

曲线拟合的逆向云改进算法

王辉,秦术,刘少英,于立君,王科俊

(哈尔滨工程大学 自动化学院,黑龙江 哈尔滨 150001)

摘要:逆向云发生器是从定量数据到定性概念的转化模型。在逆向云生成算法中, E_x 的估计值对 E_n 和 H_e 的估计有比较大的影响,直接使用样本均值进行参数估计会导致较大的误差。本文通过深入研究云模型的相关算法,对逆向云生成算法进行改进。文中利用样本均值作为非线性拟合函数的初值,把拟合函数的返回参数作为 E_x 的估计值,改善了随机选取拟合初值造成结果不稳定现象,实现对逆向云参数优化。通过与其他几种逆向云生成算法进行对比,仿真结果表明:改进后的逆向云生成算法有较好的稳定性并且有较高的精度。

关键词:逆向云发生器;云模型;曲线拟合;参数优化;稳定性

中图分类号: TP183 **文献标志码:** A **文章编号:** 1673-4785(2014)05-0590-05

中文引用格式:王辉,秦术,刘少英,等. 曲线拟合的逆向云改进算法[J]. 智能系统学报, 2014, 9(5): 590-594.

英文引用格式:WANG Hui, QIN Shu, LIU Shaoying, et al. An improved algorithm of backward cloud based on curve fitting[J].

CAAI Transactions on Intelligent Systems, 2014, 9(5): 590-594.

An improved algorithm of backward cloud based on curve fitting

WANG Hui, QIN Shu, LIU Shaoying, YU Lijun, WANG Kejun

(Herbin Engineering University College of Automation, Harbin 150001, China)

Abstract: Backward cloud generator is a model, which transforms quantitative values into qualitative concepts. The estimated value of E_x has a great impact on E_n and H_e in backward cloud generated algorithm. The direct use of the sample mean will lead to large errors. This paper makes the more thorough research to the algorithm of cloud model. Furthermore, the backward cloud generated algorithm has been improved. The sample mean is used as the initial value of the nonlinear fitting function. Then, the fitting function returns parameter as estimated value of E_x . This can improve the instability of randomly selected the fitting initial value and achieve parameters optimization of the backward cloud algorithm. Compared with other several backward cloud generation algorithms, simulation result shows that the improved algorithm has better stability and higher accuracy.

Keywords: backward cloud generator; cloud model; curve fitting; parameters optimization; stabilit

云模型是我国学者李德毅教授在结合概率论和模糊数学理论的基础上,通过赋予样本点随机确定度来统一刻画概念中的随机性、模糊性及其相互关联性。^[1]它利用期望、熵、超熵3个数字特征描述一个定性概念,通过相应的算法形成用数字特征表示

的定性概念与其相对应的定量表示之间的不确定性转化模型,主要体现了这一概念的随机性与模糊性,并把二者完全集中在一起,构成定性概念(概念内涵)和定量数据(概念外延)相互转换,深刻揭示了客观现象具有的随机性和模糊性^[2]。

在云理论中,正向云发生器是由特定的算法实现概念的内涵向其外延的转换,就是由定性概念的数字特征生成一定数量的数据。由于逆向云生成算法是把一定数量的样本数据转换成数字特征表示的定性

收稿日期:2013-4-26.

基金项目:国家自然科学基金资助项目(61174047);中央高校基本科研业务费专项基金资助项目(HEUCF041404, HEUCF041406, HEUCF041315).

通信作者:于立君.E-mail:yulijun@hrbeu.edu.cn

概念,也就是一个定性概念数字特征的还原^[3-6]。在这一个过程中误差是不可避免的,因此算法的关键是构造一个相对准确的逆向云发生器。其中文献[7]中提出了基本的逆向云发生器,但这种算法的误差较大^[7]。罗自强等提出了曲线拟合的逆向云发生器,结果的精度有所提高,但是拟合初值的选取影响结果的稳定性^[8]。本文针对曲线拟合逆向云算法中参数不规律跳变、误差较大的现象,提出了一种优化设计方案,并通过仿真验证其有效性。

1 云模型理论

设 U 是一个用精确数值表示的定量论域, C 是 U 上的定性概念,若定量值 $x \in U$,且 x 是定性概念 C 的一次随机实现, x 对 C 的确定度 $\mu(x) \in [0,1]$ 是有稳定倾向的随机数,则 x 在论域 U 上的分布称为云,每一个 x 称为一个云滴。

云模型的数字特征:期望 E_x 、熵 E_n 和超熵 H_e ,这 3 个数字反映了定性概念的定量特征。

期望 E_x :最能够表示定性概念的某些点。与 E_x 距离越近,云滴集中性越好,反映人们对定性概念的认知比较统一;与 E_x 距离越远,云滴相对分散,反映人们对概念认知越不稳定,不统一。

熵 E_n :定性概念不确定性的度量。 E_n 表达定性概念的亦此亦彼性,反映了云滴在论域空间被接受的区域范围,也就是模糊度量; E_n 还反映云滴出现的随机性,代表云滴的离散度;同时熵也反映模糊性和随机性之间关联性。一般来说,熵的取值越大,表示概念越宏观,模糊性和随机性也就越大,对概念的确定性表达就比较困难。

超熵 H_e :定性概念熵的不确定性表示,即熵的熵。由熵的模糊性和随机性共同决定。

由以上定义可知,论域中的值表示的定性概念的确定度不是恒定不变的,而是在细小变化着的。但这种细小的变化并不影响云模型的整体特征,对云来说,重点在于研究云的整体形状反映出的不确定性概念的特征,以及云滴大量出现时确定度值呈现的规律性^[9-10]。

云模型中最重要的算法是正向云发生器和逆向云发生器。由数字特征 $C(E_x, E_n, H_e)$ 生成特定定量数据的算法,称为正向云发生器 (forward cloud generator),用 FCG 表示,如图 1。云模型既不是一条清晰的隶属度曲线,也不是一个确定的概率密度函数。而云模型是由两次串接的正态发生器生成的许多云滴所组成,不仅是一对多的泛正态数学映射图像,而且是一朵可以伸缩、无确定边缘的云图,并具有定量和定性之间的转换的功能。

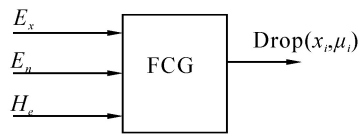


图 1 正向云发生器
Fig.1 Forward cloud generator

正向云发生器的算法描述如下。

输入:定性概念 $A \sim$ 的数字特征 E_x, E_n, H_e 以及云滴个数 N 。

输出: N 个云滴的定量数值和云滴的确定度。

算法步骤:

- 1) 以 E_n 为期望, H_e 为标准差生成一正态随机数 E_n' ;
- 2) 以期望为 E_x , 标准差为 E_n' 生成一正态随机数 x ;
- 3) 计算: $y = e^{\frac{-(x-E_x)^2}{2(E_n')^2}}$;
- 4) x 就称为一个云滴,表达定性概念 $A \sim$ 的一次具体量化; y 称作 x 属于定性概念 $A \sim$ 的确定度;
- 5) 重复 1)~4), 直至生成 N 个云滴。

逆向云发生器将一定量的样本转换为用数字特征表示的定性概念,也就是一个定性概念数字特征的还原过程,在这个过程中不可避免地会有误差产生,因此人类认知过程的关键在于构造一个相对准确的逆向云发生器^[11-13]。

2 基于曲线拟合的逆向云改进算法

逆向云发生器 (backward cloud generator, BCG) 是从定量到定性的映射,它将一定数量的精确数值有效转换为以数字特征 $C(E_x, E_n, H_e)$ 表示的定性概念,实现定量数值与定性概念之间不确定性的转换和定性概念的外延到内涵转换的过程。逆向云发生器如图 2 所示。

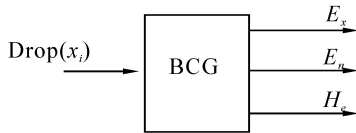


图 2 逆向云发生器
Fig.2 Backward cloud generator

依据统计原理,如果给定的样本点越多,则逆向云算法得到的参数估计误差越小。无论采用什么算法,在样本点数量有限的情况下,误差不可避免。

下面讨论正态云 (X, Y) 的期望特性曲线。

当 x_i 是一维时,正态云 (X, Y) 是二维随机变量,其联合概率密度函数为

$$f_{X,Y}(x,y) = f_Y(y)f_X(x|Y=y) =$$

$$\begin{cases} \frac{1}{2\pi H_e \ln y} e^{\frac{(x-E_x - \sqrt{-2\ln y} E_n)^2}{4H_e^2 \ln y}} (0 < y \leq 1, E_x \leq x < +\infty) \\ \frac{1}{2\pi H_e \ln y} e^{\frac{(x-E_x + \sqrt{-2\ln y} E_n)^2}{4H_e^2 \ln y}} (0 < y \leq 1, -\infty \leq x < E_x) \end{cases}$$

由 $y = e^{\frac{-(x-E_x)^2}{2(E_n')^2}}$ 知,对任意的 $0 < y \leq 1, x = E_x \pm \sqrt{-2\ln y} E_n'$ 。由于 E_n' 是随机变量,因此 X 是对称分布于 E_x 的两边的随机变量,可以只对 $x = E_x + \sqrt{-2\ln y} E_n'$ 进行分析, $x = E_x - \sqrt{-2\ln y} E_n'$ 的讨论完全类似。

由 $E_n' \sim N(E_n, H_e^2)$ 知 x 服从正态分布,期望为 $x = E_x + \sqrt{-2\ln y} E_n'$,标准差为 $B = \sqrt{D(X)} = \sqrt{-2\ln y} H_e$ 。由此知,云滴的离散程度和 H_e 成正比,和 y 成反比,即 H_e 越大,离散程度越大; y 越小(对云的位置来说是越靠近山脚),云滴越分散。由 $x = E_x + \sqrt{-2\ln y} E_n'$ 解出 $y = e^{\frac{-(x-E_x)^2}{2(E_n')^2}}$,这样得到正态云的期望曲线方程: $y = e^{\frac{-(x-E_x)^2}{2(E_n')^2}}$ 。

正态云期望曲线如图3所示,其几何形状具有明显的特点,反映了正态云模型的理想曲线。期望曲线用来表示数据集合在空间分布的统计规律。期望曲线是一条光滑、连续的曲线,刻画了云模型的整体特性,是云滴的总体轮廓。 H_e 反映了所有云滴围绕期望曲线做随机波动的程度。这里的总体轮廓不是几何意义下的中间部分,而是概率意义下的期望。

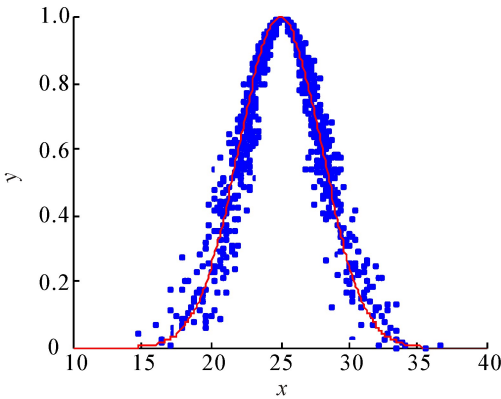


图3 正态云模型的期望曲线

Fig.3 Expectation curve of normal cloud model

图3中离散点表示云滴,光滑曲线表示的是云模型的期望曲线。

在基本逆向云生成算法中, N 个云滴的定量值 $x_1, x_2, \dots, x_i (i = 0 \sim N)$ 作为输入,输出为定性概念 $A \sim$ 的期望,熵以及超熵的估计值 $\hat{E}_x, \hat{E}_n, \hat{H}_e$ 。首先计算这些数据的均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ 以及方差 $S^2 = \frac{1}{N-1} \sum_{i=1}^N |x_i - \bar{X}|^2$;进而计算出3个参数值 $\hat{E}_x =$

$$\bar{X}, \hat{E}_n = \sqrt{\frac{\pi}{2}} \times \frac{1}{N} \sum_{i=1}^N |x_i - \bar{X}|^2 \text{ 以及 } \hat{H}_e = \sqrt{S^2 - \hat{E}_n^2}。$$

基本逆向云算法会产生较大的误差。基于曲线拟合的逆向云生成算法如图4所示。从流程图可知曲线拟合的初始值在选取的时候是随机选取的,这样会造成结果的不稳定。针对这一问题,考虑优化拟合参数的初值对算法进行改进^[14]。

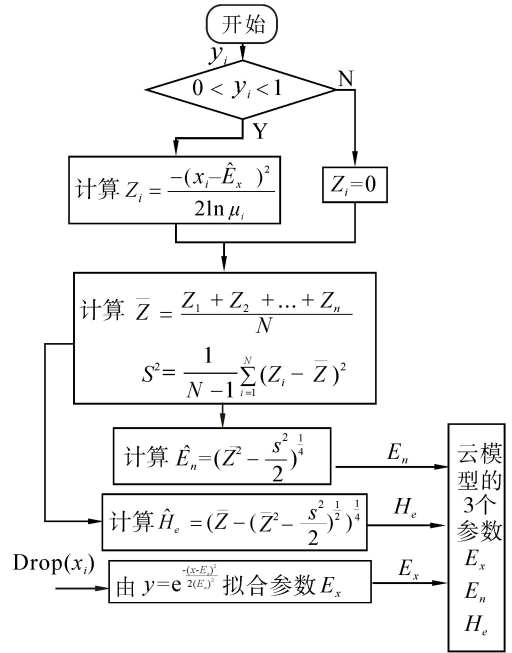


图4 曲线拟合逆向云算法流程图

Fig.4 The flowchart of curve fitting backward cloud algorithm

根据正态云模型的期望曲线编写 M 函数文件“ni”。

M—file:

function F=ni(x,xdata)

F=exp(-(xdata-x(1)).^2/(2*(x(2).^2)))

利用非线性数据拟合函数 lsqcurvefit 进行曲线拟合。 $[x, \text{resnorm}] = \text{lsqcurvefit}(@ni, x0, xdata, ydata)$

其中:ni 为拟合的 M 函数文件名, x_0 为初始向量,xdata、ydata 为曲线拟合的实验数据。在实际应用中, x_0 的选取对实验结果的影响比较大,在此算法中把 xdata 的均值作为拟合的初始值进行拟合。函数返回值 $[x, \text{resnorm}]$ 为非线性函数“ni”的拟合系数,即期望和熵。改进后的算法输入为 N 个云滴的定量值 $x_1, x_2, \dots, x_i (i = 0 \sim N)$ 以及每个云滴的确定度,输出为定性概念 $A \sim$ 的期望 E_x 、熵 E_n 和超熵 H_e 的估计值 $\hat{E}_x, \hat{E}_n, \hat{H}_e$ ^[15]。

1) 根据 x_i 计算这组数据的样本均值 $\bar{X} =$

$$\frac{1}{n} \sum_{i=1}^n x_i ;$$

2)用 \bar{X} 作为拟合曲线方程 $y = e^{\frac{-(x-E_x)^2}{2(E_n)^2}}$ 的参数进行拟合,可得到 E_x 的估计值 \hat{E}_x ,和 E_n 的估计值 \hat{E}_n ;

3)若 $0 < \mu_i < 1$,计算 $Z_i = \frac{-(x_i - \hat{E}_x)^2}{2\ln\mu_i}$,若 $\mu_i = 1, Z_i = 0, i = 1, 2, \cdots, N$;

4)计算 $\bar{Z} = \frac{Z_1 + Z_2 + \cdots + Z_N}{N}$;

5)计算 H_e 的估计值 $\hat{H}_e = (\bar{Z} - \hat{E}_n^2)^{\frac{1}{4}}$ 。
MATLAB 实现关键代码如下:

```
x0=[ mean( xdata ),1];  
[ x ,resnorm] = lsqcurvefit( @ ni,x0,xdata,ydata)  
  
y=exp(-( xdata-x(1)).^2/(2*(x(2).^2)));  
E_x = x (1)  
E_n = x (2)
```

3 实验结果比较分析

将上述的3种逆向云生成算法利用 MATLAB 进行仿真。选取的3个参数为 $E_x = 25, E_n = 3, H_e = 0.1$,先由正向云生成算法生成云滴,云滴数由1 000开始,一次增加2 000,原曲线拟合逆向云生成算法中选取拟合初值 $x_0 = [1,1]$ 进行拟合,改进的曲线拟合逆向云生成算法中用 $x_0 = [\text{mean}(xdata),1]$ 作为拟合初值进行曲线拟合。实验仿真结果对比如表1~3所示。

表1 3种逆向云生成算法的 E_x 比较

Table 1 E_x comparison of three BCG algorithms

云滴数	基本逆向云	曲线拟合	改进曲线拟合
1 000	24.928 1	24.003 0	24.996 8
3 000	25.032 9	-73.572 8	25.000 3
5 000	24.988 1	24.998 9	25.001 2
7 000	25.007 8	-383.226 2	24.999 9
9 000	25.002 2	25.002 1	25.000 8
10 000	25.001 7	25.000 0	25.000 0

由表1可知随着云滴数的增加,期望 E_x 的值越来越接近实际值。由统计学理论知,样本点数给定的愈多,逆向云生成算法得到的参数估计误差就愈小。在样本点有限的情况下,无论采用什么算法,误差都是不可避免的。但是曲线拟合逆向云生成算法中期望 E_x 的值明显比基本的逆向云生成算法的期望值要精确,但出现了结果远远偏离实际值的不稳定现象。由此可知,在曲线拟合的过程中初值的选取是非常关键的,初值选取不当会造成较大的误差,

且结果很不稳定。而改进的算法中结果相对来说比较稳定,精度也比基本的逆向云生成算法高。基本逆向云算法中只利用 N 个样本值,而没有考虑样本发生的概率,以此参数的估计值误差相对于改进算法要大至少一个数量级。

表2 3种逆向云生成算法的 E_n 比较

Table 2 E_n comparison of three BCG algorithms

云滴数	基本逆向云	曲线拟合	改进曲线拟合
1 000	2.987 2	3.001 4	3.008 9
3 000	2.981 4	118.592 2	3.003 6
5 000	3.041 0	2.998 8	3.001 9
7 000	2.941 9	495.3 311	3.000 3
9 000	2.956 6	3.000 1	3.000 8
10 000	3.009 2	2.997 3	3.001 1

由表2知,这3种逆向云生成算法中,原逆向云生成算法中熵的结果相对比较稳定,但是精度不是很高。曲线拟合算法中也出现了超熵远远大于其他值的现象,结果很不稳定。在改进的逆向云生成算法中精度较高,每一次的实现结果相差不大,在千分之一范围内变化。可见期望值的精度很大程度上影响了熵的精度,因此得到一个相对精确的期望值对熵的计算值有较大的提高。

表3 3种逆向云生成算法的 H_e 比较

Table 3 H_e comparison of three BCG algorithms

云滴数	基本逆向云	曲线拟合	改进曲线拟合
1 000	0.543 1	0.098 6	0.101 8
3 000	0.275 9	648.592 0	0.102 6
5 000	0.206 8	0.100 8	0.100 9
7 000	0.157 1	2.920 7e+003	0.101 4
9 000	0.361 8	0.101 2	0.100 2
10 000	0.337 9	0.099 9	0.099 9

由表3可以看出,基本的逆向云生成算法中超熵 H_e 的值与实际值相比有比较大的出入,而且实验的波动性比较大。曲线拟合逆向云生成算法中的有些值比较接近实际值,但是不稳定,出现较大误差的现象,改进的曲线拟合逆向云生成算法的结果稳定性比较好,而且接近实际值,效果比较理想。

5 结束语

本文对曲线拟合逆向云生成算法的初值进行优化选取,用数据的均值作为拟合初值进行曲线拟合,分析比较了3种生成算法的期望 E_x 、熵 E_n 和超熵 H_e ,得出曲线拟合算法逆向云生成算法比基本的逆向云生成算法的精度要高得多,而改进的曲线拟合算法对超熵的不稳定现象得到极大的改善,为云模

型在定性评价以及数据挖掘等方面提供了更精确的理论依据。但是,逆向云生成的过程中不可避免地会有误差产生,关键在于构造一个相对准确的逆向云发生器,因此,在后续的研究中要利用更有效的方法生成更加精确的逆向云生成算法。

参考文献:

- [1] 李德毅,刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, 6(8): 28-34.
LI Deyi, LIU Changyu. Study on the universality of the normal cloud model[J]. Engineering Sciences, 2004, 6(8): 28-34.
- [2] 龙伟哲. 基于云理论的遥感云图分类研究[D]. 南宁:广西大学, 2009: 18-35.
LONG Weizhe. Remote sensing image cloud classification research based on cloud model theory[D]. Nanning: Guangxi University, 2009: 18-35.
- [3] 刘桂花,宋承祥,刘弘. 云发生器的软件实现[J]. 计算机应用研究, 2007, 24(1): 46-48.
LIU Guihua, SONG Chengxiang, LIU Hong. Software implementation of cloud generators[J]. Application of Computers, 2007, 24(1): 46-48.
- [4] 崔天宝. 基于云模型的短期电价预测的研究[D]. 保定:华北电力大学, 2008: 10-18.
CUI Tianbao. Short-term electricity price forecasting research based on cloud model[D]. Baoding: North China Electric Power University, 2008: 10-18.
- [5] 刘常昱,冯芒,戴晓军,等. 基于云信息的逆向云新算法[J]. 系统仿真学报, 2004(11): 2417-2420.
LIU Changyu, FENG Mang, DAI Xiaojun, et al. A new algorithm of backward cloud[J]. Journal of System Simulation, 2004(11): 2417-2420.
- [6] 张仕斌,许春香. 基于云模型的信任评估方法研究[J]. 计算机学报, 2013, 36(2): 422-431.
ZHANG Shibin, XU Chunxiang. Study of trust evaluation approach based on cloud model[J]. Journal of Computer, 2013, 36(2): 422-431.
- [7] 李德毅,孟海军,史雪梅,等. 隶属云和隶属云发生器[J]. 计算机研究和发展, 1995(13): 16-21.
LI Deyi, MENG Haijun, SHI Xuemei, et al. Membership cloud and membership cloud generator[J]. Journal of Computer Research and Development, 1995(13): 16-21.
- [8] 罗自强,张光卫. 一种新的逆向云算法[J]. 计算机科学与技术, 2007, 1(2): 234-240.
LUO Ziqiang, ZHANG Guangwei. A new algorithm of backward cloud[J]. Journal of Frontiers of Computer Science and Technology, 2007, 1(2): 234-240.
- [9] 吕辉军,王晔,李德毅. 逆向云在定性评价中的应用[J]. 计算机学报, 2003, 26(8): 1009-1014.
LYU Huijun, WANG Ye, LI Deyi. The application of backward cloud in qualitative evaluation[J]. Journal of Computers, 2003, 26(8): 1009-1014.
- [10] 刘常昱,李德毅,杜鹄. 正态云模型的统计分析[J]. 信息与控制, 2005, 34(2): 236-239.
LIU Changyu, LI Deyi, DU Yi, et al. Some statistical analysis of the normal cloud model[J]. Information and Control, 2005, 34(2): 236-239.
- [11] 王辉,于立君,王科俊,等. 一种可变模糊匹配阴性选择算法[J]. 智能系统学报, 2011, 6(2): 178-184.
WANG Hui, YU Lijun, WANG Kejun, et al. An adjustable fuzzy matching negative selection algorithm[J]. CAAI Transactions on Intelligent Systems, 2011, 6(2): 178-184.
- [12] 李众,刘艳. 一维正态云模型单规则推理映射研究[J]. 系统仿真学报, 2008, 20(3): 702-705.
LI Zhong, LIU Yan. Research on one-dimension normal cloud model single rule reasoning[J]. Journal of System Simulation, 2008, 20(3): 702-705.
- [13] 王辉,王科俊,于立君. 基于模糊思想生成最有效检测器集的变阈值免疫算法[J]. 系统仿真学报, 2008, 20(17): 4596-4600.
WANG Hui, WANG Kejun, YU Lijun. Adjustable threshold immune algorithm of creating most effective detector set based on fuzzy idea[J]. Journal of System Simulation, 2008, 20(17): 4596-4600.
- [14] 刘衍民,赵庆祯,邵增珍. 基于正态云的粒子群优化算法及其应用[J]. 计算机工程与应用, 2011, 37(17): 161-162, 166.
LIU Yanmin, ZHAO Qingzhen, SHAO Zengzhen. Particle swarm optimization algorithm based on normal cloud and its application[J]. Computer Engineering, 2011, 37(17): 161-162, 166.
- [15] 李德毅,杜鹄. 不确定性人工智能[M]. 北京:国防工业出版社, 2005: 271-284.

作者简介:



王辉,女,1976年生,副教授,博士,主要研究方向为智能控制、智能算法、图像处理、人工免疫。参加国家自然科学基金项目2项,发表学术论文30余篇,其中被EI检索10篇,编著教材2部。



秦术,男,1987年生,硕士研究生,主要研究方向为智能控制、人工免疫。



刘少英,女,1987年生,硕士研究生,主要研究方向为智能控制算法、船舶运动控制。