

利用遗传算法和支持向量机测算农用地理论单产和可实现单产

王海军^{1,2}, 柳敏燕¹, 高娟¹

(1. 武汉大学资源与环境科学学院, 武汉 430079; 2. 香港大学城市规划与设计系, 香港 999077)

摘要: 为了解决传统的采用一元线性回归法计算农用地理论单产和可实现单产的局限性, 该文将遗传算法和支持向量回归机理论应用于农用地产能核算, 构建理论单产和可实现单产核算模型。通过建立样点分等因素质量分与理论单产的遗传算法支持向量回归机 (genetic algorithm-support vector machine, GA-SVM) 模型进行理论单产预测; 建立样点分等因素质量分和利用系数之积与可实现单产的遗传算法支持向量回归机模型进行可实现单产预测。以广东省揭西县产能核算为例, 分别采用 GA-SVM 模型和一元线性回归模型测算, 并将测算结果进行对比分析。研究结果证明, GA-SVM 对于理论单产和可实现单产的测算精度更高, 适于单个样本值的预测, 可以作为农用地产能核算的一种新方法。

关键词: 土地利用, 产能, 遗传算法, 支持向量机, 理论单产, 可实现单产, 农用地

doi: 10.3969/j.issn.1002-6819.2013.19.030

中图分类号: F301.21

文献标志码: A

文章编号: 1002-6819(2013)-19-0244-09

王海军, 柳敏燕, 高娟. 利用遗传算法和支持向量机测算农用地理论单产和可实现单产[J]. 农业工程学报, 2013, 29(19): 244-252.

Wang Haijun, Liu Minyan, Gao Juan. Calculation of theoretical and accessible yields of agricultural land based on genetic algorithm and support vector machine[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(19): 244-252. (in Chinese with English abstract)

0 引言

农用地理论单产^[1]是在农业生产条件得到充分保证, 光、热、水、土等环境因素均处于最优状态, 技术因素所决定的农作物所能达到的单位面积最高产量。可实现单产^[1]是在农业生产条件得到基本保证, 光、热、水、土等环境因素均处于正常状态、技术条件可以满足, 由政策、投入等因素决定的正常年景下农作物能够获得的单位面积最高产量。农用地产能核算主要是依据所有分等单元的单产乘以相应分等单元的耕地面积进行, 因此, 农用地理论单产和可实现单产是测算农用地理论产能和可实现产能的基础, 也是农用地产能核算的研究重点^[2-5]。

农用地产能核算是国土资源大调查的重要组成部分。近年来, 关于产能核算的研究层出不穷^[6-9]。目前, 全国范围内开展的产能核算采用的方法是一元线性回归法。但是, 随着农用地产能核算工作的不断深入, 传统的产能核算方法的局限性也逐步显现出来。传统农用地产能核算主要是基于农用地分

等成果, 采用的建模方法是一元线性回归模型, 这种建模方法是用一条直线对大量分散的样本进行拟合, 其拟合效果往往不太理想^[10-13]。并且, 这种建模方法在计算样本理论单产与可实现单产时, 选取的自变量分别是自然等指数以及利用等指数, 但是自然等指数与利用等指数是根据农用地分等因素质量分加权测算得来的, 其测算结果的准确性受权重选取影响较大, 经过多步计算得到单产, 误差进一步累加, 从而导致产能核算结果偏差加大。

针对上述问题, 本文结合遗传算法 (genetic algorithm) 和支持向量机理论 (support vector machine), 构建遗传算法支持向量回归机 (genetic algorithm-support vector machine, GA-SVM) 模型进行理论单产与可实现单产的测算。其输入变量可以是多维的, 可以直接将各个分等单元的分等因素质量分作为输入项来进行预测, 由此得出的测算结果可以避免多步计算所带来的误差累加问题。近年来, 有学者^[14-16]尝试运用遗传算法对支持向量机参数进行优化, 建立遗传算法支持向量回归机 (GA-SVM)。目前, GA-SVM 模型已运用于工农业生产^[17-18]、经济学^[19-20]、环境^[21]、医学研究^[22-23]等领域, 进行预测分析^[24-28]、参数优化^[29]、遥感反演^[30]等, 取得了比较理想的试验结果。采用 GA-SVM 模型计算理论单产和可实现单产, 进一步可进行农用地产能核算, 获得更加科学准确的产能

收稿日期: 2013-03-10 修订日期: 2013-08-26

基金项目: 国家自然科学基金项目 (40871179)

作者简介: 王海军 (1972-), 男, 陕西凤翔人, 副教授, 博士, 主要从事土地资源评价和土地利用覆盖变化、地理元胞自动机理论与方法等方面的研究。武汉 武汉大学资源与环境科学学院, 430079。

Email: landgiswhj@163.com

核算结果，更加真实地反映中国粮食综合生产能力状况，以期为国家科学合理制定政策提供依据。

1 研究区概况与数据来源

揭西县位于广东省东部，揭阳市之西，榕江南河的中上游，地处 $115^{\circ}36'22'' \sim 116^{\circ}11'15''E$ ， $23^{\circ}18'53'' \sim 23^{\circ}41'13''N$ 。东连揭东县，西接汕尾市的陆河县，南邻普宁市，北与五华、丰顺接壤。全县东西长约 51 km，南北宽约 36.6 km，北回归线恰好通过县城北隅，横贯全县中部，属亚热带季风气候区。揭西县隶属于揭阳市，共辖 16 个镇 1 个乡。

本文数据以 2011 年揭西县农用地分等成果和农用地产能核算成果为基础。通过对指定作物审定品种的区域试验单产进行调查，作为农用地理论单产的样本值。对样本单元指定作物近 3~5 a 正常年景下的单产进行调查，并把这些年来农用地的最高单产看作农用地可实现单产样本值。由揭西县统计局提供统计年鉴，将各乡（镇）指定作物的统计单产作为指定作物的实际单产。

2 研究方法

近年来，支持向量机被广泛应用于回归问题。核函数的选择是支持向量机的核心问题，直接关系到它在高维空间中的推广性。本文采用支持向量机来进行回归预测，使用的核函数是径向基向量函数（radial basis function, RBF）。径向基向量核函数能够逼近任意非线性函数，在高维空间具有较强的推广性。

以径向基向量函数为核的支持向量机的参数选择对于支持向量机的优劣有着决定性的作用，必须综合考虑惩罚系数 C 和核参数 σ 。传统支持向量机在进行参数选择的时候一般选取的是经验值，但是这并不一定是最适合的参数，如果多次选取不同参数进行调试，则耗时耗力，并且也不一定能得到最优解。因此，考虑采用遗传算法^[31]优化参数，以获取最优解。

本文利用 Matlab 软件，将支持向量机与遗传算法结合，构建基于 GA-SVM 的理论单产和可实现单产预测模型。GA-SVM 算法的大致实现过程如图 1 所示。

将 GA-SVM 算法与农用地产能核算结合起来，构建基于 GA-SVM 的理论单产和可实现单产预测模型过程如下：

1) 理论单产建模样点选取

本次产能核算共划定分等单元 2 525 个，理论单产样点根据农业主管部门或农科院提供的指定作物审定品种区域试验田数据以及测土配方数据

设置，将各审定品种的区域试验单产视作农用地理论单产的样本值。揭西县按照两头等取 1~2 个样点，中间等取 3~4 个样点的规则，结合各等级进行布设样点。

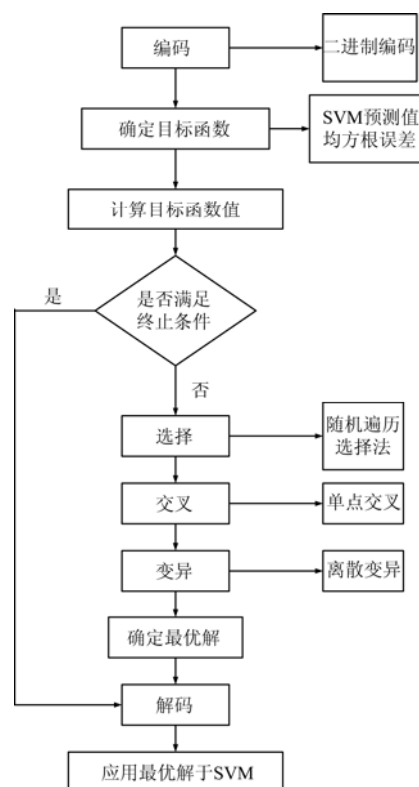


图 1 GA-SVM 算法实现流程图
Fig.1 GA-SVM algorithm flowchart

根据收集的指定作物理论单产样点，结合自然质量等和利用等进行分析，整理分析区域内农用地的理论产能，并且对比筛选前后的样点产能范围，检验所选取的样点是否符合实际要求。最终选取水田理论产能核算样点 24 个，旱地理论产能核算样点 26 个，如图 2a 所示。

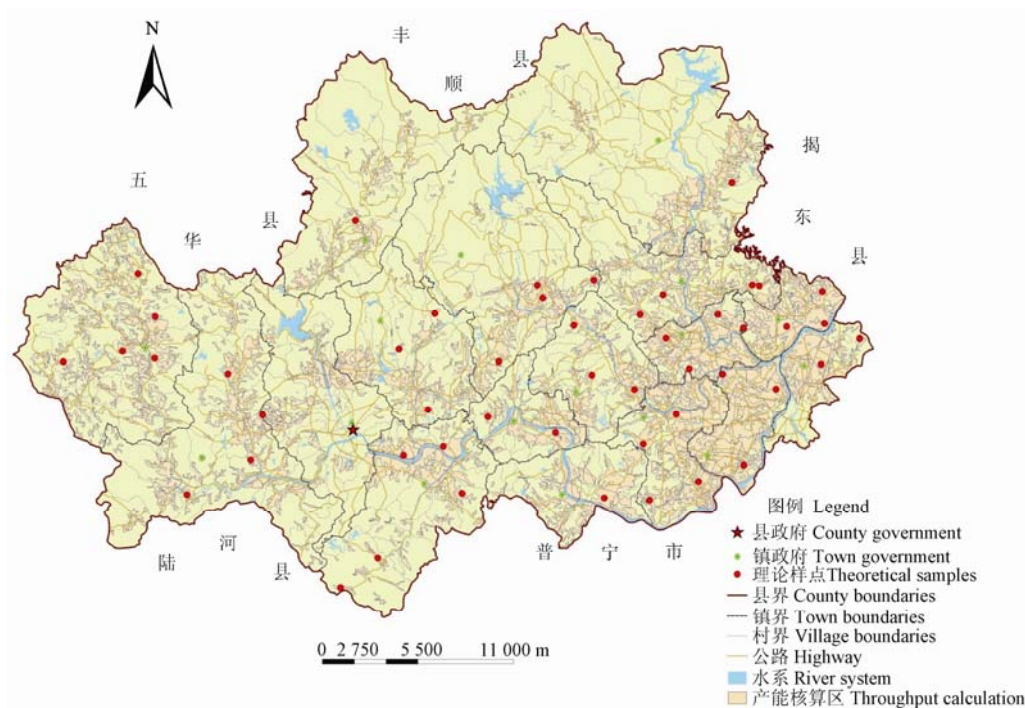
2) 可实现单产建模样点选取

可实现单产样点调查优先将指标区内国家级和省级标准样地作为调查点。揭西县按照两头等不少于 10 个，中间等不少于 20 个的原则计算。根据收集的指定作物可实现产能样点，结合自然质量等和利用等进行分析，整理分析区域内农用地的可实现产能，并且对比筛选前后的样点产能范围，检验所选取的样点是否符合实际要求。最终选取水田可实现产能核算样点 148 个，旱地可实现产能核算样点 53 个，如图 2b 所示。

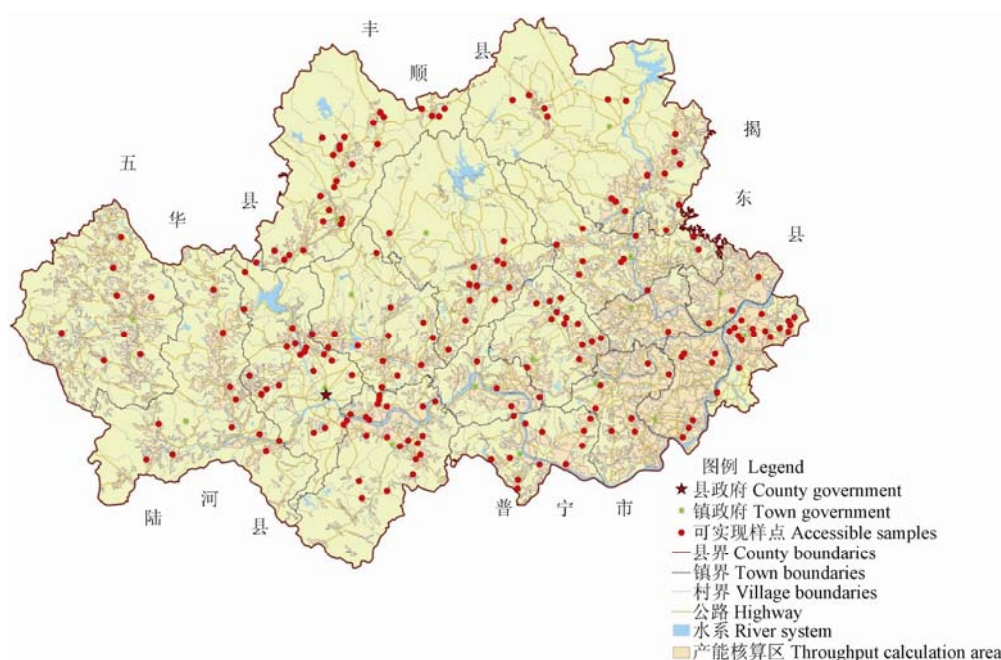
3) 分别确定理论单产 GA-SVM 模型和可实现单产 GA-SVM 模型的输入变量。根据农用地分等成果，获取分等单元的分等因素相关数据，并参照“指定作物—分等因素—农用地质量分”关系表，

计算在指定作物条件下各分等因素相应的质量分。理论单产 GA-SVM 模型的输入变量是理论单产样点的分等因素质量分, 可实现单产 GA-SVM 模型

的输入变量是可实现单产样点的分等因素质量分与其利用系数的乘积, 它们的输出变量分别是理论单产和可实现单产。



a. 理论产能样点
a. Theoretical productivity samples



b. 可实现产能样点
b. Accessible productivity samples

图 2 样点分布图

Fig.2 Samples distribution pattern

4) 进行惩罚系数和核参数的编码。在设定的值域范围内, 随机产生一定组数惩罚系数和核参数, 对其进行二进制编码, 由此产生的编码串就构

成了遗传算法的初始种群。

5) 确定目标函数。在单产测算中, 将样点预测单产和实际单产之间的方差 (这里指真实值与预

测值的均方根误差的平方，下同）作为目标函数。目标函数的值越小，就说明预测值更加接近真实值，结果更加准确。

6) 计算适应度函数值。将初始种群的数值代入支持向量机模型，分别计算各组数值所构成的支持向量机的预测结果，然后计算方差。

7) 计算遗传算法算子。首先是选择算子：按个体适应度大小进行选择，采用随机遍历方法进行个体选择，它是对轮盘赌方法进行改进后的方法，不过此方法只需要进行一次轮盘旋转。 n 为需要选择的个体数，采用的指针数量也为 n ，选择指针的距离的 $1/n$ ，第一个指针由 $[0,1/n]$ 区间的均匀随机数决定。其次是交叉算子：本模型采用单点交叉，即在随机设置的交叉点处互换 2 个配对个体的部分染色体。最后是变异算子：由于本文采用的是二进制编码，即将某些位上的 0 变为 1、1 变为 0，生成新的个体。

8) 最后是终止条件判断，遗传算法达到指定的代数并且获得最优惩罚系数和核参数，使支持向量回归机预测准确率达到最大，则终止判断。

3 结果与分析

3.1 基于 GA-SVM 揭西县单产预测

本文采用 Matlab 软件，结合 GA 和 SVM 工具箱进行编程，构建基于 GA-SVM 的理论单产和可实现单产预测模型，分别将其用于水田单产预测和旱地单产预测。

3.1.1 水田理论单产预测

经过多次试验，早稻、晚稻和冬甘薯理论单产 GA-SVM 模型在 500 代时已达到收敛条件，获得其最佳惩罚系数 C 和最佳核参数 g 。因此，设置早稻、晚稻和冬甘薯的理论单产 GA-SVM 模型最大进化代数为 500，最终获得的早稻、晚稻和冬甘薯的最佳惩罚系数 C 分别为 127.42、57.99 和 72.29，早稻、晚稻和冬甘薯的最佳核参数 g 分别为 0.32、28.15 和 59.28（为了便于计算取 $g=1/(2\sigma^2)$ ，下同）。获得了最优参数以后，将其代入支持向量机，利用支持向量机模型来进行早稻、晚稻和冬甘薯的理论单产样点回归预测，得到回归预测结果。其中，早稻理论单产样点预测结果如图 3 所示。

将采用 GA-SVM 测算指定作物产量折算为标准粮产量，根据耕作制度，揭西县农用地中水田作物为两年五熟，因此，水田的标准粮实际产量具体计算公式为：水田标准粮实际产量=早稻单产×早稻产量比系数+晚稻单产×晚稻产量比系数+半季冬甘薯单产×冬甘薯产量比系数。在实际操作时，

调查得到的样点甘薯产量均为鲜质量，结合《广东省农用地产能核算》中相关技术说明，将鲜质量按照五折一得到标准粮产量。得到理论单产结果如表 1 所示。

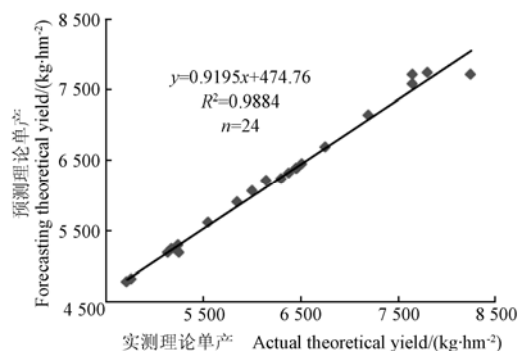


图 3 早稻理论单产 GA-SVM 预测结果图
Fig.3 Prediction results of theoretical yield of early rice using GA-SVM algorithm

3.1.2 水田可实现单产预测

在计算可实现单产时，将选取样点的分等质量分与土地利用系数之积作为属性项，将选取样点的可实现单产作为预测项进行样本训练，获得最优的 GA-SVM 模型，然后将其运用于其余分等单元的可实现单产预测。

设置早稻、晚稻和冬甘薯可实现单产 GA-SVM 模型的最大进化代数为 500，分别获得最佳惩罚系数 C 和最佳核参数 g 。获得了最优参数后，将其代入支持向量机，利用支持向量机模型来进行早稻、晚稻和冬甘薯可实现单产样点回归预测，得到所有样点预测结果。以早稻为例，可实现单产 GA-SVM 预测结果如图 4 所示。同样地，计算揭西县各乡镇水田可实现单产，得到结果如表 1 所示。

3.1.3 旱地理论单产预测和可实现单产预测

采用 GA-SVM 模型来测算揭西县旱地的理论单产和可实现单产。由于旱地的理论单产和可实现单产预测方法与水田基本一致，这里不再赘述，只给出最后预测结果。

同样地，采用 GA-SVM 测算指定作物产量折算为标准粮产量。根据耕作制度，揭西县农用地中旱地作物为一年两熟，因此，旱地的标准粮实际产量具体计算公式为：旱地标准粮实际产量=花生单产×花生产量比系数+秋甘薯单产×秋甘薯产量比系数。在实际操作时，调查得到的样点甘薯产量均为鲜质量，结合《广东省农用地产能核算》中相关技术说明，将鲜质量按照五折一得到标准粮产量。计算揭西县各乡镇旱地理论单产和可实现单产得到结果如表 1 所示。

表 1 揭西县各乡镇水田和旱地理论单产、可实现单产成果表
Table 1 Theoretical and accessible yields of paddy field and dry land of all towns in Jiexi

乡(镇)编号 Number of town	乡(镇)名称 Name of town	水田 Paddy field			旱地 Dry land		
		面积 Acreage/hm ²	理论单产 Theoretical yield/(kg·hm ⁻²)	可实现单产 Accessible yield/(kg·hm ⁻²)	面积 Acreage/hm ²	理论单产 Theoretical yield/(kg·hm ⁻²)	可实现单产 Accessible yield/(kg·hm ⁻²)
1	河婆镇	1193.71	14955.75	13024.78	460.26	10432.74	9541.30
2	龙潭镇	1265.52	14858.19	13509.00	140.02	10697.12	10553.35
3	南山镇	1050.11	14822.54	12970.98	74.36	10628.38	9583.38
4	五经富镇	1395.60	14712.85	13649.91	70.07	10545.06	9710.03
5	京溪园镇	1012.83	14797.83	13974.34	165.42	10576.43	9613.28
6	灰寨镇	1048.42	15135.09	12892.14	103.37	10668.67	9655.08
7	塔头镇	1254.72	15110.74	13848.56	253.98	10646.95	10483.50
8	东园镇	881.92	14843.72	13778.86	160.57	10643.30	10557.20
9	凤江镇	1834.54	14783.35	14047.21	310.11	10671.98	10640.64
10	棉湖镇	1208.40	14852.14	14786.20	146.95	10642.69	9752.89
11	金和镇	2106.19	15617.58	13950.01	196.18	10656.81	10636.10
12	大溪镇	693.58	14972.80	13850.86	158.40	10663.12	10607.01
13	钱坑镇	1027.08	14748.66	14227.73	110.01	10649.47	9676.49
14	坪上镇	1339.17	14985.44	14946.53	329.83	10507.00	10178.91
15	五云镇	1896.84	14625.16	13508.67	415.76	10645.22	10267.58
16	上砂镇	1779.61	14682.54	13956.83	484.39	10817.67	10019.93
17	良田乡	653.02	14749.63	12094.04	168.39	11119.99	9368.61
全县	—	21641.26	253254.01	233016.65	3748.07	181212.60	170845.28

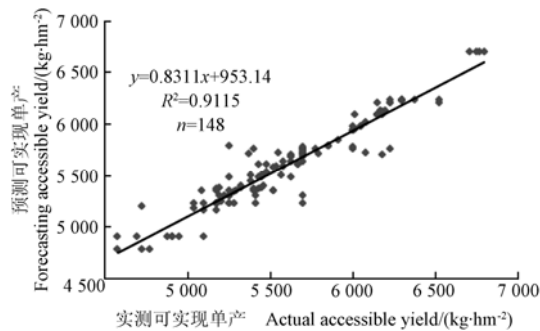


图 4 早稻可实现单产 GA-SVM 预测结果图

Fig.4 Prediction results of accessible yield of early rice using GA-SVM algorithm

3.2 结果对比分析

3.2.1 水田样点单产结果对比

分别采用 GA-SVM 方法和一元线性回归法测算了水田跟旱地的理论样点和可实现样点的产值，并将测算结果的精确度进行对比分析。

水田样点理论单产误差分析结果如表 2 所示。从表 2 中可以看出，采用 GA-SVM 方法测算的结果比采用一元线性回归法测算的结果具有较小的最大相对误差、最大绝对误差、平均绝对误差以及均方根误差，具有较大的最小相对误差以及最小绝对误差。虽然在最小绝对误差以及相对误差上，一元线性回归法值更小，但从总体上来看，用 GA-SVM 方法测算结果具有更高的精确度。

表 2 水田样点理论单产和可实现单产预测结果误差表
Table 2 Deviations of theoretical yield and accessible yield of paddy field samples

	预测方法 Forecasting methods	最小相对误差 Minimum relative error/%	最大相对误差 Maximum relative error/%	最小绝对误差 Minimum absolute error	最大绝对误差 Maximum absolute error	平均绝对误差 Mean absolute error	均方根误差 Root mean square error
理论单产 Theoretical yield	GA-SVM	0.05	8.28	0.54	105.00	16.13	26.53
	一元线性回归法 Linear regression	0.02	25.57	0.26	147.29	64.69	81.48
可实现单产 Accessible yield	GA-SVM	0.01	9.01	0.11	75.31	13.18	19.36
	一元线性回归法 Linear regression	0.08	14.94	0.68	135.90	49.07	58.87

为了直观的将一元线性回归法测算结果、GA-SVM 方法测算结果进行对比，描绘了 2 种方法的预测结果与标准粮真实值的对比图，如图 5 所示。从图 5b 中可以看出 GA-SVM 预测结果跟

真实值十分接近， $R^2=0.9801$ ；而图 5a 中用一元线性回归法预测的结果重合度相对较低， $R^2=0.5703$ 。GA-SVM 的拟合结果明显优于一元线性回归法。

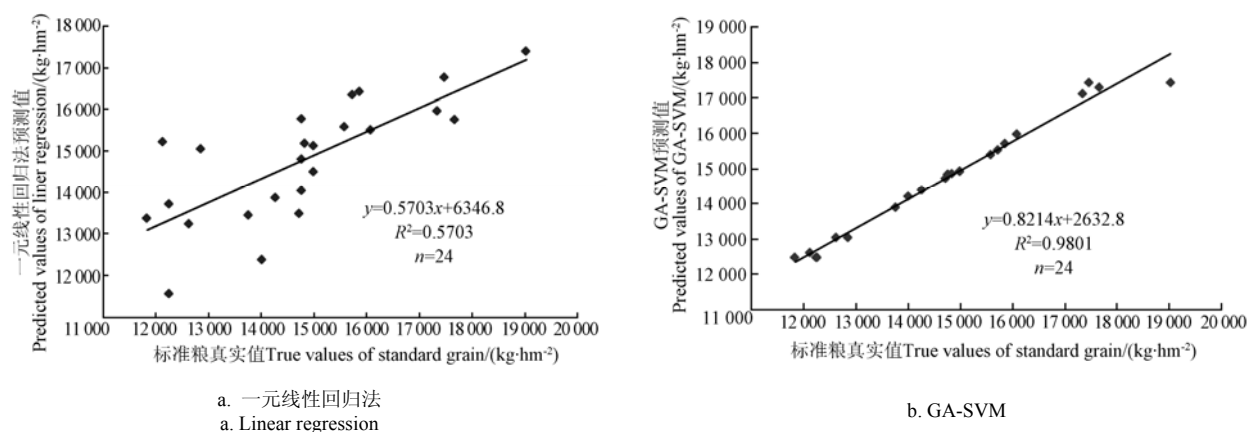


图 5 揭西县水田样点理论单产预测结果对比图

Fig.5 Prediction results of theoretical yield of paddy field samples in Jiexi

水田样点可实现单产误差结果如表 2 所示。从表 2 中可以看出,采用 GA-SVM 方法测算的结果比采用一元线性回归法测算的结果各个误差统计项的值都要小。这表明,采用 GA-SVM 预测所得到的结果更加接近真实值,其预测结果较之一元线性回归法更为准确。

图 6a、图 6b 分别是揭西县水田 148 个可实

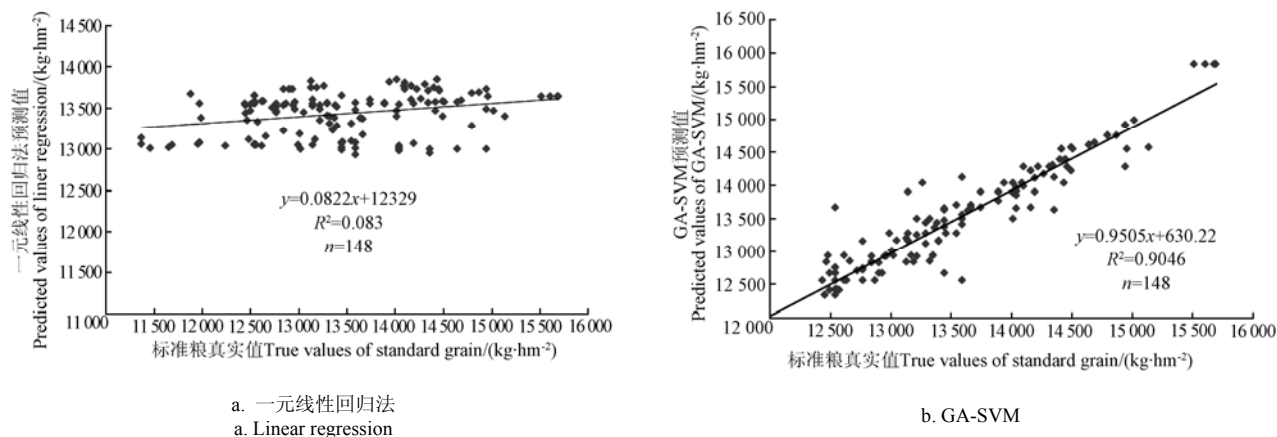


图 6 揭西县水田样点可实现单产预测结果对比图

Fig.6 Prediction results of accessible yield of paddy field samples in Jiexi

3.2.2 旱地样点单产结果对比

同样地,对揭西县旱地的理论单产和可实现单产结果分别进行分析。旱地样点理论单产误差分析结果如表 3 所示。从表 3 中可以看出,GA-SVM 法与一元线性回归法相比,具有较大的最小绝对误差,具有相同的最小相对误差,具有较小的最大相对误差、最大绝对误差、平均绝对误差以及均方根误差。这表明,GA-SVM 在进行单个样点理论单产预测时比一元线性回归法具有更大的准确性。

用同样的方法对旱地可实现单产进行分析,旱地样点可实现单产误差分析结果如表 3 所示。从表 3 中可以看出,GA-SVM 所得结果各项误差分析值都比一元线性回归法要小。对于旱地样点可实现单

现产能样点的一元线性回归法预测值与标准粮真实值、GA-SVM 预测值与标准粮真实值的对比图。从图 6b 中可以看出,GA-SVM 预测结果与真实值更为接近, $R^2=0.9046$; 而图 6a 中用一元线性回归法预测的结果重合度较低, R^2 仅为 0.083。GA-SVM 的拟合结果明显优于一元线性回归法。

产的单点预测,GA-SVM 法优于一元线性回归法,具有更高的准确度。

对各样点理论单产和可实现单产预测结果误差进行分析可知,一元线性回归法的平均绝对误差是 GA-SVM 法的 2~4 倍;GA-SVM 法预测结果比一元线性回归法更接近标准粮真实值,前者的决定系数 R^2 均达到 0.90 以上,而后的决定系数 R^2 较小。原因为一元线性回归法在计算样本理论单产与可实现单产时,选取的自变量分别是自然等指数以及利用等指数,经过多步计算,单产、产能核算结果的偏差加大。采用 GA-SVM 进行建模,直接将各个分等单元的分等因素质量分作为输入项来进行预测,能避免多步计算带来的误差累加的问题。

表 3 旱地样点理论单产和可实现单产预测结果误差表

Table 3 Deviations of theoretical yield and accessible yield of dry land samples

	预测方法	最小相对误差	最大相对误差	最小绝对误差	最大绝对误差	平均绝对误差	均方根误差
	Forecasting methods	Minimum relative error/%	Maximum relative error/%	Minimum absolute error	Maximum absolute error	Mean absolute error	Root mean square error
理论单产 Theoretical yield	GA-SVM	0.15	5.61	1.10	37.59	12.73	16.06
	一元线性回归法 Linear regression	0.15	7.65	1.03	53.51	25.83	30.39
可实现单产 Accessible yield	GA-SVM	0.02	17.53	0.17	108.53	14.59	22.95
	一元线性回归法 Linear regression	0.28	26.56	2.00	176.14	59.59	72.86

4 结 论

对于理论单产和可实现单产的预测,一元线性回归法的平均绝对误差是 GA-SVM 法的 2~4 倍; GA-SVM 法预测结果比一元线性回归法的预测结果更接近标准粮真实值,前者的决定系数 R^2 均达到 0.90 以上,而后的决定系数 R^2 较小,用 GA-SVM 模型预测农用地单产是可行性,预测结果更准确。

基于一元线性回归法的理论单产与可实现单产测算需要进行大量计算,过程繁琐,易出现人为错误;而 GA-SVM 法主要是通过运行程序来获取结果,过程简单,省时省力,并可避免人为错误,适用于大量样本的理论单产与可实现单产测算,具有较强的推广性。尽管本文采用 GA-SVM 模型对县域尺度的农用地理论单产和可实现单产进行了测算,并取得了理想的试验结果。然而,对于更大尺度,如市域、省域尺度的农用地单产核算等,该方法的适用性有待今后进一步的研究。

[参 考 文 献]

- [1] 赵永华,崔璐,卫新东,等. 泾阳县农用地产能及其空间差异性[J]. 中国农学通报, 2011, 27(24): 112—118.
Zhao Yonghua, Cui Lu, Wei Xindong, et al. Study on cultivated land productivity calculation of Jingyang county in Shaanxi province[J]. Chinese Agricultural Science Bulletin, 2011, 27(24): 112—118. (in Chinese with English abstract)
- [2] 吴克宁,程先军,黄勤,等. 基于分等成果的农用地综合生产能力[J]. 农业工程学报, 2008, 24(11): 51—56.
Wu Kening, Cheng Xianjun, Huang Qin, et al. Comprehensive productivity of agricultural land based on the agricultural land classification[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2008, 24(11): 51—56. (in Chinese with English abstract)
- [3] 李团胜,张艳,闫颖,等. 基于农用地分等成果的陕西周至县耕地粮食生产能力测算[J]. 农业工程学报, 2012, 28(15): 193—198.
Li Tuansheng, Zhang Yan, Yan Ying, et al. Calculation of farmland grain potential productivity of Zhouzhi County based on farmland classification[J]. Transactions

of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(15): 193—198. (in Chinese with English abstract)

- [4] 门明新,张俊梅,刘玉,等. 基于综合生产能力核算的河北省耕地重点保护区划定[J]. 农业工程学报, 2009, 25(10): 264—271.
Men Mingxin, Zhang Junmei, Liu Yu, et al. Division of key cultivated land regions based on calculation of comprehensive productivity in Hebei province[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(10): 264—271. (in Chinese with English abstract)
- [5] 曹操,张云涛. 浅析农用地产能核算及其成果应用[J]. 商业文化: 学术版, 2011(8): 270—270.
- [6] 李天杰,郇文聚,赵焯,等. 土地质量、生产能力与粮食安全相关研究的现状及展望[J]. 资源与产业, 2006, 8(1): 19—23.
Li Tianjie, Yun Wenju, Zhao Ye, et al. Situation and prospects of researches on land quality, land productivity and food security[J]. Resources and Industries, 2006, 8(1): 19—23. (in Chinese with English abstract)
- [7] 张千五,王数,张凤荣,等. 基于农用地分等的粮食生产能力田间质量限制研究[J]. 农业工程学报, 2008, 24(10): 85—88.
Zhang Qianwu, Wang Shu, Zhang Fengrong, et al. Field quality limit to grain potential productivity based on farmland classification[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2008, 24(10): 85—88. (in Chinese with English abstract)
- [8] 李翠珍,孔祥斌,秦静,等. 大都市区农户耕地利用及对粮食生产能力的影响[J]. 农业工程学报, 2008, 24(1): 101—107.
Li Cuizhen, Kong Xiangbin, Qin Jing, et al. Cultivated land use of peasant households in metropolitan suburbs and its effect on grain production capacity[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2008, 24(1): 101—107. (in Chinese with English abstract)
- [9] 周健民. 加强我国粮食安全保障能力建设的思考[J]. 中国科学院院刊, 2004, 19(1): 40—44.
- [10] 伍育鹏,郇文聚,邹如. 耕地产能核算模型的研究[J]. 农业工程学报, 2008, 24(增刊 2): 108—113.
Wu Yupeng, Yun Wenju, Zou Ru. Model for calculation of cultivated land productivity[J]. Transactions of the Chinese Society of Agricultural Engineering

- (Transactions of the CSAE), 2008, 24(Supp.2): 108—113. (in Chinese with English abstract)
- [11] 孔祥斌, 李翠珍, 赵晶, 等. 乡镇尺度耕地生产能力实现程度分析与实证[J]. 农业工程学报, 2010, 26(12): 345—351.
Kong Xiangbin, Li Cuizhen, Zhao Jing, et al. Method and empirical research on the realization degree of arable land production capacity at town level[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(12): 345—351. (in Chinese with English abstract)
- [12] 张蕾娜, 郇文聚, 苏强, 等. 基于农用地分等成果的产能核算研究[J]. 农业工程学报, 2008, 24(增刊 1): 133—136.
Zhang Leina, Yun Wenju, Su Qiang, et al. Methods for productivity calculation based on agricultural land classification[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2008, 24(Supp.1): 133—136. (in Chinese with English abstract)
- [13] 王晓峰, 曹建农, 卫新东, 等. 基于农用地分等成果的耕地产能核算研究: 以陕西省扶风县为例[J]. 陕西师范大学学报: 自然科学版, 2010, 38(6): 89—93.
Wang Xiaofeng, Cao Jiannong, Wei Xindong, et al. Study on cultivated land productivity calculation based on agricultural land classification: A case study of Fufeng county, Shaanxi province[J]. Journal of Shaanxi Normal University: Natural Science Edition, 2010, 38(6): 89—93. (in Chinese with English abstract)
- [14] 王禾军, 邓飞其. 基于模糊最小二乘支持向量机的区域粮食安全性预警分析[J]. 农业工程学报, 2011, 27(5): 190—194.
Wang Hejun, Deng Feiqi. Analysis on region grain security warning based on fuzzy least squares support vector machines[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(5): 190—194. (in Chinese with English abstract)
- [15] 张豪, 罗亦泳, 张立亭, 等. 基于遗传算法最小二乘支持向量机的耕地变化预测[J]. 农业工程学报, 2009, 25(7): 226—231.
Zhang Hao, Luo Yiyong, Zhang Liting, et al. Cultivated land change forecast based on genetic algorithm and least squares support vector machines[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(7): 226—231. (in Chinese with English abstract)
- [16] 王定成, 乔晓军, 汪春秀, 等. 基于作物响应的温室环境 SVMR 控制仿真[J]. 农业工程学报, 2010, 26(增刊 2): 290—293.
Wang Dingcheng, Qiao Xiaojun, Wang Chunxiu, et al. Simulation of greenhouse environment SVMR control based on plant response[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2010, 26(Supp.2): 290—293. (in Chinese with English abstract)
- [17] 周竹, 李小昱, 李培武, 等. 基于 GA-LSSVM 和近红外傅里叶变换的霉变板栗识别[J]. 农业工程学报, 2011, 27(3): 331—335.
Zhou Zhu, Li Xiaoyu, Li Peiwu, et al. Near-infrared spectral detection of moldy chestnut based on GA-LSSVM and FFT[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2011, 27(3): 331—335. (in Chinese with English abstract)
- [18] 唐晓旭, 刘平礼, 赵立强, 等. 基于 GA-SVM 的油田注水动态预测模型[J]. 实验研究, 2011, 30(6): 1—3.
- [19] 李云飞, 鄢飞, 张谦. 基于 GA-SVM 的上市公司财务危机预警研究[J]. 西安工程大学学报, 2010, 24(6): 822—826.
Li Yunfei, Yan Fei, Zhang Qian. Research on financial distress early-warning of listed companies based on GA-SVM[J]. Journal of Xi'an Polytechnic University, 2010, 24(6): 822—826. (in Chinese with English abstract)
- [20] 李艳. GA 和 SVM 在供应链绩效评价中的应用[J]. 计算机工程与应用, 2010, 46(1): 246—248.
Li Yan. Application of GA and SVM in performance evaluation of supply chain[J]. Computer Engineering and Applications, 2010, 46(1): 246—248. (in Chinese with English abstract)
- [21] 向先全, 陶建华. 基于 GA-SVM 的渤海湾富营养化模型[J]. 天津大学学报, 2011, 44(3): 215—220.
Xiang Xianquan, Tao Jianhua. Eutrophication model of bohai bay based on GA-SVM[J]. Journal of Tianjin University, 2011, 44(3): 215—220. (in Chinese with English abstract)
- [22] Wang Yanqiu, Chen Xiaowen, Jiang Wei, et al. Predicting human microRNA precursors based on an optimized feature subset generated by GA-SVM[J]. Genomics, 2011, 98(2): 73—78.
- [23] Eslam Pourbasheer, Siavash Riahi, Mohammad Reza Ganjali, et al. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity[J]. European Journal of Medicinal Chemistry, 2009, 44(12): 5023—5028.
- [24] 张稳, 张丽丽. 基于 GA-SVM 的银行客户流失预测分析[J]. 计算机与数字工程, 2010, 38(4): 55—58.
- [25] 宋芳. 基于 GA-SVM 模型的福建省城镇登记失业率预测[J]. 科技和产业, 2009, 30(6): 82—85.
Song Fang. The prediction of the urban registered unemployment rate in Fujian province based on GA-SVM model[J]. Science Technology and Industry, 2009, 30(6): 82—85. (in Chinese with English abstract)
- [26] Gu Jirong, Zhu Mingcang, Jiang Liuguangyan. Housing price forecasting based on genetic algorithm and support vector machine[J]. Expert Systems with Applications, 2011, 38(4): 3383—3386.
- [27] Lu Sheng, Cai Zhongjian, Zhang Xiaobin. Application of GA-SVM Time Series Prediction in Tax Forecasting[C]// 2009 2nd IEEE International Conference on Computer Science and Information Technology. Beijing: ICCSIT, 2009: 34—36.
- [28] Xu Chen. Urban water supply quantity forecasting by GA-SVM[C]// 2009 Proceedings of the 2nd International

- Conference on Advanced Computer Theory and Engineering. Cairo: ICACTE, 2009: 265—269.
- [29] 周娟, 杨鼎才. 基于 GA-SVM 的说话人辨认的参数优化[J]. 电子技术, 2008, 45(2): 54—56.
- Zhou Juan, Yang Dingcai. Parameter optimization of speaker identification based on GA-SVM[J]. Technology Research, 2008, 45(2): 54—56. (in Chinese with English abstract)
- [30] 汪西莉, 周兆永, 延军平. 应用 GA-SVM 的渭河水质参数多光谱遥感反演[J]. 遥感学报, 2009, 13(4): 740—744.
- [31] 吉根林. 遗传算法研究综述[J]. 计算机应用与软件, 2004, 21(2): 69—73.
- Ji Genlin. Survey on genetic algorithm[J]. Computer Applications and Software, 2004, 21(2): 69—73. (in Chinese with English abstract)

Calculation of theoretical and accessible yields of agricultural land based on genetic algorithm and support vector machine

Wang Haijun^{1,2}, Liu Minyan¹, Gao Juan¹

(1. School of Resources and Environment Science, Wuhan University, Wuhan 430079, China; 2. Department of Urban Planning and Design, The University of Hong Kong, Hong Kong 999077, China)

Abstract: Grain security is a complex, resource-intensive problem being addressed by governments, international organizations, and scientific community. Ensuring adequate grain supply is vital to the survival of humanity, and its key lies in improving total agricultural land productivity. The level of the comprehensive productive capacity of agriculture is directly related to the effective supply of grains. In 2011, China conducted a nationwide agricultural productivity survey in order to improve its overall capacity, protect farm quality, and accomplish the National Food Security Strategy. The core components of agricultural land productivity calculations are the theoretical and accessible yields. The theoretical yield calculation was traditionally performed by establishing samples' linear regression models between natural level index and theoretical yield. The accessible yield, meanwhile, was performed by establishing samples' linear regression models between usage level index and accessible yield. Then the agricultural land productivity can be calculated by the yield multiplying with the total area of classification units. However, these models suffer from significant calculative limitations. A novel approach based on the Genetic Algorithm and Support Vector Machine (GA-SVM) is proposed herein. The theoretical yield can be calculated by establishing a GA-SVM model between the theoretical yield and quality scores of gradation factors. The accessible yield can be calculated by establishing a GA-SVM model between the accessible yield and the product of usage factor and quality scores of gradation factors. The theoretical and accessible yield prediction models are built using Matlab software, taking full advantage of the GA and SVM toolbox, to test the rationality and accuracy of the new GA-SVM model. Jiexi County in Guangdong Province is provided as an example to verify the theoretical and accessible yields based on the GA-SVM models. The GA-SVM results as well as the traditional linear regression method results are then compared with the actual yields. The analysis reveals several interesting conclusions. First, GA-SVM requires less computational time to provide more accurate results. The linear regression models turn to cumbersome process and need many calculations. Some human errors occurred. However, the GA-SVM models acquire results by running programs. This method can save time and effort, also can avoid human errors. Second, the paper compared the results of GA-SVM model and the actual values, also compared the results of linear regression method and the actual values. The study found that the GA-SVM yield predictions are much closer to the actual values than those provided by the linear regression method. In addition, GA-SVM model is more suitable in forecasting a single sample value, which is more accurate than the linear regression method. Therefore, the GA-SVM model can forecast the theoretical and accessible yield better. Since the GA-SVM model is faster and more accurate, it should be used as a new method to calculate agricultural land productivity. Based on the theoretical and accessible yields calculated by GA-SVM model, the theoretical and accessible productivities can be enumerated with the theoretical and accessible yield multiplying with the total area of classification units, while the accuracy of agricultural land productivity can be guaranteed.

Key words: land use, productivity, genetic algorithms, support vector machines, theoretical yield, accessible yield, agricultural land

(责任编辑: 张俊芳)