

基于支持向量机的室内舒适度评价方法

潘伟强, 李长云, 胡盛龙

(湖南工业大学 计算机与通信学院, 湖南 株洲 412007)

摘 要:针对室内环境因素多元化、动态变化的特点和目前评价方法的不足,建立了基于支持向量机的室内舒适度混合评判模型。首先将从真实环境中采集的数据集进行数据规范化处理;然后根据群体和个体感觉,分别用离线训练和在线训练的方法训练分类器;最后使用训练好的分类器预测样本的标签。以 Matlab 为开发工具,编写了基于支持向量机的室内舒适度评价算法,并与 BP 神经网络和概率神经网络等室内舒适度评价算法进行了比较,仿真结果表明,该方法是可行且有效的。

关键词:室内舒适度;支持向量机;BP 神经网络;概率神经网络

中图分类号:TP335

文献标识码:A

文章编号:1673-629X(2013)06-0214-05

doi:10.3969/j.issn.1673-629X.2013.06.055

Indoor Comfort Evaluation Method Based on Support Vector Machine

PAN Wei-qiang, LI Chang-yun, HU Sheng-long

(College of Computer and Communication, Hunan University of Technology, Zhuzhou 412007, China)

Abstract: For the diversity and dynamic changes of indoor environmental factors, and lack of the current evaluation methods, put forward an indoor comfort mixed-evaluation model based on support vector machine. Firstly, the method handled dataset collected from the real environment with the method of data standardization. Secondly, according to the group feeling and individual feeling, it used off-line and on-line to train the classifier. At last, it used the trained classifier to predict the label of the sample. The indoor comfort evaluation algorithm based on support vector machine was written with the developing tool Matlab, it compared with the BP neural network and probabilistic neural network (PNN). The simulation results of training and validation show that the method is feasible and effective.

Key words: indoor comfort; support vector machine; BP neural network; probabilistic neural network

0 引言

随着人们生活水平的提高,人们对室内环境的要求也不断提高,而评价室内舒适度是改善室内环境的前提条件。对室内环境舒适度的研究是当前研究的一个热点。

针对室内环境因素多元化、动态变化的特点,通过对室内环境的评价和预测进行探讨,根据所选指标的不同,目前已经提出了物元分析、熵权理论、数值预测、模糊评判、神经网络等多源数据融合方法。例如,张甫仁等人建立了室内环境的评价物元模型,给出了基于关联函数的可拓评价方法^[1]。刘大江利用熵与未知测度理论建立舒适度综合评价模型^[2]。Dounis 提出了应用模糊逻辑控制热舒适度标准的方法^[3];Bauke 等

人用模糊逻辑推理方法评价人们的工作条件^[4];谢东坡等人基于模糊理论完成了温度、湿度数值-符号转换以及舒适度合成,并探讨了在考虑室外环境参数以及人的心情等主观因素的情况下,舒适度合成方法的改进思路^[5]。赵博等人以温度、湿度、空气流速、平均辐射温度做输入量建立室内热舒适度的 BP 神经网络评判模型^[6];张方方等人提出了基于神经网络的室内环境热舒适度融合算法^[7];李婷等人采用 BP 神经网络(经 Levenberg-Marquart 算法优化)方法实时地确定 SET * 指标,分析热环境因素与 SET * 指标的关系,以 SET * 指标作为输出,影响 SET * 的环境变量作为输入,建立 SET * 指标的神经网络模型^[8]。

这些方法大都只是针对室内热、光、声环境舒适度或空气品质的某一方面,没有将其结合起来;另一方面

收稿日期:2012-09-20

修回日期:2012-12-24

网络出版时间:2013-03-05

基金项目:国家住建部科研项目(2010FJ3041);国家技术创新基金资助项目(11C26214302856);国家自然科学基金资助项目(6077311);湖南省科技计划项目(2012GK3086);2011年度湖南工业大学自然科学研究项目(2011HZX31)

作者简介:潘伟强(1987-),男,硕士研究生,研究方向为物联网感知技术;李长云,教授,博士,研究方向为物联网情境融合服务提供技术。

网络出版地址:http://www.cnki.net/kcms/detail/61.1450.TP.20130305.0817.039.html

这些方法是在群体感觉或者个体感觉上建立起来的单一评判模型,不能同时适用于个体和群体。针对上述问题,根据室内环境的特点,文中建立基于支持向量机的室内舒适度混合评判模型,并编写基于支持向量机的室内舒适度评价算法。仿真结果表明,在文中采集的数据集基础上,该方法性能稳定且分类准确率高。

1 基于支持向量机的室内舒适度评价模型

1.1 室内环境指标

室内环境质量主要取决于室内空气质量、热环境、噪声水平和照明采光等四个方面。通常而言,室内空气质量和热环境比其他两项对人的影响较为显著。影响室内环境质量的各指标具体如图1所示。

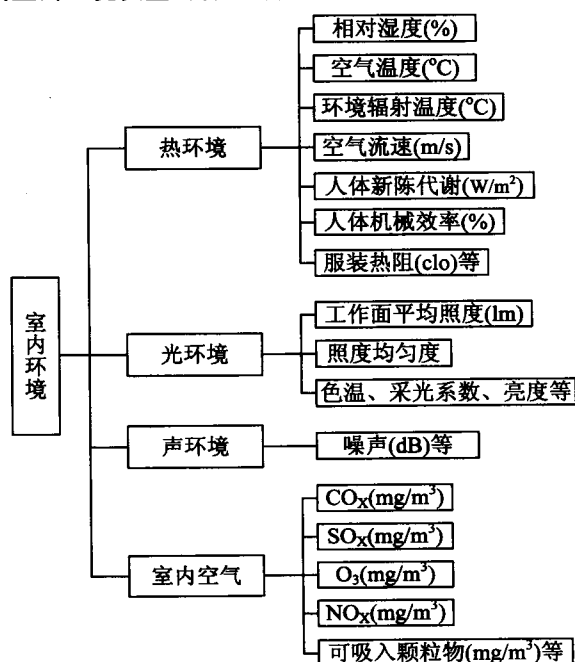


图1 室内环境指标

1.2 模型建立

室内舒适度是一个模糊的概念,是人们在众多因素作用下的主观反映。随着因素数量的增多,评价的难度越大;还有一些因素难以直接测量,评价方法难以将所有因素包括在内。所以文中选取既能直接采集,又能反映室内环境质量的主要因素作为室内舒适度评价指标,包括空气温度(°C)、空气相对湿度(%)、平均照度(lx)、CO₂浓度(%)、噪声(dB)等5个因素。该模型的目的是在群体感觉的基础上建立一个训练集,用这个训练集训练网络,得到一个初始化的网络;在使用网络过程中,根据房间主人的个体感觉,再训练网络,使得网络慢慢适用于个体。

室内舒适度作为模型的输出参数,该输出参数的取值范围为{4,3,2,1},分别表示舒适度的等级,即优、良、中、差。根据模型的设计目的,采用基于支持向量机的离线训练(off-line)和在线训练(on-line)两种

方法分别训练网络。室内舒适度混合评判模型的流程图如图2所示。

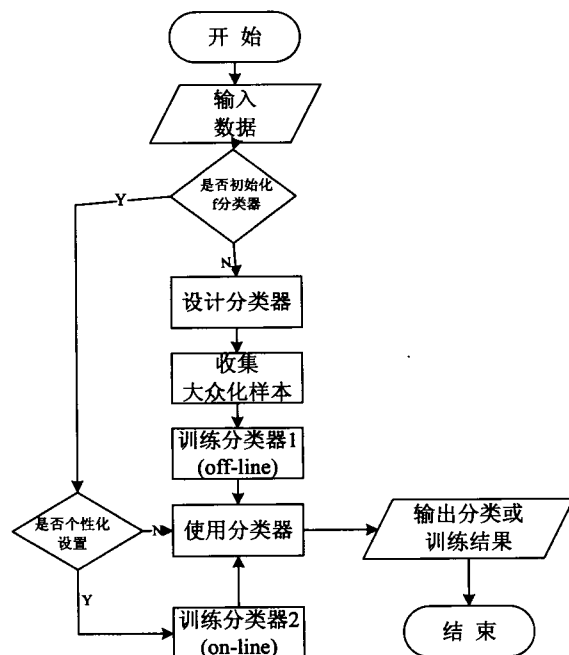


图2 室内舒适度混合评判模型的流程图

1.3 非线性分类器

支持向量机(Support Vector Machine,简称SVM)的概念是前苏联学者Vapnik等人在1974年提出来的,其主要思想是建立一个超平面作为决策曲面,使得正例和反例之间的隔离边缘被最大化。

与其他机器学习相比,SVM有以下特征:

1) SVM是以统计学习理论和结构风险最小化为基础,提高了分类器的泛化能力;

2) SVM中使用核函数,有效地克服了“维数灾难”问题;

3) SVM的训练问题实质上是一个二次规划(QP)问题,因此其局部最优解定为全局最优解^[9]。

文中采用SVM非线性分类模型C-SVC进行分类。其分类器决策函数为:

$$y = f(x) = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right) \\ = \text{sgn} \left(\sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \right) \quad (1)$$

其中, $K(x_i, x) = \exp \{ -(x - x_i)^2 / 2\sigma^2 \}$ 为高斯核函数,令参数 $g = 1/2\sigma^2$; x 为待分类的样本,训练集为 (x_i, y_i) , $i = 1, 2, \dots, n$, n 为训练样本个数, $x_i \in R^d$ 为训练样本, $y_i \in \{1, -1\}$ 是样本 x_i 的类标签, SV 为支持向量集是训练集的一个子集。参数 α_i 是通过训练(解优化问题(2)及其对偶问题(3))得到。

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \\ \text{subject to: } y_i (w^T \Phi(x_i) + b) \geq 1 - \xi_i \\ \xi_i \geq 0, i = 1, \dots, l \quad (2)$$

将输入空间映射到 Hilbert 空间, 得到原始问题 (2) 的对偶问题:

$$\begin{aligned} \max: L_D &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{subject to: } &C \geq \alpha_i \geq 0 \\ &\sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (3)$$

类与类之间交迭由惩罚系数 $C > 0$ 控制, $C = 0$ 则不允许交迭。 C 为常数, 它实际上起控制错分样本惩罚程度的作用, 目的在错分样本的比例与算法复杂度间的折中。

b 是分类的阈值, 因为支持向量 SV 满足公式 (4) 的等式, 所以可由任一个支持向量用下式求得 b 。

$$\begin{aligned} w &= \sum_{i=1}^n \alpha_i \\ y_i x_i \cdot y_i &= (w \cdot x_i + b) - 1, i = 1, 2, \dots, n \end{aligned} \quad (4)$$

算法步骤如下:

输入: 训练数据集; 输出: SVM 分类器。

步骤 1: 设置惩罚系数 C 和核参数 g , 并初始化分类器;

步骤 2: 导入训练数据集, 解二次规划方程 (3), 得到 α 向量;

步骤 3: 根据 α 向量寻找支持向量 SV , 按照公式 (4) 求 b ;

步骤 4: 将 SV, b, α 等带入决策函数 (1), 得到 SVM 分类器。

2 实验与分析

2.1 数据采集

实验数据是从湖南工业大学计算机与通信学院物联网实验室得到, 该实验室包括温湿度、光照、噪声、 CO_2 等五个环境指标的测量传感器和一些控制设备。通过实时监控室内环境各指标, 并将数据存入数据库中。从数据库中分层抽取样本以获取实验所用的包含 96 条记录的训练数据集、包含 24 条记录测试数据集^[10], 两个数据集没有重复的样本。为了确保数据的状态完好, 对采集到的数据进行数据预处理, 包括数据集成、数据清理、数据变换。其中数据清理包括对奇异值、缺失值等处理; 数据变换采用式 (5) 进行规范化, 达到扩大异类之间的距离、缩小同类之间距离的目的。由于数据集只有 5 个特征, 特征数目较少, 所以没有采用主成分分析法对数据集进行降维等数据归约处理。

$$y_i = (y_{\max} - y_{\min}) * \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} + y_{\min} \quad (5)$$

其中 y_{\max}, y_{\min} 分别为 0.9 和 0.1; x_{\max}, x_{\min} 为数据集的最大值和最小值。使用此规范化, 意味着测试数据

集的每一模式的所有特征分量的最大值 (最小值) 不会大于 (小于) 训练数据集的每一模式的所有特征分量的最大值 (最小值)。

根据 PMV-PPD 热舒适度方程、《建筑采光设计标准 GB/T50033-2001》、《声环境质量标准 GB3096-2008》和《室内空气质量标准 GB/T18883-2002》等领域知识, 规定如下规则:

- 1) 如果 R、K、S、G 都为舒适, 则 $D = \text{“优”}$;
- 2) 如果 R、K、S、G 有且只有一个为不舒适, 则 $D = \text{“良”}$;
- 3) 如果 R、K、S、G 有两个或三个为不舒适, 则 $D = \text{“中”}$;
- 4) 如果 R、K、S、G 都为不舒适, 则 $D = \text{“差”}$ 。

注: R、S、K、G 分别表示热舒适度、声舒适度、光舒适度、空气品质, 暂且粗略地分为舒适和不舒适, D 表示室内环境舒适度, 分为优、良、中和差四个等级。

参照文献[7]结论与展望部分, 根据领域知识 (即上述规则), 标记部分样本的标签, 利用半监督模糊聚类分析的方法对数据集进行多次聚类, 进而推测出其他样本的标签。图 3 表示 120 行 5 列的数据集的二维散点图矩阵, 从图中可以发现特征之间的简单关系, 如空气温度越高, 空气相对湿度越高; 同标签的数据集间具有结构上某种相同点等知识。表 1 记录了 24 条数据的测试数据集和对应的标签。

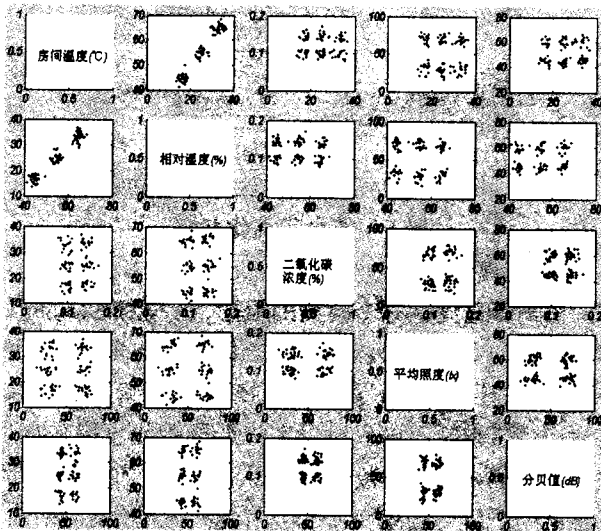


图 3 二维散点图矩阵

2.2 比较分析

为了检验基于 SVM 的室内舒适度评价方法的可行性和有效性, 文中进行了大量的仿真实验。实验的硬件平台是 CPU 为 Intel2.93GHz、内存为 2.00GB 的 PC 机; 软件平台是 Matlab R2009a, 把从真实环境中采集的夏季工况数据作为训练数据集和测试数据集。文中的方法与基于概率神经网络 (PNN) 的室内舒适度评价方法和文献[8]中提出的基于 BP 神经网络 (经 L-M

算法优化)的室内舒适度评价方法进行比较分析。

表1 测试数据集和对应的标签

编号	空气温度(℃)	相对湿度(%)	CO ₂ 浓度(%)	照度(lx)	噪声(dB)	标签
1	15.6	46.4	0.159	36	59	1(差)
2	30.7	64.8	0.147	34	62	1(差)
3	17.5	42.4	0.112	30	65	2(中)
4	17.3	43.2	0.131	34	43	2(中)
5	16.9	47.9	0.103	30	42	2(中)
6	19.1	41.2	0.138	73	60	2(中)
7	18.7	45.5	0.092	63	60	2(中)
8	13.2	46.4	0.130	65	43	2(中)
9	21.9	56.2	0.138	28	57	2(中)
10	26.3	54.4	0.135	34	45	2(中)
11	28.9	56	0.138	70	60	2(中)
12	32.7	64.8	0.095	28	60	2(中)
13	29.9	63.4	0.136	31	42	2(中)
14	31.5	62.7	0.119	27	43	2(中)
15	37	65.7	0.153	70	65	2(中)
16	33.5	64.1	0.092	78	57	2(中)
17	30.3	65.2	0.129	66	43	2(中)
18	15.4	45.5	0.096	77	43	3(良)
19	24.7	56	0.094	21	64	3(良)
20	26.7	55.3	0.1	31	41	3(良)
21	24.8	53.1	0.14	61	43	3(良)
22	32.6	64.5	0.086	63	43	3(良)
23	25.8	53.6	0.086	68	60	4(优)
24	25.3	54.6	0.094	72	43	4(优)

这三种方法有相同点也有不同点。其中概率神经网络(PNN)是径向基(RBF)神经网络的变化形式,同BP神经网络一样,都是前馈型网络,但它的设计思想同BP神经网络完全不同,与SVM的设计原理相似。PNN和SVM的设计原理是将复杂的模式分类问题非线性地投射到高维空间比投射到低维空间更可能是线性可分的。也就是说这个问题在低维空间不一定是线性可分的,但如果把它映射到高维度的空间去,在那里就可能是线性可分的。

文中采用Matlab中的神经网络工具箱编程实现了基于BP神经网络、PNN的室内舒适度评价方法,使用国立台湾大学林智仁的LibSVM工具箱^[11]编程实现了基于SVM的室内舒适度评价方法。分别进行如下两组实验:

(1)首先用训练数据集对分类器进行训练,再利用测试数据集来测试训练得到的模型,以平均训练时间和分类准确率做为评价分类器的性能指标,每组实验运行20次。

(2)采用交叉验证(Cross Validation,简称CV)的方法进一步检验三种模型的性能,首先将原始数据分

成5组(均分),然后将每个数据子集分别做一次验证集,其余的4组子集数据作为训练集,这样得到5个模型,用这5个模型最终的验证集的分类准确率的平均值作为5-CV分类器的性能指标,每组实验运行1次。

其中BP神经网络(经L-M算法优化)的参数设置为5个输入层、50个隐含层、1个输出层,拟合精度为0.0001;文中的SVM惩罚系数 $C=2$,高斯核函数参数 $g=8$;PNN等其他参数均采用默认值。上述实验的结果如表2和表3所示,其中加粗的值表示预测错误的值。

表2 预测结果的比较

样本编号	理论值	最坏结果			最好结果		
		BP	PNN	SVM	BP	PNN	SVM
1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1
3	2	1	2	2	2	2	2
4	2	2	2	2	2	2	2
5	2	2	2	2	2	2	2
6	2	1	2	2	2	2	2
7	2	2	2	2	2	2	2
8	2	3	3	2	2	3	2
9	2	2	2	2	2	2	2
10	2	2	2	2	2	2	2
11	2	1	2	2	2	2	2
12	2	2	2	2	2	2	2
13	2	2	2	2	2	2	2
14	2	2	2	2	2	2	2
15	2	2	2	2	2	2	2
16	2	2	2	2	2	2	2
17	2	2	2	2	4	2	2
18	3	4	3	3	3	3	3
19	3	2	3	3	3	3	3
20	3	4	3	3	3	3	3
21	3	3	3	3	3	3	3
22	3	3	3	3	3	3	3
23	4	3	4	4	4	4	4
24	4	4	4	4	4	4	4

分析表3可知:平均训练时间一栏,SVM训练花费的时间最少,几乎可以忽略不计,而BP神经网络训练花费的时间最多。BP神经网络是一种误差反向传播网络,为了达到预期的误差精度,需要经过成千上万次的训练,而SVM和PNN训练次数均远少于BP神经网络,所以BP神经网络花费的时间比其他方法的多。不管是平均预测准确率、最高预测准确率,还是交叉验证准确率,SVM的准确率(跟训练集有关)都是最高的。SVM和PNN以结构风险最小化为基础,而文中采用的数据集是用基于欧氏距离的半监督模糊聚类方法预处理过的,使得同标签的数据集有某种结构上的相

同点,所以 SVM 和 PNN 的各种分类准确率很高。BP 神经网络以经验风险最小化为基础,一味地减少与训练集的误差,出现了“过拟合”的现象,导致泛化能力降低,所以 BP 神经网络的各种分类准确率不高。综上所述,在文中采集的数据集的基础上,运用三种方法对室内舒适度进行综合评价,其中基于 SVM 舒适度评价方法的性能稳定且分类准确率高。

表 3 三种方法性能的比较

方法	平均训练时间(s)	平均预测准确率(%)	最高预测准确率(%)	交叉验证准确率(%)
SVM	0.0039	100.00	100.00	100.00
BP	2.1094	76.67	95.83	72.67
PNN	0.0398	95.83	95.83	99.17

3 结束语

文中根据群体感觉和个体感觉建立了一个基于支持向量机的室内舒适度评价方法混合评判模型,并编写了基于 SVM 的室内舒适度评价算法。通过与其他神经网络评价方法对比分析发现,该方法的训练时间少且分类准确率高。

实现了制作室内环境数据集的方法。参照文献[7]的展望部分,根据已标签的样本和领域知识,并利用半监督模糊聚类分析的方法对数据集进行聚类,进而推测出其他样本的标签。

由于文中方法数据预处理时使用半监督模糊聚类算法给原始数据集聚类,提高了分类准确率,所以用支持向量机和半监督模糊聚类算法的融合算法分层处理原始数据集^[12]是今后研究的一个重点。

参考文献:

- [1] 张甫仁,杨昭,郁文红.室内环境评价物元模型及可拓评价方法[J].天津大学学报,2005,38(4):307-309.
- [2] 刘大江.基于熵-未知测度模型的住宅建筑节能评价研究[J].价值工程,2011(2):75-76.
- [3] Dounis A I, Caraiscos C. Advanced control systems engineering for energy and comfort management in a building environment: A review[J]. Renewable and Sustainable Energy Reviews, 2009, 13(6): 1246-1261.
- [4] de Vries B, Steins R J. Assessing working conditions using fuzzy logic[J]. Automation in Construction, 2008, 17(5): 584-591.
- [5] 谢东坡,秦华锋,余成波.基于模糊理论的室内环境舒适度监测系统研究[J].中国测试技术,2008,34(4):126-128.
- [6] 赵博,连之伟,周湘江.基于神经网络的室内热舒适评判模型[J].哈尔滨工业大学学报,2003,35(12):1437-1438.
- [7] 张方方.室内环境多信息融合算法的研究[D].重庆:重庆理工大学,2010.
- [8] 李婷,陈渊馨.室内环境舒适度的神经网络建模与仿真[J].计算机仿真,2011,28(6):189-192.
- [9] Vapnik V N. 统计学习理论[M].许建华,张学工,译.北京:电子工业出版社,2004.
- [10] 室内舒适度训练集和测试集[EB/OL]. 2012-08-21. <http://www.datatang.com/datasets/detail.aspx?id=43623>.
- [11] Chang Chih-Chung, Lin Chih-Jen. LIBSVM: a library for support vector machines[EB/OL]. 2001. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [12] 李向阳,李玲娟,陈建新.面向情感感知的不确定性数据融合策略[J].计算机技术与发展,2012,22(2):127-130.

(上接第 213 页)

参考文献:

- [1] Markowitz H. Portfolio Selection[J]. Journal of Finance, 1952, 7(1): 77-91.
- [2] Syam S S. A dual ascent method for the portfolio selection problem with multiple constraints and linked proposals[J]. European Journal of Operations Research, 1998, 108(1): 196-207.
- [3] Andre F P. Large-scale portfolio optimization[J]. Management Science, 1984, 31(10): 1143-1159.
- [4] Hamza F, Janssen J. The mean-semivariances approach to realistic portfolio optimization subject to transaction costs[J]. Applied Stochastic Models and Data Analysis, 1998, 14(4): 275-283.
- [5] Brorsen B W, Buck D W, Konntz S R. Hedging Hard Red Winter Wheat: Kansas City versus Chicago[J]. The Journal of Futures Markets, 1998, 18(4): 449-466.
- [6] 陈叔平,李胜宏,吴雄伟.一类投资组合优化问题的求解及实证分析[J].高校应用数学学报 A 辑,2000,15(4):491-498.
- [7] 陈炜,张润彤,杨玲.存在融资条件下证券组合选择的一种模糊决策方法[J].北京交通大学学报,2007,6(1): 67-70.
- [8] 华仁海,仲伟俊.对我国期货市场价格发现功能的实证分析[J].南开管理评论,2002,5(5):57-60.
- [9] 华仁海,陈百助.我国期货市场价格收益及波动方差的长记忆性研究[J].金融研究,2004(2):52-61.
- [10] 张卫国.不相关证券组合投资优化模型及实证分析[J].系统工程理论与实践,1998,18(4):34-40.
- [11] 华仁海,仲伟俊.我国期货市场价格收益、交易量、波动性关系的动态分析[J].统计研究,2003,7(7):25-29.
- [12] 于维杰.组合证券投资的有效边界[J].数理统计与管理, 1996, 15(3): 27-31.

基于支持向量机的室内舒适度评价方法

作者: [潘伟强](#), [李长云](#), [胡盛龙](#), [PAN Wei-qiang](#), [LI Chang-yun](#), [HU Sheng-long](#)
作者单位: [湖南工业大学计算机与通信学院, 湖南株洲, 412007](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2013, 23 (6)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjtz201306055.aspx