

8-7-2020

Mufti-model Soft Modeling Based on Clustering in Evaporation Process

Peng Chen

1. She Wai School, Central South University of Forestry and Technology, Changsha 410004, China;;

Xiaoshan Qian

2. Physical Science and Technology College, Yichun University, Yichun 336000, China;;3. School of Information Science & Engineering, Central South University, Changsha 410083, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research](#), [Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Mufti-model Soft Modeling Based on Clustering in Evaporation Process

Abstract

Abstract: Based on the idea approach based on multi-model with a stronger robustness and precision, combining with strong coupling and the highly nonlinear characteristics in multi-effect evaporation process, *a soft sensor modeling method of fuzzy kernel clustering was proposed based on differential evolution and multi-squares support vector machines. The method made clustering samples under different conditions by use of DE-KFCM, each sub-models were constructed using LSSVM using fuzzy clustering algorithm to cluster the input data under different conditions, and finally get the system output by weighting.* Taking alumina evaporation process as an example, for the soft sensor modeling of export of feed concentration, the better results have been obtained.

Keywords

DE algorithm, least squares support vector machine, evaporation process, KFCM

Recommended Citation

Peng Chen, Qian Xiaoshan. Mufti-model Soft Modeling Based on Clustering in Evaporation Process[J]. Journal of System Simulation, 2015, 27(9): 2050-2055.

基于聚类的多模型蒸发过程软测量建模

彭琛¹, 钱晓山^{2,3}(1.中南林业科技大学涉外学院, 湖南 长沙 410004; 2.宜春学院物理科学与工程技术学院, 江西 宜春 336000;
3.中南大学信息科学与工程学院, 湖南 长沙 410083)

摘要: 基于多模型方法具有更强的鲁棒性和精度的思想, 结合多效蒸发过程强耦合和高度非线性特征, 提出一种基于差分进化的模糊核聚类的多最小二乘支持向量机的软测量建模方法。该方法依据运用 DE-KFCM(差分进化核模糊聚类)对不同工况下的数据样本进行聚类, 得到的子集分别用 LSSVM(最小二乘支持向量机)构建模型不同工况使用模糊核聚类算法对输入数据进行聚类划分, 针对每个聚类子集用最小二乘支持向量机方法建立子模型, 通过加权得到系统输出。以氧化铝蒸发过程为例, 对出口料液浓度的进行软测量建模, 获得了较好的实验结果。

关键词: 差分进化算法; 最小二乘支持向量机; 蒸发过程; 模糊核聚类

中图分类号: TP18

文献标识码: A

文章编号: 1004-731X (2015) 09-2050-06

Mufti-model Soft Modeling Based on Clustering in Evaporation Process

Peng Chen¹, Qian Xiaoshan^{2,3}(1. She Wai School, Central South University of Forestry and Technology, Changsha 410004, China;
2. Physical Science and Technology College, Yichun University, Yichun 336000, China;
3. School of Information Science & Engineering, Central South University, Changsha 410083, China)

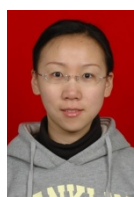
Abstract: Based on the idea approach based on multi-model with a stronger robustness and precision, combining with strong coupling and the highly nonlinear characteristics in multi-effect evaporation process, a soft sensor modeling method of fuzzy kernel clustering was proposed based on differential evolution and multi-squares support vector machines. The method made clustering samples under different conditions by use of DE-KFCM, each sub-models were constructed using LSSVM using fuzzy clustering algorithm to cluster the input data under different conditions, and finally get the system output by weighting. Taking alumina evaporation process as an example, for the soft sensor modeling of export of feed concentration, the better results have been obtained.

Keywords: DE algorithm; least squares support vector machine; evaporation process; KFCM

引言

早在 60 年代的时候, 有学者就提出来了多模

型建模思想, 对于一些比较复杂的系统, 在一定的条件下, 多模型方法的控制和辨识会具有更强的鲁棒性和较高精度^[1-2]。在冶金、造纸、电力、石油、化工等复杂工业过程中, 往往模型的结构或参数往往随着生产操作条件的变化而变化, 这一类模型可以视为多模型系统。在双酚 A 生产过程质量指标的软测量建模中, 文献[3]运用了局部重构融合流形聚类方法克服了异常样本点对模型的影响且验证了该方法有效; 在花生四烯酸(ARA)发酵过程软



收稿日期: 2015-04-08 修回日期: 2015-07-03;
基金项目: 国家自然科学基金项目(60634020, 60874069, 60804037); 国家 863 项目(2006AA04Z181); 宜春学院校级科研课题(XJ1314);
作者简介: 彭琛(1982-), 女, 湖南湘潭, 硕士, 讲师, 研究方向为智能信息处理; 钱晓山(1980-), 男, 江西九江, 博士, 讲师, 研究方向为复杂过程的建模、优化与控制。

<http://www.china-simulation.com>

• 2050 •

测量建模和芳烃异构化过程建模, 文献[4-5]运用反射传播聚类的多支持向量机模型进行模型的构建, 训练及测试结果表明回归精度和泛化能力都有所提高, 但难以克服异常样本点的影响; 由上可知, 上述方法构建的模型在应用于工业生产过程中虽然取得了较好的结果, 但是对于聚类算法中的对初始聚类中心敏感和依赖数据的分布形式的问题的解决效果不好且易使聚类目标函数陷入局部极值^[6]。基于此本文在核模糊聚类中引进差分进化算法, 克服了目标函数陷入局部最优的问题, 然后对每个聚类子集构建 LSSVM 模型并且由后验概率加权多个 LSSVM 模型得到最终的模型, 并将其应用于氧化铝多效蒸发过程, 取得了很好的效果。

1 基于差分进化的核模糊 C 均值聚类

熟知模糊聚类中的 2 个重要研究问题是聚类中心的初始化和聚类的有效性, 迄今为止, 很难找到一种通用的方法来解决这 2 个问题, 聚类数目的确定仍需聚类的有效性函数来完成。而判断聚类有效性函数又是模糊划分矩阵的函数, 所以在模糊聚类之后才能对其函数有效性的检验。

初始值的确定大多是采用多次初始值的随机赋值, 根据算法运算结果选择最优解, 运算量大、很难保证全局最优解且易陷入局部最优, 文献[7]在聚类原型先验知识中运用遗传算法对模糊聚类进行初始化, 虽然能避免陷入局部最优, 但是聚类数目的确定问题还是依靠事先给定, 通常情况下, 事先给定的聚类数目并不一定是最佳个数。文献[8]给出了聚类数目的经验公式: $C \leq \sqrt{N}$

文献[9]依据信息论观点虽然给出了该经验公式的合理解释, 而忽略了给定的聚类中心完全依赖于数据的数目——数据本身的几何特性这一弊端, 造成聚类数目随着 N 的增加而较快增长, 增加了计算复杂度。

以上分析可知, 核模糊聚类算法有效地避免了传统聚类算法对数据分布的依赖性, 然而实际应用中发现 KFCM 收敛速度慢且易陷入局部。基于此,

在核模糊聚类算法中引入差分进化算法(收敛速度快、全局搜索能力强等特点)能够在对 KFCM 的初始值进行优化的同时寻求更好的聚类效果, 本文利用 DEKFCM 在高维特征空间进行聚类, 提高了聚类的精确度, 降低了聚类时间。DEKFCM 的算法流程图如图 1 所示。

已知样本空间 $X=\{x_1, x_2, \dots, x_N\}$, x_i 为 M 维向量, 一个聚类中心由差分进化算法中的一个种群来代替, 即 $V=\{v_1, v_2, \dots, v_N\}$, v_j 和 x_i 维数相同。选择适应度函数如式(1):

$$f(x_i) = \frac{1}{J_\phi(U, V) + 1} = \frac{1}{2 \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m (1 - K(x_i, v_j)) + 1} \quad (1)$$

从式(1)可以看出适应度函数和聚类效果成正比。

DEKFCM 具体步骤如下:

- S1: 设定聚类数目 C , 误差 ε , $t=1$;
- S2: 初始化种群 V 、变异因子 F 、交叉因子 C_R 、DE 最大迭代次数 $G_{\max}$;
- S3: 核矩阵的计算;
- S4: 计算每个样本的 $U(t)$;
- S5: 进行变异、交叉、选择操作, 计算适应度函数值, 更新种群;
- S6: 若达到事先给定的最大迭代次数, 停止迭代, 得到最优聚类中心。否则转 S3, $t=t+1$;
- S7: 根据式(1)更新群体的隶属度和聚类中心
- S8: 计算相邻两代隶属度矩阵之差 E , 若小于 ε , 则停止, 否则转 S7。

2 基于 DEKFCM 的多 LSSVM 蒸发过程软测量

基于前面的研究基础, 本节将提出的基于 DEKFCM 的多 LSSVM 软测量建模方法, 应用于氧化铝多效蒸发过程, 旨在进一步提高软测量精度。

由于复杂工业过程的样本数据或多或少的存在噪声和数据缺失、孤立点(如蒸发过程的水洗和酸洗过程中的生产调度以及调整造成的数据丢失、异常), 这些样本数据如果不做处理直接用于支持向量机模型的训练, 必将会造成模型预测不准确, 主要是因为支持向量机对于训练集中数据噪音或者孤立点又过分敏感^[10]。而模糊核聚类可以解决

这一问题, 这样就减少了数据异常点对模型的影响。然后对每个子类构建 LSSVM 模型, 最后应用后验概率的平均加权得到模型的最终输出。软测量建模流程图如图 2 所示。

另外, 本文选择混沌粒子群算法优选 LSSVM 的惩罚系数 C 和核函数宽度 σ , 得到最优的参数集。

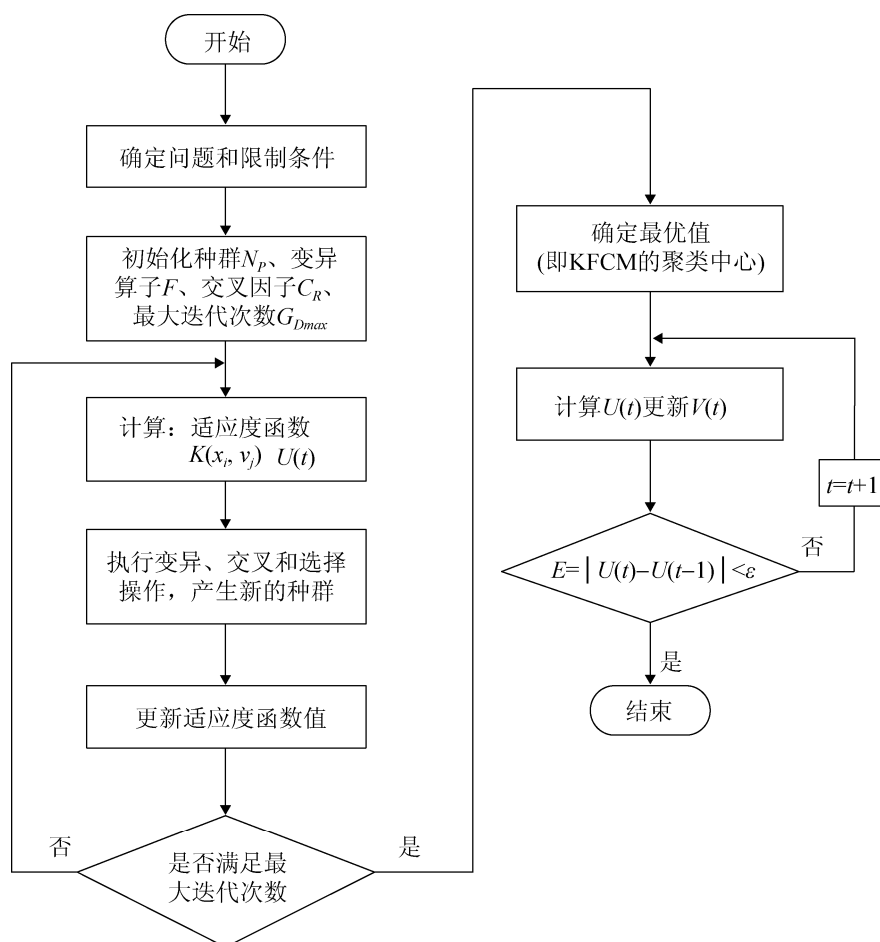


图 1 DEKFCM 算法流程图

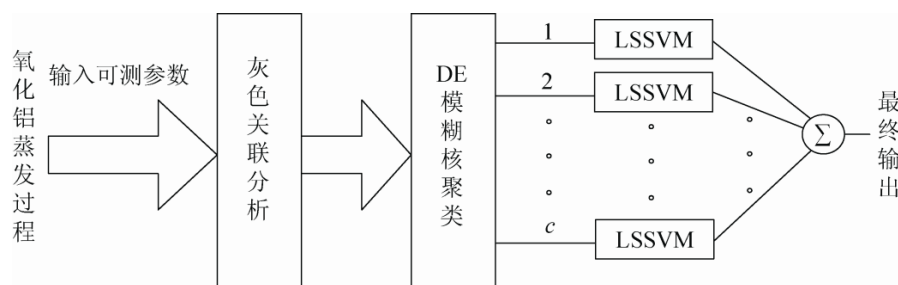


图 2 基于 DEKFCM-LSSVM 软测量建模过程

3 工业验证及分析

为了验证本文提出的建模方法的有效性,以河南某氧化铝厂蒸发过程为实际对象,根据采集的实时数据和数据库保存的历史数据,随机选取经过数据滤波和数据插值等处理后的 5 个月 360 组数据,其中的 300 组作为训练集,60 组作为测试集,用来验证并比较算法的性能。依据蒸发过程的数据分布特性,初步设定聚类个数为 4,差分进化算法的参数设置为:初始种群为 30,交叉因子 F 为 0.6,变异因子 C_R 为 0.3, ε 为 10^{-5} , $G_{\max}$ 为 1000。为了比较聚类算法的效果,分别采用 FCM、KFCM

和 DEKFCM 进行聚类,结果如表 1 所示。每一个样本集采用 LSSVM 进行训练,采用相对均方根误差(RMSE)作为评价指标,惩罚系数和核函数宽度参数对 (C, σ) 初始值为(30, 0.5)。运用混沌粒子群算法最终得到最优参数对 $(C, \sigma)=(42, 0.7)$ 。三种模型的仿真结果如图 2~3 所示。

表 1 聚类结果比较

算法	正确率	错误率	平均迭代次数	总计算时间
FCM	135	15	23	8.153 7
KFCM	143	8	67	16.056 8
DEKFCM	151	2	18	10.231 8

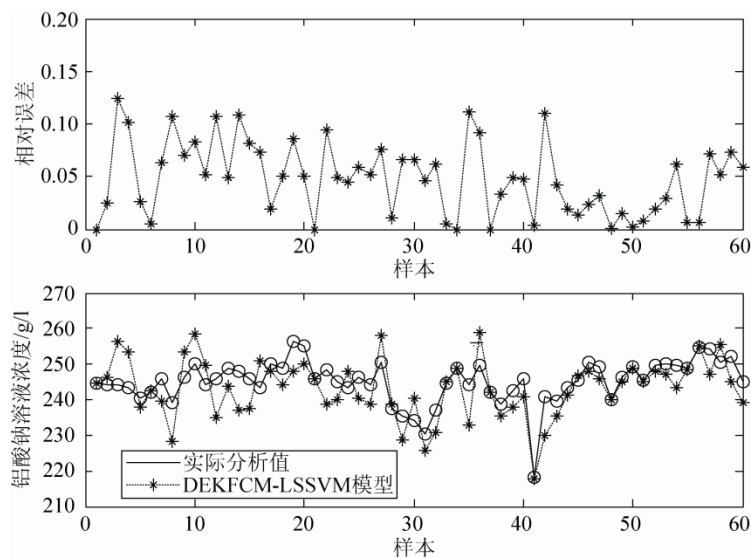


图 3 DEKFCM-LSSVM 仿真结果

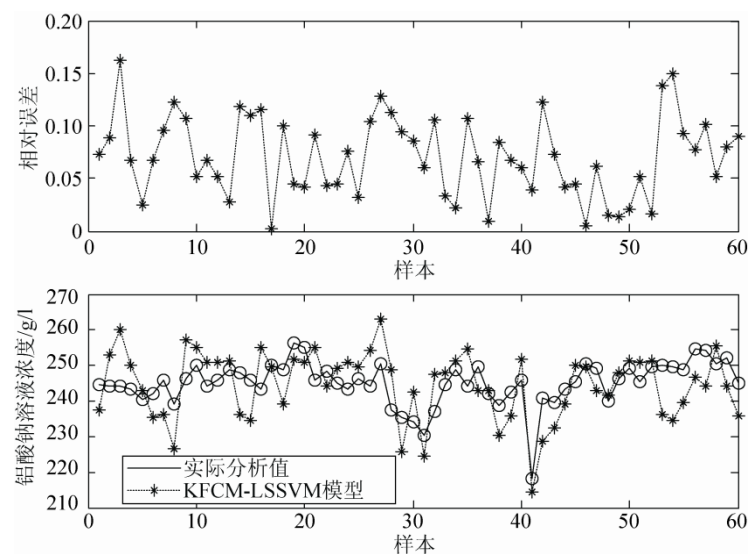


图 4 KFCM-LSSVM 的仿真结果

<http://www.china-simulation.com>

由表 1 看出, DEKFCM 聚类错误率最低, 而 KFCM 要循环计算核函数, 平均迭代时间最高。总的来说, DEKFCM 相比于 FCM 和 KFCM 聚类正确率最高, 且运行时间最短, 从而也验证了本文提出的方法有效且效果好。

综合图 3~5 和表 1~2 可以看出, GRA-KFCM-LSSVM 模型相比于 KFCM-LSSVM 软测量效果更好, 且最大相对误差仅为 12% 表明模型软测量精度高, 均方根差为 0.005 95 表明模型可信度高, 均方根相对误差为 1.6% 表明模型的跟踪性能较强。

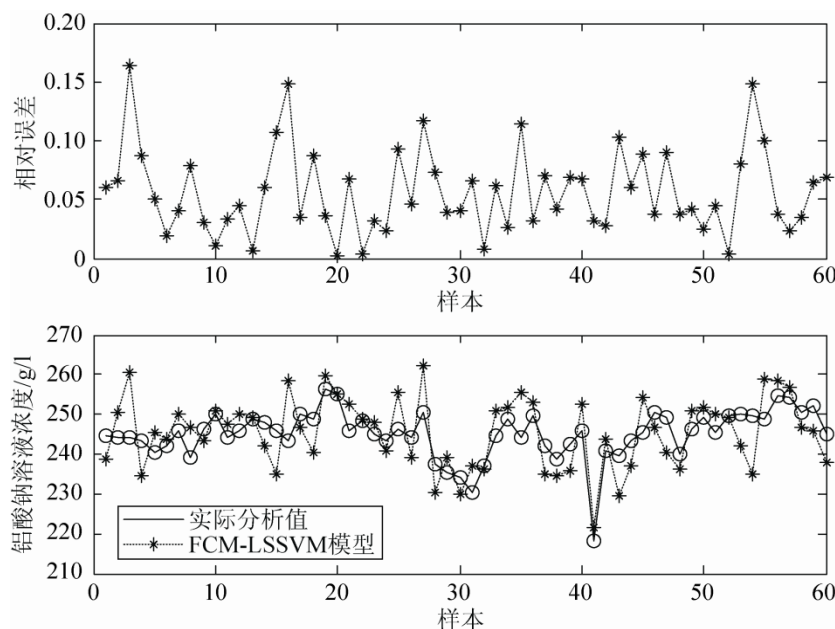


图 5 FCM-LSSVM 的仿真结果

另外, DEKFCM-LSSVM 模型的最大相对误差仅为 6.67%, 说明模型跟踪样本数据变化趋势的能力强且精度高, 自适应能力强。

表 2 误差结果分析

模型	Emax%	RMSE	RRMSE%
DEKFCM-LSSVM	12	0.00595	2.44
KFCM-LSSVM	16.5	0.00687	2.62
FCM-LSSVM	17.2	0.008494	9.21

4 结论

针对蒸发过程具有工艺机理复杂、生产流程长、影响因素多、大滞后等特点以及铝酸钠溶液浓度难以在线检测, 提出了一种基于灰色关联分析的 DE 核模糊聚类和最小二乘支持向量机相结合的蒸发过程多模型软测量方法。运用灰关联分析方法筛选出通过对经过灰关联筛选出输入样本集, 再通过 DE 核模糊聚类, 减少孤立点和缺失数据对支持向

量机模型训练的影响。依据生产数据进行训练和测试, 结果表明本文所提方法基于尽可能全面选择样本的前提, 建立的蒸发过程软测量模型精度高、自适应能力强, 完全可以应用于蒸发过程的实时操作优化, 也为其他类似复杂过程的软测量建模提供了很好的借鉴。

参考文献:

- [1] 李卫, 杨煜普, 王娜. 基于核模糊聚类的多模型 LSSVM 回归建模[J]. 控制与决策, 2008, 23(5): 560-564. (Li Wei, Yang Yupu, Wang Na. Multi-model LSSVM regression modeling based on kernel fuzzy clustering [J]. Control and Decision, 2008, 23(5): 560-564.)
- [2] 李修亮, 苏宏业, 褚健. 基于在线聚类的多模型软测量建模方法[J]. 化工学报, 2007, 58(11): 2834-2839. (Li Xiuliang, Su Hongye, Zhu Jian. Multiple models soft-sensing technique based on online clustering arithmetic [J]. Journal of Chemical Industry and Engineering, 2007, 58(11): 2834-2839.)

- [3] 陈定三, 杨慧中. 基于局部重构融合流形聚类的多模型软测量建模[J]. 化工学报, 2011, 62(8): 2281-2285. (Chen Dingsan, Yang Huizhong. Multiple model soft sensor based on local reconstruction and fusion manifold clustering [J]. Journal of Chemical Industry and Engineering, 2011, 62(8): 2281-2285.)
- [4] 李丽娟, 宋坤, 赵英凯. 基于仿射传播聚类的 ARA 发酵过程建模 [J]. 化工学报, 2011, 62(8): 2116-2122. (Li Lijuan, Song Kun, Zhao Yingkai. Modeling of ARA fermentation based on affinity propagation clustering [J]. Journal of Chemical Industry and Engineering, 2011, 62(8): 2116-2122.)
- [5] 李丽娟, 刘君. 芳烃异构化过程的多模型建模 [J]. 化工学报, 2011, 62(8): 2350-2354. (Li Lijuan, Liu Jun. Multi-modeling of aromatics isomerization process [J]. Journal of Chemical Industry and Engineering, 2011, 62(8): 2350-2354.)
- [6] 徐海霞, 刘国海, 周大为, 等. 基于改进核模糊聚类算法的软测量建模研究[J]. 仪器仪表学报, 2009, 30(10): 2226-2231. (Xu Haixia, Liu Guohai, Zhou Dawe, et al. Soft sensor modeling based on modified kernel fuzzy clustering algorithm [J]. Chinese Journal of Scientific Instrument, 2009, 30(10): 2226-2231.)
- [7] Boulet Benoit, Lalli Gino, Ajersch Mark. Modeling and control of an electric arc furnace [C]// Proceedings of the American Control Conference. USA: IEEE, 2003, 4: 3060-3064.
- [8] Goodfellow Howard D, Pozzi Marcello, Maiolo Joe. Dynamic process control and optimization for EAF steelmakers [J]. MPT Metallurgical Plant and Technology International (S0935-7254), 2006, 29(6): 24-30.
- [9] MacRosty R D, Swartz. C L. Dynamic optimization of electric arc furnace operation [J]. AIChE Journal (S0001-1541), 2007, 53(3): 640-653.
- [10] 李卫, 杨煜普, 王娜. 基于核模糊聚类的多模型 LSSVM 回归建模[J]. 控制与决策, 2008, 23(5): 560-562. (Li W, Yang Y P, Wang N. Multi-model LSSVM regression modeling based on kernel fuzzy clustering [J]. Control and Decision, 2008, 23(5): 560-562.)

《系统仿真学报》荣获“2014 中国最具国际影响力学术期刊”证书

由中国学术期刊(光盘版)电子杂志社与清华大学图书馆联合成立的中国学术文献国际评价研究中心, 以美国汤森路透 Web of Science 收录的 1.2 万余种期刊为引文统计源, 首次研制发布了 2012《中国学术期刊国际引证年报》(CAJ-IJCR 年报)。第一次给出了我国 5600 余种中外文学学术期刊总被引频次、影响因子、半衰期等各项国际引证指标, 并采用了新的国际影响力综合评价指标 CI 对期刊排序, 发布了“中国最具国际影响力学术期刊”(排序 TOP5%)和“中国国际影响力优秀学术期刊”(排序 TOP5-10%), 在国内外学术界产生了较大反响。之后, 2013 年版年报, 将引文统计源期刊扩展到 1.44 万多种。目前, 2014 版国际、国内年报与 TOP5%和 TOP5-10%期刊的遴选业已完成,《系统仿真学报》被列入“2014 中国最具国际影响力学术期刊”行列。

我学报连续 2 年被列入 TOP5%国内一流的中国最具国际影响力学术期刊, 走向世界, 进入国际一流, 指日可待!