

ров P , характеризующих ПА типа i : $f_{k,i}^1 = f(x_{1,i}, x_{2,i}, \dots, x_{q,i})$, где q – число параметров ПА типа i .

Предложенная модель может быть с успехом использована для описания и моделирования реальных самоорганизующихся систем, функцио-

нирование которых может быть выражено целевым функционалом. Кроме того, агенты, программно реализующие предложенную модель, могут быть автономными и территориально распределенными, повышая тем самым степень достоверности результатов симуляции модели.

СПИСОК ЛИТЕРАТУРЫ

1. Кузнецова В. Л., Раков М. А. Самоорганизация в технических системах. Киев: Наук. думка, 1987.
2. Иващенко Г. И., Зайченко Ю. П., Димитров В. Д. Принятие решений на основе самоорганизации. М.: Сов. радио, 1976.
3. Gorodetskii V. I. Self-organization and multiagent systems: I. Models of multiagent self-organization // J. of Computer and Systems Sciences Intern. 2012. Vol. 51, iss. 2. P. 256–281.
4. Gorodetskii V. I. Self-organization and multiagent systems: II. Applications and the development technology // J. of Computer and Systems Sciences Intern. 2012. Vol. 51, iss. 3. P. 391–409.

M. S. Kupriyanov, A. V. Kochetkov
Saint Petersburg Electrotechnical University «LETI»

THE MULTIAGENT MODEL OF SELF-ORGANIZING DISTRIBUTED SYSTEM

Discusses the issues relating to the current management problems in complex technical systems. The analysis of the existing solutions in this area on the basis of self-organization methods and algorithms is given. As a practical implementation of self-organizing methods the selection of the multi-agent approach is proven. Next, the article describes the proposed multiagent model for building self-organizing distributed system.

Self-organization, self-organizing system, multiagent system, system model

УДК 004.942

И Вэньлун, И. В. Герасимов, С. А. Кузьмин
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Хэ Хоцзяо, Ян Хунюнь
Аграрный университет провинции Цзянси (г. Наньчан, Китай)

Применение метода опорных векторов для задачи восстановления регрессии в обработке данных листьев поливного риса

Рассматривается метод опорных векторов, позволяющий сгруппировать собранные в ходе исследований данные моделирования листьев поливного риса в образцы обучающего набора и тестового набора. На их основе были сконструированы 3 модели машинного обучения для прогнозирования условий выращивания поливного риса на основе длины, ширины и значения SPAD листа.

Лист поливного риса, физиологическая экология, машинное обучение, метод опорных векторов

В листе поливного риса, являющегося одним из основных питательных органов поливного риса, происходит фотосинтез и испарение влаги, что влияет на урожайность риса и других его функций. Его рост является сложным и нелинейным процессом, на который влияют не только темпе-

ратура окружающей среды, наличие питательных веществ в почве, влажность и другие внешние условия выращивания, но и внутренние факторы, которые также взаимодействуют между собой.

В 60-х гг. XX в. начались исследования по разработке компьютерной модели для количе-

ственного анализа и прогнозирования площади листа агрокультуры, содержания в нем хлорофилла, урожайности агрокультуры и других показателей в различных условиях выращивания. Через 50 лет аналоговые модели выращивания агрокультуры получили быстрое развитие. Среди них есть ряд наиболее передовых моделей: Wit впервые разработал компьютерную модель фотосинтеза в листьях [1], Sinclair и другие ученые исследовали не только физиологию процесса роста в жизненном цикле агрокультуры, но и баланс питательных веществ в почве [2], Lemmon и Chuk с помощью компьютера проводили моделирование процесса роста хлопка [3], Wenlong спроектировал сельскохозяйственную экспертную систему, которая вырабатывает рекомендации для принятия решений в выращивании поливного риса [4].

Так как данные физиологической экологии листа поливного риса имеют высокую размерность, а потому требуют большого объема вычислений, в данной статье рассматривается метод опорных векторов для решения задачи восстановления регрессии и создания модели машинного обучения для регрессии внешних окружающих сред и данных физиологической экологии листа поливного риса с целью уменьшения сложности расчетов и повышения надежности модели регрессии.

Метод опорных векторов для задачи восстановления регрессии. Метод опорных векторов (Support Vector Machine, SVM) является популярным методом машинного обучения, который создан на основе теории статистического обучения, созданной русскими учеными В. Н. Вапником и др. Его основной идеей было то, что распределенные в исходном пространстве данные посредством специальных нелинейных функций отображаются в пространстве высокой размерности. В пространстве высокой размерности устанавливается оптимальная разделяющая гиперплоскость, описываемая следующей функцией: $f(x) = \omega^T \varphi(x) + b$, где ω – нормальный вектор разделяющей гиперплоскости; T – знак транспонирования вектора; b – координата высоты разделяющей гиперплоскости; φ – функция отображения входного вектора x из исходного пространства в пространство высокой размерности. При этом наиболее удаленные образцы различных классов легче отделяются друг от друга. Таким образом, лучше достигается принцип минимизации риска структуры. Также он хорошо показал уникальные преимущества, которые проявляются при рассмотрении небольшого ко-

личества образцов и нелинейности в машинном обучении [5]. В настоящее время этот метод широко применяется в классификации текстов, распознавании изображений, биоинженерии и т. д.

Метод опорных векторов для задачи восстановления регрессии (Support Vector Regression, SVR) был предложен на основе метода опорных векторов для стандартной задачи классификации. Элементарные образцы у SVR принадлежат к одной категории. Определяемая оптимальная гиперплоскость строится таким образом, чтобы все обучающие образцы имели минимальное общее расстояние друг от друга. Вапник ввел так называемую нечувствительную функцию потерь ε (insensitive loss), которая допускает погрешность в пределах какого-то значения пороговой величины. Если обучающие образцы находятся на расстоянии, не большем ε от оптимальной разделяющей гиперплоскости, то погрешность равна нулю. Погрешность элементарного образца вне этой зоны равна $|y - f(x)| - \varepsilon$ [6], [7], как показано на рис. 1, где $f(x)$ – прогнозируемые значения в соответствии с реальными значениями y в обучающем наборе образцов SVR, линия 1 – оптимальная разделяющая гиперплоскость: $f(x) = \omega^T \varphi(x) + b$, а две линии 2 – ε -границы гиперплоскости: $f(x) = \omega^T \varphi(x) + b \pm \varepsilon$. С учетом этого в пространстве высокой размерности некоторые обучающие образцы приводят, возможно, к линейно неразделимым результатам, в SVM и SVR использована дополнительная скользящая переменная (slack variable) ξ (см. приведенные далее формулы).

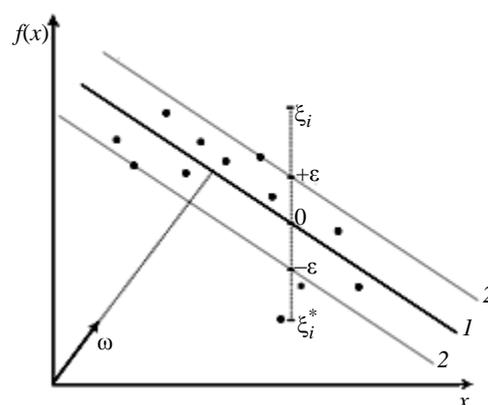


Рис. 1

Искомая оптимальная разделяющая гиперплоскость (separating hyperplane) является гиперплоскостью, до которой в пространстве высокой размерности совокупность расстояний от всех обучающих образцов минимальна, т. е. у задачи восстановления регрессии будет самая мини-

мальная суммарная ошибка. Тогда эта разделяющая гиперплоскость должна находиться в центре всех обучающих образцов. Согласно формуле расстояния двух параллельных плоскостей в пространственной аналитической геометрии и с точки зрения методов оптимизации это можно записать в виде следующей формулы [7]:

$$\begin{aligned} & \text{minimize}_{\omega, b, \xi, \xi^*} \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n \xi_i + c \sum_{i=1}^n \xi_i^* \\ & \text{subject to} \begin{cases} y_i - (\omega^T \varphi(x_i) + b) \leq \varepsilon + \xi_i, \\ -y_i + (\omega^T \varphi(x_i) + b) \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, n, \end{cases} \end{aligned} \quad (1)$$

где n – количество обучающих образцов; c – штрафной коэффициент; ξ_i – значение дополнительной скользящей переменной i -го образца в положительную сторону, а ξ_i^* – ее значение в отрицательную сторону. Через метод множителей Лагранжа проблема оптимизации с ограничением превращается в функцию Лагранжа без ограничения:

$$\begin{aligned} & \text{minimize } L(\omega, b, \xi_i, \xi_i^*, \alpha, \alpha^*, \eta, \eta^*) = \\ & = \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n \xi_i + c \sum_{i=1}^n \xi_i^* + \\ & + \sum_{i=1}^n \alpha_i (y_i - \omega^T \varphi(x_i) - b - \varepsilon - \xi_i) + \\ & + \sum_{i=1}^n \alpha_i^* (-y_i + \omega^T \varphi(x_i) + b - \varepsilon - \xi_i^*) - \\ & - \sum_{i=1}^n \eta_i \xi_i - \sum_{i=1}^n \eta_i^* \xi_i^*, \end{aligned} \quad (2)$$

где $\alpha, \alpha^*, \eta, \eta^* \geq 0$ – множители Лагранжа; $\varphi(x_i)$ – нелинейная функция отражения образцов из пространства x_i в пространство высокой размерности. С помощью $\omega, b, \xi_i, \xi_i^*$ нужно определить частный дифференциал функции Лагранжа, потом ввести результаты в (2) – и можно получить дуальную проблему формулы (1):

$$\begin{aligned} & \text{maximize}_{\alpha, \alpha^*} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)^T (\alpha_j - \alpha_j^*) \times \\ & \times \varphi(x_i)^T \varphi(x_j) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \\ & \text{subject to} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha, \alpha^* \leq c. \end{cases} \end{aligned}$$

Потом через условие Каруша–Куна–Таккера (ККТ) можно получить уравнение оптимальной разделяющей гиперплоскости

$$y = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \kappa(x_i, x) + b,$$

где $\kappa(x_i, x) = \varphi(x_i)^T \varphi(x)$ – ядерная функция. Это отображение скалярного произведения образцов первичного пространства в соответствующем пространстве высокой размерности, что не требует поиска нелинейной функции отображения φ (тем самым достигается упрощение вычислений) [7]. Кроме того, формула оценки среднеквадратической ошибки (mean square error):

$$\text{MSE} = \frac{\sum_{i=1}^n (f(x_i) - y_i)^2}{n}, \quad \text{формула коэффициента корреляции (correlation coefficient):}$$

$$R^2 = \frac{\left(n \sum_{i=1}^n f(x_i) y_i - \sum_{i=1}^n f(x_i) \sum_{i=1}^n y_i \right)^2}{\left(n \sum_{i=1}^n f(x_i)^2 - \left(\sum_{i=1}^n f(x_i) \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)},$$

где n – количество образцов обучающего или же тестового набора; $i \in [1, 2, \dots, n]$ – номер образца, $f(x_i)$ – прогнозируемое значение образца; y_i – соответствующее ему реальное значение образца [7].

Применение моделирования листьев поливного риса в задаче обработки данных.

Метод опорных векторов. В описываемой работе была использована библиотека libsvm¹, разработанная Линь Чжижэном, которая реализует функции классификации образцов и регрессии у SVM [8]. Процесс восстановления регрессии показан на рис. 2.

¹ A Library for Support Vector Machines. URL: <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.



Рис. 2

В процессе создания модели обучения использована kern-функция Гаусса: $k(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\delta^2}\right)$.

Источник данных в эксперименте. Ведущая лаборатория аграрных информационных технологий при высших учебных заведениях провинции Цзянси (Китай) на своем опытном полигоне, находящемся в 28°46' северной широты, 115°55' восточной долготы и 48 м над уровнем моря, проводила эксперимент по выращиванию поливного риса сорта «Цзинью». После снятия результатов измерений и фильтрации данных окончательно было получено 198 образцов. Формат записей: дата сбора, количество дней роста, количество часов солнечного освещения, уровень внесенных удобрений, расположение листьев, максимальная длина жилки листа, максимальная ширина жилки листа, значение SPAD листа. Используя метод опорных векторов, случайным образом 198 собранных образцов были разделены на набор из 158 обучающих образцов и набор из 40 тестовых образцов. Потом к ним были присоединены данные внешних факторов выращивания риса сорта «Цзинью» в качестве многомерного аргумента (дата сбора, количество дней роста, количество часов солнечного освещения, уровень внесенных удобрений и расположение листьев). А затем к каждому набору была по очереди при-

менена тройка данных – максимальная длина жилки листа, максимальная ширина жилки листа и значение SPAD листа. Каждый из параметров выступал в качестве зависимой переменной для образования, соответственно, трех моделей прогнозирования. Полученная после машинного обучения имитационная модель показана на рис. 3.

Внешние факторы роста риса – это совокупность таких данных выращивания поливного риса сорта «Цзинью», как дата сбора, количество дней роста и т. д., представленная в виде вектора значений. На рис. 3, а показана диаграмма имитационной модели на основе максимальной длины жилки листа для обучающего набора, а на рис. 3, б – для тестового набора. На рис. 3, в приведена диаграмма имитационной модели на основе максимальной ширины жилки листа для обучающего набора, а на рис. 3, г – для тестового набора. На рис. 3, д и е изображены диаграммы имитационной модели на основе значения SPAD листа для обучающей и тестовой сборки соответственно.

Регулировка штрафного коэффициента c , параметра kern-функции Гаусса σ и порогового значения погрешности ϵ существенно влияет на среднеквадратическую ошибку и коэффициент корреляции обучающей модели регрессии (таблица). После проведения многократных экспериментов были получены более качественные результаты обучения.

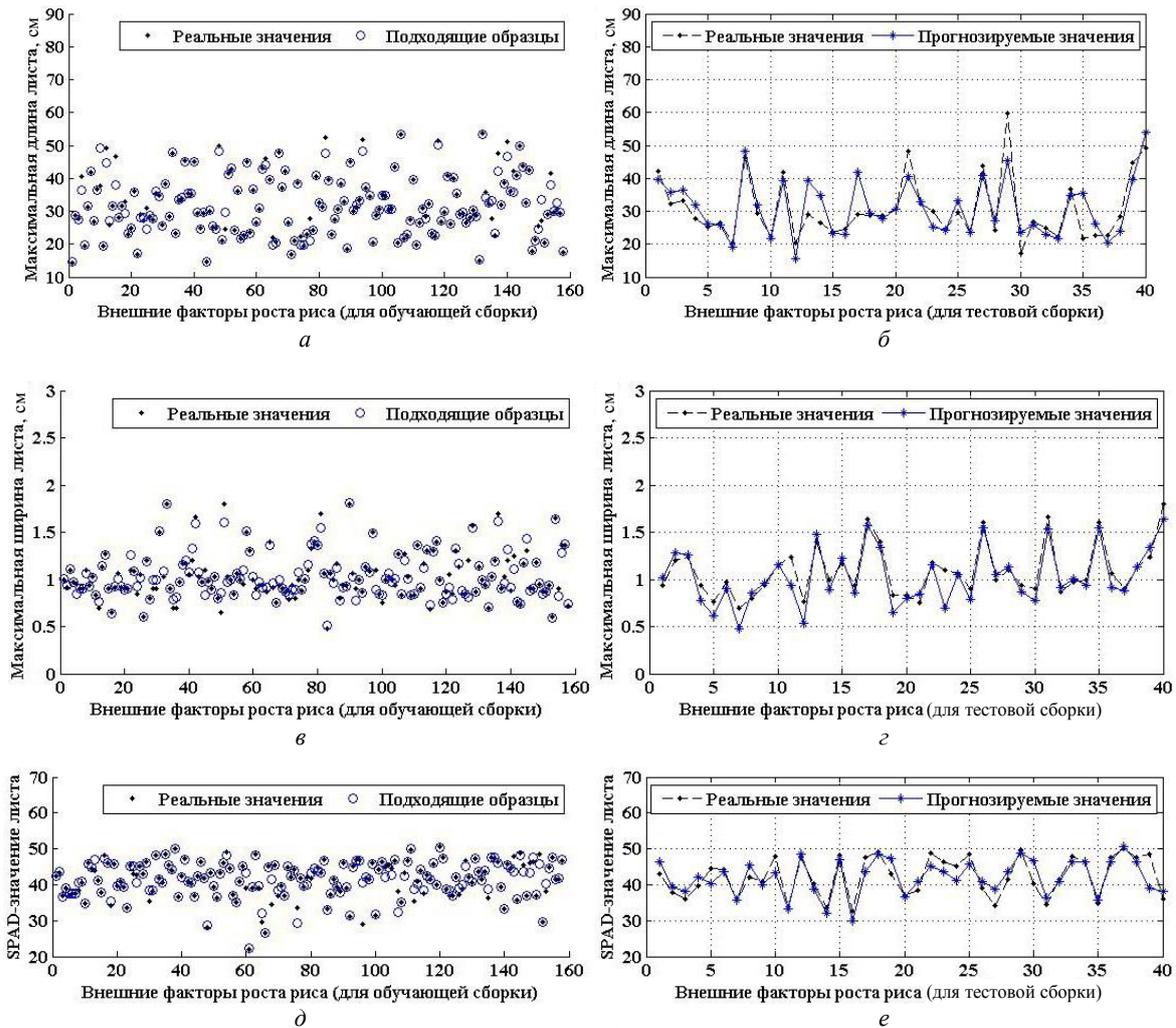


Рис. 3

Объект	Параметры SVG			Результаты обучения образца		Результаты тестирования модели	
	c	σ	ϵ	MSE	$R^2, \%$	MSE	$R^2, \%$
Максимальная длина жилки листа	2.8284	4	0.01	0.010112	95.351	0.068869	70.588
Максимальная ширина жилки листа	32	0.35355	0.01	0.01022	92.667	0.036078	85.521
SPAD-значение листа	1.4142	4	0.01	0.012312	90.601	0.042468	71.59

В данной статье предпринята попытка использовать метод опорных векторов в обработке данных физиологической экологии листьев поливного риса. Полученные экспериментальные результаты свидетельствуют о том, что метод опорных векторов способен обрабатывать сложные модели нелинейной связи высокой размерности, обеспечивая тем самым поддержку фундаментальных моделей данных для визуализации листьев поливного риса в будущем. Благодаря тому, что он отображает данные из пространства меньшей размерности в пространство большей размерности, мож-

но хорошо обрабатывать сложную нелинейную модель прогноза роста агрокультуры с высокой размерностью. Будущие исследования следует проводить по двум приоритетным научным направлениям: визуализация физиологической экологии листьев риса, более точное установление параметров SVG посредством их оптимизации для дальнейшего повышения точности и надежности модели машинного обучения.

Проект создан при поддержке государственного фонда естественных наук Китая (грант № 61363041, 61562039).

СПИСОК ЛИТЕРАТУРЫ

1. Wit C. T. Photosynthesis of leaf canopies. Wageningen, Netherlands: Centre for Agricultural Publications and Documentation, 1965. 57 p.
2. Sinclair T. R., Seligman N. G. Crop modeling: from infancy to maturity // *Agronomy J.* 1996. № 88(5). P. 698–704.
3. Lemmon H., Chuk И. Object-oriented design of a cotton crop model // *Ecological Modelling.* 1997. № 94(1). P. 45–51.
4. Вэньлун И. Проведение научных исследований по проекту «Разработка веб-ориентированной экспертной системы по выращиванию гибридного риса // 68-я Науч.-техн. конф. профессорско-преподавательского состава СПбГЭТУ «ЛЭТИ»: сб. докл. студентов, аспирантов и молодых ученых. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2015. С. 122–125.
5. Vapnik V. N. *Statistical Learning Theory. Adaptive and Learning Systems for Signal Processing, Communications and Control.* New York, USA: Wiley, 1998. 768 p.
6. Vapnik V. N. *The Nature of Statistical Learning Theory. Statistics for Engineering and Information Science.* 2nd edition. New York, USA: Springer, 2000. 314 p.
7. Scholkopf B., Smola A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond.* Cambridge, Massachusetts, USA: The MIT Press, 2001. 626 p.
8. Chang C. C., Lin C. J. LIBSVM: A Library for Support Vector Machines // *ACM Transactions on Intelligent Systems and Technology.* 2011. Vol. 2, iss. 3. 27 p. URL: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>.

Yi Wenlong, I. V. Gerasimov, S. A. Kuzmin
Saint Petersburg Electrotechnical University «LETI»
He Huojiao, Yang Hongyun
Jiangxi Agricultural University (Nanchang, China)

APPLICATION OF SUPPORT VECTOR REGRESSION IN MODELING AND DATA PROCESSING OF RICE LEAVES

By applying support vector regression, the modeling data of rice leaves collected in our study were grouped into sample training set and test set, and three machine learning prediction models on rice growing environment against leaf blade length, width and SPAD value were constructed.

Rice leaf, physiological ecology, machine learning, support vector machine

УДК 519.7

Н. А. Перязев,
Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Ю. В. Перязева,
Гимназия № 24 (Санкт-Петербург)

И. К. Шаранхаев
Бурятский государственный университет

Минимальные алгебры унарных мультиопераций

Получено матричное представление алгебр унарных мультиопераций конечного ранга и список тождеств, выполняемых в таких алгебрах. Эти результаты используются для доказательства основного результата: описания минимальных алгебр унарных мультиопераций конечного ранга. Как следствие этой теоремы получен список всех таких минимальных алгебр для небольших рангов.

Мультиоперация, алгебра, минимальная алгебра, матрица, операция, подстановка

Рассматриваемые в статье алгебры унарных мультиопераций являются конечными алгебрами. Для изучения их строения большое значение имеет описание минимальных алгебр [1]. В работе [2] получено описание всех алгебр унарных мультиопераций ранга 3. Основной результат данной работы анонсирован в [3]. Отметим, что алгебры унарных

мультиопераций находят применение при изучении суперклонов, а значит, и клонов [4].

Пусть $B(A)$ – множество всех подмножеств множества A . Отображение из A в $B(A)$ называется унарной мультиоперацией на A . Использо-