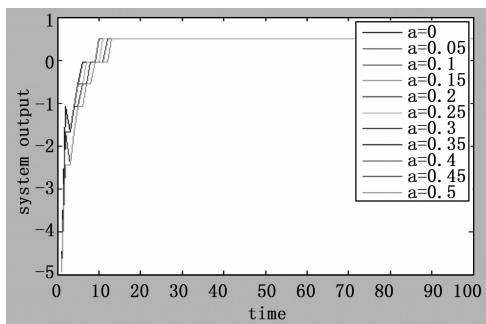


图 4 三种算法设备画像推荐准确性比较

因此达到了期望的控制效果, 验证了该方法的有效性。

4 结束语

本文提出一种基于大数据技术的任务画像推荐算法。将大规模多领域专家学者资源作为推荐资源, 通过 Hadoop 建立大数据信息存储和处理平台, 通过开源 MapReduce 技术完成对大规模数据的计算与处理。把开源怀卡托智能分析环境和建立的 Hadoop 大数据处理平台结合在一起, 为设备画像推荐算法提供依据。

任意确定 k 个聚类中心, 通过相似度求出某用户和各聚类中心的相似性, 同时把用户划分至和聚类中心最相似的簇中, 直至全部用户均被划分完成, 产生 k 个簇。

完成聚类后, 在相似度最高的簇中为用户找到最近邻集合, 依据经奇异值分解处理后的填充评分矩阵, 向用户推荐合理的专家学者。

实验结果表明, 所提算法主观推荐结果可靠, 将平均绝对误差、均方根误差和 P 值作为指标进行衡量, 发现所提算法推荐精度高。在机器学习和专家知识库的不断更新和迭代中, 相信工业控制系统中的故障诊断系统会更加完善。

参考文献:

- [1] 工业控制系统场景指纹及异常检测 [J]. 清华大学学报 (自然科学版), 2016, 56 (1): 14-21.
- [2] 李佳伟, 郝悍勇, 李宁辉. 工业控制系统信息安全防护 [J]. 中国电力, 2015, 48 (10): 139-143.
- [3] 蓝冬梅. 大数据量图书下多数据集的二部图多样化推荐 [J]. 情报理论与实践, 2016, 39 (2): 69-72.
- [4] 田贤忠, 沈杰. 大数据环境下基于概率矩阵分解的个性化推荐 [J]. 计算机科学, 2017, 44 (s1): 438-441.
- [5] 董爱美, 王士同, 蒋亦樟, 等. 基于最小包含球的异质空间大数据集快速相似度学习算法 [J]. 控制与决策, 2014, 29 (9): 1553-1561.
- [6] 吴博文, 陈曦. 一种改进 Minhash 的分布式协同过滤推荐算法 [J]. 计算机工程与应用, 2016, 52 (12): 95-100.
- [7] 高升, 任思婷, 郭军. 基于潜在因子模型的跨领域信息推荐算法 [J]. 电信科学, 2015, 31 (7): 75-79.
- [8] 陆坤, 谢玲, 李明楚. 一种融合隐式信任的协同过滤推荐算法 [J]. 小型微型计算机系统, 2016, 37 (2): 241-245.
- [9] 王永贵, 宋真真, 肖成龙. 基于改进聚类和矩阵分解的协同过滤推荐算法 [J]. 计算机应用, 2018, 38 (4): 1001-1006.
- [10] 李改, 陈强, 李磊. 基于评分预测与排序预测的协同过滤推荐算法 [J]. 电子学报, 2017, 45 (12): 3070-3075.
- [11] 贾忠涛, 吴颖川, 刘志勤. 一种协同过滤算法在网络干扰过滤中的应用 [J]. 计算机仿真, 2016, 33 (1): 284-287.
- [12] 黄璐, 林川杰, 何军, 等. 融合主题模型和协同过滤的多样化移动应用推荐 [J]. 软件学报, 2017, 28 (3): 708-720.