

基于 cox_stuart 趋势检验算法的 MVR 系统升级

马苏上¹ 刘明剑² 包彦宇³ 郭富旺⁴ 张贺⁴ 贾向东⁴ 陈苏超³ 袁思佳⁵

1. 大连海洋大学信息工程学院; 2. 大连海洋大学设施渔业教育部重点实验室; 3. 华润三九医药股份有限公司;
4. 华润三九(枣庄)药业有限公司; 5. 西南科技大学生命科学与工程学院

摘要: 针对中药提取流程使用的机械蒸汽再压缩装备 (MVR) 能耗优化环节, 本文研究并提出了一种对于该装备控制系统的优化设计模型。其主要思路首先是将系统中的实时数据链接数据库, 在数据处理系统进行数据清洗, 完成初步的筛选。其次利用可视化分析对数据进行处理, 实现更直观的数据展示。最后根据数据的变化情况权衡产品生产、OEE (设备综合效率)、能耗节约之间的关系, 设计出合适的清洁时间, 投料次数的优化算法, 最终提高中药提取浓缩液的产出效率, 降低成本。通过实验测试, 该系统的升级在保证安全的前提下, 提取液的生产效率提升, 证明了该设计的正确性和有效性。

关键词: 中药提取流程; MVR; cox_stuart 趋势检验算法; MVR 设备清洁方案

现阶段, 中药生产主要包括中药提取、药液浓缩成膏、制剂生产等步骤。其中, 药液浓缩是中药生产的一个关键步骤, 其工业化生产过程会消耗大量的新鲜蒸汽, 这也是中药加工过程中的主要耗能单元^[1], 其浓缩方法的优化和改进对于提高中药质量和生产效率具有重要意义。传统的中药蒸发浓缩工艺存在一些问题如能耗高、热能损耗高、操作复杂, 高温度还容易影响热敏性较高药品中有效成分的性质^[2]等。为了满足现阶段生产对于高效、节能的需求, MVR (机械蒸汽再压缩) 技术现如今逐渐被制药企业广泛应用。

MVR 是一项节能技术装备, 它通过机械压缩做功, 将蒸汽能量从低品位提升至高品位, 使压缩机只消耗少量电能就能获得更大的供热量, 从而大幅减少对外界能源的需求。这种实现了持续蒸发状态的工艺过程, 使 MVR 广泛应用于制造业。在医药领域, 梁林等^[3]对 MVR 技术应用于维生素 C 溶液蒸发

的实验特性进行了研究分析。结果表明, MVR 工艺过程可以实现低温蒸发, 节能效果明显, 经济效益显著。此外, 连晓岚^[4]译自国外的一篇期刊指出 MVR 应用于酒精蒸发中所需的能耗要远低于市面其他已知的蒸发技术。随着学者对 MVR 技术的深入研究以及制造业的现代化进程, MVR 已经成为蒸发领域的主流应用技术。

MVR 技术中蒸汽压缩机的工作介质是水蒸汽, 这一特点带来了重要的问题, 即与水直接接触的零部件需要防腐防锈保护^[5]。除此之外, 在蒸发过程中, 残留的杂质会附着在换热管道壁上, 导致换热效率降低, 能耗增加, 因此对于设备仪器进行及时的清洁维护是必要的。

但是在中药生产行业中, 由于 MVR 技术的广泛应用产生了大量的回顾数据, 导致了数据丰富而信息匮乏的局面^[6]。因此, 如何对 MVR 过程进行数据建模, 用于优化其运行策略, 以实

现中药提取过程的智能化和精细化控制, 仍是当今制药装备智能化升级的重点研究方向。

本研究旨在通过深入分析中药提取 MVR 过程的特点, 结合数据挖掘和算法设计等技术^[7], 建立准确可靠的 MVR 数据模型, 并提出相应的设备控制系统优化策略。这将为中药提取过程中 MVR 设备清洁方案的优化提供理论支持和实践指导。

1 提取环节改进实现

1.1 前期调研及归因分析

通过在工厂提取车间进行调研, 结合具体的数据与车间工人的实操经验, 绘制影响该装备能耗的鱼骨图, 如图 1 所示。

根据图 1 并结合实际调研可以得出以下信息:

(1) 外部环境影响: 气候影响冷却设备的使用, 当室外温度过高时, 冷却水系统会增加耗电量;



图1 能耗鱼骨图

(2) 员工操作影响: 在实际的数据统计中, 为人工记录数据, 容易造成误差。同时 MVR 设备操作繁琐, 不同员工操作导致设备能源消耗不同;

(3) 药液品种影响: 不同的药液品种密度不同、工艺参数不同, 因此 MVR 设备的生产周期、清洁方式、耗电量亦不相同;

(4) 在 MVR 界面: 人为记录数据主要包括用电量 (主要能耗), 冷凝流量 (蒸发量), 电流 (电流量为 0 则代表生产结束), 以及根据原始数据计算获得的平均电流, 蒸发耗电量, 小时蒸发量, 小时耗电量为主要的;

(5) 生产情况: 药液的生产周期长, 同时过程繁琐。随着生产批次增加, 换热效率降低, 导致设备蒸发效率降低, 进而导致生产时间延长, 耗电量提高。

1.2 数据预处理与可视化分析

为处理海量 MVR 数据, 本研究设计以下处理方法。首先对装备控制系统中的原始数据进行清洗, 确保最大精确的保留重要数据。其次, 根据每一批次生产结束后仪器电流会暂时归 0 的特性, 将数据划分为不同的批次, 并得出每一批次内的数据变化。之后对重要的影响数据进行分析, 根据数据变化找出合适的仪器清洁时间。

1.2.1 数据清洗

将原始数据中的乱码筛出, 之后根据周期的开始时间与结束时间进行数

据处理, 同时进行可视化的处理。

通过调研得知, 在实际生产中, 电量是主要的能耗来源, 因此将电流与电量的影响占比设为 MAX, 其余 MVR 变量的影响占比计算方式为:

$$P_{No} = \sum_{k=0}^n \frac{Q_{No.time(k)} - Q_{No.time(k+1)}}{Q_{22.time(k)} - Q_{22.time(k+1)}} \quad (1)$$

$$P_V = \frac{P_{No}}{n} \times 100\% \quad (2)$$

式中: P_{No} 代表 MVR 变量数据的影响占比, n 代表 MVR 变量数据的总量, $Q_{No.time(n)}$ 代表 MVR 变量数据在第 k 位置的数值, P_V 代表平均影响占比。

本研究规定当 $P_V < 0.01\%$ 时计为弱影响不计入 P_V , 当 $0.01\% < P_V < 10\%$, $P_V > 10\%$ 时分别记为普通影响, 强影响并计入 P_V 。

在进行数据清洗时, 保留表 1 中的强影响量, 以简化后续操作并使结果更精准。

1.2.2 数据处理

在实际生产中, 工厂数据的记录形式大多通过手动抄录设备的启动时间与结束时间的数据变化, 完成数据的统计。

为判断数据之间的联系, 本研究在数据处理中加入时间维度, 利用 mysql 数据库以及操作工具 SQLyog, 根据生产药液的起始时间划分批次, 批序号记为 i , 获取处理这一批次的生产耗费的时间 t , 后续通过计算获取该批次平均耗电量 $A_{i,e}$, 平均蒸发量 $A_{i,w}$, 其公式

表 1 MVR 数据提取中变量影响

No.	MVR 变量数据	影响分类	PV
1	进料累计	强影响	43.26%
2	冷凝水累计	强影响	38.64%
3	出料累计	弱影响	—
4	频率	弱影响	—
5	电流	强影响	MAX
6	进料温度	普通影响	3.18%
7	蒸发器温度	普通影响	2.12%
8	分离器温度	弱影响	—
9	出口温度	弱影响	—
10	循环管温度	弱影响	—
11	循环水温度	弱影响	—
12	冷凝水液位	普通影响	0.17%
13	分离器液位	弱影响	—
14	排水液位	弱影响	—
15	油箱温度	弱影响	—
16	油路温度	弱影响	—
17	高速前轴温	弱影响	—
18	高速后轴温	弱影响	—
19	低速前轴温	弱影响	—
20	低速后轴温	弱影响	—
21	密度	普通影响	4.17%
22	电量	强影响	MAX

如下:

$$t = T_{i,b} - T_{i,o} \quad (3)$$

$$A_{i,e} = \frac{E_{i,T} - E_{i,O}}{t} \quad (4)$$

$$A_{i,w} = \frac{W_{i,T} - W_{i,O}}{t} \quad (5)$$

式中: $T_{i,b}$ 表示 i 批次的开始时间, $T_{i,o}$ 表示 i 批次的结束时间。 $E_{i,T}$ 表示开始时间总电量。

为直观反映耗电量的变化情况, 本研究根据 $A_{i,e}$ 以及 $A_{i,w}$ 设计蒸发耗电量 $E_{i,n}$ 。

$$A_{i,n} = \frac{E_{i,T} - E_{i,O}}{W_{i,T} - W_{i,O}} \quad (6)$$

式中: $E_{i,O}$ 表示结束时间总电流。 $W_{i,T}$ 表示开始时间蒸发量, $W_{i,O}$ 表示结束时间蒸发量。

利用数据分析工具处理后其部分数据表现形式如表 2 所示。

1.2.3 可视化分析

特征数据分析:

为便利本研究后续的分析, 需确立

能直观表示设备作业状态变化的变量，本研究取原始数据中 2023 年 1 月 1 日至 2023 年 10 月 1 日原始数据清洗与数据处理，选择 5 段不同周期的数据进行数据可视化，图 2 为变量蒸发耗电量的周期变化对比，图 3 为小时蒸发量的周期变化。

(1) 情况说明：随着仪器的生产时间增加 / 生产批次增加，仪器小时蒸发量会逐渐降低，而蒸发耗电量会逐渐增加；

(2) 量化评估：由于蒸发量数据变化幅度 θ 过小，往往一个周期内变化 $0.5 < \theta < 2$ ，所以多数周期内所展示图形较为平滑，而蒸发耗电量数据变化幅度处于 $5 < \theta < 15$ ，数据变化明显，更能直观展示 MVR 设备的作业状态；

(3) 处理方法：针对于不同周期的蒸发耗电量变化构造散点图，结合实际情况进行分析以获取不同情况下的处理方式。

实际处理分析：

在设备的实际生产中，其正常增长趋势如图 4 所示，异常增长趋势如图

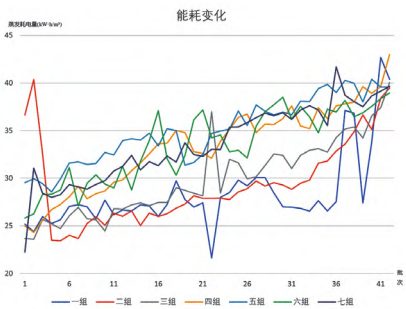


图 2 吨蒸发耗电量变化

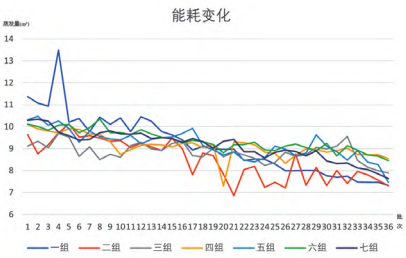


图 3 小时蒸发量变化

5 所示。

(1) 情况说明：在一轮生产后会进行相应清洁，但换热设备多轮次生产后会降低换热效率，会造成后续的生产蒸发耗电量增加，从而导致图 5（1）的情况，当设备进行周期清洁后，若长时间未使用，会导致设备开机时前几轮

生产蒸发耗电量增加，即出现图 5（2）的情况；

(2) 量化评估：设备生产时出现简单的生产问题对于总体生产影响极小，但是数据的波动会导致算法处理出现问题，本研究将针对异常进行算法设计；

(3) 处理方法：cox_stuart 趋势检测可以更直观的展示数据的变化趋势，同时本研究将对于异常情况进行处理，以完善算法设计。

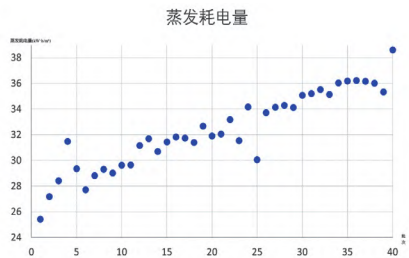
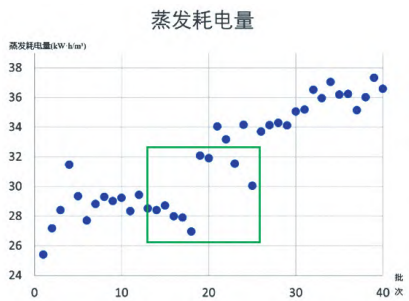
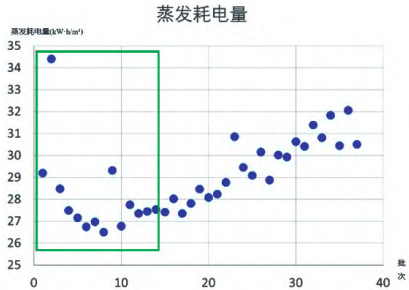


图 4 正常增长变化



(1)



(2)

图 5 数据异常图

表 2 整合数据表

批次	时长 (h)	用电量 (kW·h)	进液量 (m³)	冷凝水量 (m³)	蒸发耗电 (kW·h/m³)	小时耗电量 (kW·h)	小时蒸发量 (m³/h)
1	5.67	1406	35.8	35.9	39.16	247.97	6.33
2	6.50	1555	39.7	40.2	38.68	239.23	6.18
3	6.00	1451	37.2	37.5	38.69	241.83	6.25
4	5.83	1413	35.2	35.6	39.69	242.37	6.11
5	6.33	1453	38	38	38.24	229.54	6.16
6	6.17	1458	37.9	37.6	38.78	236.3	6.09
7	6.33	1493	38.3	38.3	38.98	235.86	6.05
8	6.17	1432	37.2	37.1	38.6	232.09	6.01
9	6.33	1450	36.7	37.3	38.87	229.07	5.89
10	6.67	1601	40.5	40.5	39.53	240.03	6.07
11	6.33	1498	38.2	37.5	39.95	236.65	5.92
12	6.67	1547	38.4	38.7	39.97	231.93	5.8
13	6.67	1586	38.6	39	40.67	237.78	5.85
14	6.83	1587	39.5	39.3	40.38	232.36	5.75
15	6.83	1565	38.8	38.2	40.97	229.14	5.59
16	7.17	1681	40.4	41.1	40.9	234.45	5.73
17	7.00	1597	39.6	38.7	41.27	228.14	5.53
18	6.67	1543	37.6	38.8	39.77	231.33	5.82

1.3 算法设计

1.3.1 调研分析

实际生产中，工厂的设备清洁时间主要由设备生产的批数、设备生产的时间两大指标确定。通过调研发现，在工作者确认清洁时间时，只需对设备生产表的数据变化进行简单的确认。然而数据的变化会最直观反应设备状态，如

果忽视对于数据的处理,会造成大量重要信息的浪费。

本研究的重点就是对于数据的深度处理,增添新的清洁指标。由 1.2.3 可以得知平均蒸发量与蒸发耗电量可以直观地展示仪器的状态。根据吨蒸发耗电量的表示,可以得出,设备的蒸发耗电量会在下降时存在一定的起伏,针对此行为,本研究根据原始的 cox_stuart 趋势检测,通过增添控制要求与时间输出,设计了一种算法针对处理,以获取最准确的清洁时间。同时为确保安全有效,如果工厂规定的清洁指标先达成,立刻进行仪器的清洁。

1.3.2 cox_stuart 趋势检验算法改进

原始的 cox_stuart 趋势检验算法直接考虑数据的变化趋势,若数据有上升趋势,那么排在后面的数据的值要比排在前面的数据的值显著增大;反之,若数据有下降趋势,那么排在后面的数据的值要比排在前面的数据的值显著的小。利用前后两个时期不同数据的差值正负来判断数据总的变化趋势。该算法不依赖趋势结构,因此展现的趋势变化更加观,也更适合判断最佳清洁时间。

算法介绍:

可以把每一个观察值和后面的另一个观察值配对作比较,即独立观测时间序列 $x_1, x_2, \dots, x_n$ 合理选择数据间距 c , 获得数对 $(x_1, x_{1+c}) (x_2, x_{2+c}) \dots$, 本研究规定间距 c 为 MVR 数据个数的一半,其具体公式如下:

$$c = \begin{cases} \frac{n}{2} & n \text{ 为偶数} \\ \frac{n+1}{2} & n \text{ 为奇数} \end{cases} \quad (7)$$

数对个数:

$$n' = \begin{cases} c & n \text{ 为偶数} \\ c-1 & n \text{ 为奇数} \end{cases} \quad (8)$$

对得到的数对,看增长的对子与减少的多少来判断总的趋势,令计算的每一对两个元素的差,记为:

$$D_i = \begin{cases} x_i - x_{i+c} > 0 & x_i > x_{i+c} \\ 0 & x_i = x_{i+c} \\ x_i - x_{i+c} < 0 & x_i < x_{i+c} \end{cases} \quad (9)$$

根据 D_i 正负情况判断该数据组是否具有增长的趋势。

算法改进:

为了选取合适的清洁时间,本研究对原始算法中的 D_i 数列进行处理,规定其中正数个数为 S^+ , 负数个数为 S^- , 其公式如下:

$$\begin{cases} S^+ = \sum_{D_i > 0} (\text{sign} D_i) \sim B(n', \frac{1}{2}) \\ S^- = \sum_{D_i < 0} (-\text{sign} D_i) \sim B(n', \frac{1}{2}) \end{cases} \quad (10)$$

MVR 设备作业周期内,通常按照处理批次达到 40 次开始对设备进行处理,另外当设备即将达到清理时间时,其数据波动较为明显,具体清洁时间也是由此区间进行判断。因此本研究规定算法的数据基数 μ 为批次指标的 90%, 其具体公式如下:

$$\mu = 40 \times 90\% = 36 \quad (11)$$

当数据个数小于 μ 时,该数据不具备普适性,因此将 S^+ 与 μ 作比。当对原始数据首次进行 cox_stuart 趋势检验算法后,若数据呈现的趋势符合清洁时的变化趋势,则应对其算法中所得 S^+ 进行处理。

其具体公式如下:

$$\eta = \begin{cases} \frac{S^+}{36} \times 100\% & N_{\text{总}} < 36 \\ \frac{S^+}{N_{\text{总}}} \times 100\% & N_{\text{总}} > 36 \end{cases} \quad (12)$$

当 η 的值大于 85% 时,可以确定设备到达情节时刻,当 η 的值小于 85% 时,进行第二次处理。

在实际工作中可能由于外部环境,人工操作失误,设备长时间未开机等行为影响数据的上升趋势并不明显,但是在设备的处理周期中其最大值数据与最

小值数据差值较为稳定,当数据在此周期内本研究规定前 n 个数据中的最小值为 $x_{\min}$, 后 n 个数据中的最大值为 $x_{\max}$, 所取阈值 λ 公式如下:

$$\lambda = x_{\max} - x_{\min} \quad (13)$$

其流程图如图 6 所示。

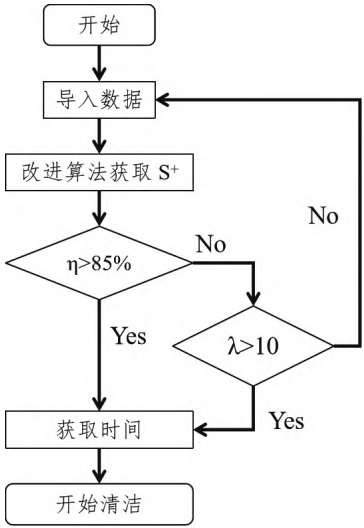


图 6 算法流程图

2 实验验证

2.1 实验设计

取一台 MVR 设备在 2023 年上半年的生产数据(该设备主要生产药品 1), 为判断算法可以将异常情况减少, 并使生产耗电稳定, 本研究根据其周期的平均蒸发耗电量, 进行数据可视化, 以便于后续分析。其具体内容如图 7 所示。

(1) 情况说明: 图 7 中上部分数据多半由于周期过长导致蒸发耗电量较高, 本研究规定为高耗生产, 下部分则是生产周期过短, 产生数据较少, 同时设备运行处于健康状态, 蒸发耗电量较低但是产量较低。本研究规定为低量生产;

(2) 量化评估: 药物生产周期过长虽然确保周期结束时药物产量高, 但是其中的能耗极大, 不利于设备长期健康生产。而生产周期过短则导致药物产量较低, 设备增加使用次数, 若要使生产药物达到规定产量, 能耗同样会增加;

(3) 处理方式: 本研究将对 2023

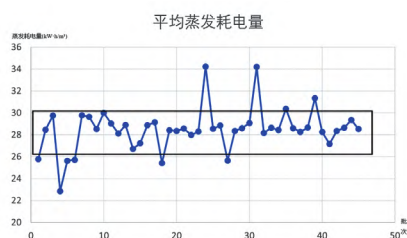


图 7 周期对比图

年上半年的数据进行处理，假设上半年的数据处于已应用改进算法的理想状态，将改进算法处理高耗生产，低量生产，健康生产所获取的数据与未处理数据作对比，进行科学的分析。

2.2 实验对比

使用算法的设备生产耗能变化以及清洁时间变化如图 8 所示。

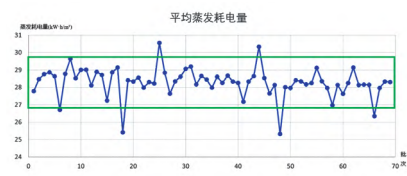


图 8 改进算法对比图

(1) 情况说明：通过图 8 与图 7 进行对比，除去设备面对突发情况可能导致能耗中途升高或者周期过短。导致能耗过高或过低。而通过图 9 可以观察到代表原始生产批次的黄线药略高于代表改进生产的红线，同时由于改进批次后每一批次的药物生产有所下降，因此改进生产会多生产 2~3 个周期；

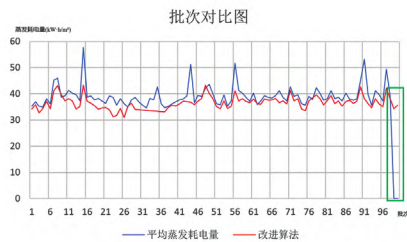


图 9 批次对比图

(2) 量化评估：可以观察到利用算法处理的数据更为稳定，能很好地把控能耗；

(3) 验证方式：为了明显展示数据变化，根据其耗电量，蒸发量，蒸发耗电量等数据，制作如下数据表 3。

表 3 数据对比图

数据对比	原有清洁方式	改进清洁方式	变化率
小时耗电量 (kW·h)	253.040	244.007	-3.71%
小时蒸发量 (m³)	9.43	9.54	1.16%
批次	72	75	4.1%
生产 (m³/L)	0.001475	0.001507	2.1%
蒸发耗电量 (kW·h/m³)	29.24	28.45	-2.77%

根据上述图表可以观察到，改进的清洁方式虽然导致药物生产周期增加，但是在节能与生产都要由于原始清洁算法，为了进一步验证算法的合理性，本研究将对 2024 年上半年的某一台生产药物 1 的 MVR 设备应用算法处理 MVR 设备清洁后的数据进行可视化

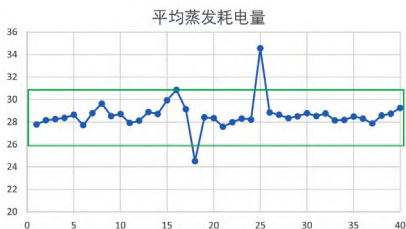


图 10 实时数据处理

处理，其具体数据展示如图 10 所示。

(1) 情况说明：由于算法的改进使得数据波动广分范在一个稳定的区间；

(2) 量化评估：可以观察到利用算法处理的数据更为稳定中药集中于 26-30 的范围之内；

(3) 结果说明：通过对于实时数据的监督以及设备的处理，避免了低效生产，在确保了药物稳定生产的前提下，很好地把控能耗。

3 结论

基于中药提取 MVR 数据分析进行生产过程的优化，是当前企业绿色可持续发展领域的研究热点。通过建立数据处理系统，完成对数据的清洗、可视化与分析，可以实现对中药提取过程的精准控制和优化。本研究通过对于算法的改进，实现了一种判断合适清洁时间的策略，在一定程度上提高生产效率，降低生产的能耗。

未来，主要研究方向将侧重于技术

融合与装备的智能化升级，将单机装备控制系统与生产车间 DCS 集成在一起，整体优化生产效率、提高产品质量、降低生产成本，来进一步提高中药提取浓缩流程的智能化水平。

参考文献

[1] 王如猛. 基于 PCA-BP 的煤层底板突水预测研究与防治 [J]. 华北科技学院报, 2017,14(4):29 ~ 33

[2] 李舒艺, 伍振峰, 岳鹏飞, 等. 中药提取液浓缩工艺和设备现状及问题分析 [J]. 世界科学技术: 中医药现代化, 2016, 18(10):6. DOI: 10.11842/wst.2016.10.019.

[3] 连晓岚译. 机械蒸汽再压缩蒸发器蒸发酒精蒸馏废液的八年经验 [M]. 福建糖业, 1992:61-67.

[4] 梁林, 韩东. 蒸汽机械再压缩蒸发器的实验 [J]. 化工进展, 2009, 28 (增刊):358-360.

[5] 杨俊玲; 杨鲁伟; 张振涛; MVR 热泵节能技术的研究进展 [J]; 风机技术; 2016 年 04 期

[6] 陈胜, 陈厚柳, 瞿海斌, 中药产品质量回顾数据挖掘方法与研究 [J]. 中国中药杂志. [http://doi/ 10.19540/j.cnki.cjcmm.20221128.301](http://doi/10.19540/j.cnki.cjcmm.20221128.301)

[7] Jiawei Han, Micheline Kanber. Data mining: Concepts and techniques [M]. San Francisco: Morgan Kaufman Publishers Inc, 2001