

УДК 004.5

MSC 68T20

DOI: 10.35330/1991-6639-2020-6-98-91-96

## ПРИМЕНЕНИЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СЕЛЕКЦИИ\*

М.И. АНЧЁКОВ, З.И. БОГОТОВА

ФГБНУ «Федеральный научный центр  
«Кабардино-Балкарский научный центр Российской академии наук»  
360002, КБР, г. Нальчик, ул. Балкарова, 2  
E-mail: kbncran@mail.ru

*В работе проведен обзор методов искусственного интеллекта, применяемых в селекции. Рассмотрены работы, в которых сравнивались классические статистические методы и методы на основе искусственного интеллекта. Выявлены основные проблемы, препятствующие внедрению методов на основе искусственного интеллекта, и предложены пути их решения.*

**Ключевые слова:** искусственный интеллект, машинное обучение, селекция, генетика.

### ВВЕДЕНИЕ

В условиях постоянного роста численности населения задача повышения эффективности производства сельхозпродукции актуализируется все сильнее. Одним из способов решения данной проблемы является выведение новых сортов, линий и пород, которые обладают высокими конкурентными характеристиками. Однако процесс селекции может занимать достаточно долгое время. Например, на создание новой линии кукурузы уходит порядка 7-10 лет. Такие временные затраты совершенно не отвечают реалиям высококонкурентного рынка.

Целью данной работы является обзор применения методов искусственного интеллекта в селекции растений и животных.

### МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СЕЛЕКЦИИ

В работах [1, 2] авторы отмечают, что для удовлетворения растущего спроса на продукты питания необходимо уменьшить время разработки новых сортов (линий, пород). Однако для проведения качественного анализа необходимо собрать и обработать большое количество данных. В данный момент наибольшее распространение получили методы геномного отбора (Genomic Selection, GS) и методы геномной оценки племенной ценности (Genomic Estimated Breeding Value, GEBV). В указанных работах рассматривается задача оптимизации обучающей выборки с помощью генетического алгоритма [3]. Авторы указывают на тот факт, что процедура фенотипирования является дорогостоящей в отличие от процедуры генотипирования. Последняя операция дает очень много данных, которые трудно связать с хозяйственно полезными признаками. Предлагаемые методы основываются только на данных, полученных с помощью генотипирования. Информация о фенотипе используется только для проверки метода. В работе [2] указывается, что классический генетический алгоритм имеет свойство «застрять» в локальных точках экстремума, и для решения этой проблемы авторы предлагают использовать оригинальный генетический алгоритм – адаптивный генетический алгоритм (Adaptive Genetic Algorithm, AGA). Пред-

---

\* Работа выполнена при финансовой поддержке грантов РФФИ №№ 18-01-00658, 19-01-00648

ложенный генетический алгоритм подстраивает вероятности использования операторов кроссинговера и мутации на каждой итерации.

В работе [4] отмечено, что в случае, когда необходимо «улучшать» одновременно несколько хозяйственно полезных признаков, использование индекса отбора дает лучшие результаты по сравнению с прямым отбором. Авторы предлагают использовать генетический алгоритм для оптимизации процесса отбора по нескольким признакам. Функция пригодности (fitness function) генетического алгоритма рассчитывается как линейная комбинация выбранных признаков, которые оцениваются по определенной шкале. Предложенный метод показывает сопоставимые с другими методами результаты, но выполняется за более короткое время и позволяет исключить человеческий фактор.

В работе [5] проводят сравнение таких методов машинного обучения, как эластичная сеть (elastic net), метод регуляризации Тихонова (ridge regression), метод регрессии «лассо» (lasso regression), случайный лес (random forest), градиентный бустинг (gradient boosting machines, GBM), машина опорных векторов (support vector machine, SVM) с двумя современными методами классической статистической генетики: наилучший линейный несмещенный прогноз (best linear unbiased prediction, BLUP) и двухступенчатый регрессионный анализ. Исследование проводилось на трех культурах: дрожжи, рис и пшеница. При этом первые две имеют полностью секвенированные геномы, а третья – частично. В работе отмечено, что подобные задачи имеют отличительную черту: количество характеристик, по которым принимается решение, значительно превышающее количество элементов в выборке. Например, входные данные по дрожжам были представлены информацией о 1008 гаплоидных штаммах дрожжей, а данные по каждому штамму были представлены 11 623 маркерами. Авторы указывают, что методы машинного обучения менее требовательны к пониманию процессов, происходящих в предметной области. Этот факт является как положительным качеством методов на основе глубокого обучения, так и отрицательным. Авторы отмечают, что существует необходимость разработки методов, которые учитывают специфику таких процессов, как мейоз, доминантность, эпистаз и т.д.

В работе [6] авторы применяют машинное обучение для решения задачи предсказания урожайности гибридов. Предложенный метод основывается на искусственных нейронных сетях и методе матричной факторизации. Отличительной чертой подхода является то, что авторы, кроме генетических данных, используют еще и данные о месте проведения испытаний. Применяемый метод показал результаты, которые выше, чем классические методы статистического анализа, применяемые в настоящее время селекционерами. Хочется отметить, что исследование проводилось на данных, которые предоставляла фирма Syngenta, в рамках Syngenta Crop Challenge In Analytics – конкурса, в котором участники должны были разработать метод прогноза времени посадки конкретных семян кукурузы, – с целью оптимизации процесса уборки урожая.

В [7] авторы исследуют связь между генотипом и фенотипом посредством искусственных нейронных сетей на примере коров породы Джерси и пшеницы. Информация была представлена по 297 коровам и для каждой были известны 35 798 маркеров и фенотипические признаки, такие как надой молока, содержание белка и жира в молоке и т.д. Пшеница была представлена 599 линиями, каждая из которых имеет 1279 маркеров и информацию о средней урожайности. Авторы анализируют влияние количества нейронов в скрытом слое на качество полученных решений. Из полученных результатов сделаны выводы, что, варьируя количество нейронов в скрытом слое, можно бороться с *проклятием размерности*, что очень актуально для задач с большой размерностью входных данных. Авторы доказывают гипотезу о том, что нелинейные функции активации улучшают качество

сформулированного прогноза, но данный результат без существенного превышения стандартных методов. Такая осторожность в выводах связана с тем, что проведенное исследование апробировалось пока на небольшом количестве данных.

Применение методов молекулярной селекции и генетики должно ускорить создание новых сортов и гибридов, а применение методов искусственного интеллекта может оптимизировать этот процесс. Например, трудозатратной является процедура удаления метелок у материнской формы, которую, как правило, проводят вручную. Автоматизация и роботизация этого процесса связана с определенными трудностями: применение высокоспециализированной техники, низкий уровень автоматизации и т.д.

В работе [8] используют предобработку данных с использованием суперпиксельной кластеризации, которая заключается в разбиении исходных изображений метелок риса на суперпиксели (фрагменты), которые представляют собой множество относительно однородных расположенных рядом пикселей. После этого всю выборку разбивают на обучающую и тестовую. Обучающую часть используют для обучения конволюционной (сверточной) нейронной сети. Предложенный авторами подход превосходит аналогичные методы по точности.

В [9] авторы указывают на то, что на данный момент мониторинг процесса вегетации производят непосредственно люди. Кроме того, что это очень трудоемкий процесс, результаты могут в значительной степени носить субъективный характер. В работе предлагают использовать систему машинного зрения для определения характеристик метелок кукурузы, которая получает изображения со стационарных камер, установленных на мачте. Авторы собрали базу данных, которая состоит из 16031 изображения, полученного в течение 4 сезонов. В качестве обучающей выборки было использовано 10164 изображения, остальное – тестовая выборка. Предложенный авторами метод основывается на том, что классификация происходит в несколько этапов:

- 1) определение области, в которой находится метелка;
- 2) определение на заданной области метелки и заднего фона;
- 3) определение формы метелки;
- 4) определение характеристик метелки (длина, ширина, цвет, количество ветвей).

В работе указывается, что предложенный метод в значительной степени превосходит существующие аналоги. Следует заметить, что авторы провели достаточно большую работу для получения обучающей выборки, которая заключалась в ручной разметке изображений в графическом редакторе. Это можно отнести к недостаткам метода.

На начальном этапе селекционной работы первоочередной и очень важной является задача определения структуры исходного материала. Селекционер должен выбрать перспективные образцы, на основе которых будут создаваться сорта и линии. В работе [10] для решения задачи классификации исходного материала пшеницы использовалась сеть Кохонена [11]. Авторами из 36 признаков было выбрано 6 основных: зимостойкость, урожайность, масса зерна с колоса, содержание клейковины, масса 1000 зерен, число зерен с колоса. Эти данные были получены по 249 образцам за три года исследований по сортоиспытанию озимой пшеницы в условиях лесостепи ЦЧР. В результате исследований авторам удалось провести гено-средовую корреляцию и установить влияние условий вегетации в разные годы на изменчивость признаков. К недостаткам метода можно отнести то, что сеть Кохонена основана на линейном сумматоре, который плохо работает с существенно нелинейными задачами.

Одним из перспективных направлений генетики является феномика, которая занимается выявлением закономерностей формирования и изменения фенотипов [12]. В [13] авто-

ры разработали программный продукт, который не только мог фенотипировать декоративные растения, но и позволял оценивать их физиологическое состояние. Комбинируя на разных этапах метод опорных векторов и сверточную нейронную сеть MobileNet, авторы добились точности порядка 96 процентов.

#### ЗАКЛЮЧЕНИЕ

Проведенный анализ российских и зарубежных публикаций показал, что методы искусственного интеллекта находят применение в задачах селекции.

Это связано с тем, что системы на основе искусственного интеллекта могут обрабатывать большие объемы информации за короткое время и находить нетривиальные зависимости. Однако пока рано говорить о том, что искусственный интеллект получил широкое распространение в селекции, и для данного факта есть несколько причин:

- отсутствие специалистов, которые достаточно хорошо одновременно понимают методы селекции и методы искусственного интеллекта;
- сложность получения обучающей выборки;
- отсутствие методов искусственного интеллекта, которые могут «понимать» специфику вегетации и погодных условий;
- сложность и дороговизна проверки гипотез.

Для решения этих проблем необходимо:

1. Автоматизировать и/или роботизировать процесс получения лабораторных и полевых данных.
2. Разработать организационную и техническую систему сбора и хранения полученных данных.
3. Провести исследования на уже полученных данных по селекции и семеноводству.

#### ЛИТЕРАТУРА

1. Akdemir D., Sanchez J.I., Jannink J.-L. Optimization of genomic selection training populations with a genetic algorithm // *Genet Sel Evol.* 2015. V. 47. № 1.
2. Purbarani S. C., Wasito I., Kusuma I. Adaptive genetic algorithm for reliable training population in plant breeding genomic selection // 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS): IEEE, 2016.
3. Holland J.H. Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence. London: Bradford book edition, 1994. 211 p.
4. Azimzadeh M. etc. Computer aided selection in breeding programs using genetic algorithm in MATLAB program // *Span J Agric Res.* 2010. T. 8. № 3. P. 672.
5. Grinberg N.F., Orhobor O.I., King R.D. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat // *Mach Learn.* 2019. V. 109. № 2. Pp. 251-277.
6. Khaki S., Khalilzadeh Z., Wang L. Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach // *PLoS ONE.* 2020. V. 15. № 5.
7. Gianola D. etc. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat // *BMC Genet.* 2011. V. 12. № 1. P. 87.
8. Xiong X. etc. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization // *Plant Methods.* Springer Science and Business Media LLC, 2017. V. 13. № 1.
9. Lu H. etc. Fine-grained maize tassel trait characterization with multi-view representations // *Computers and Electronics in Agriculture.* Elsevier BV. 2015. V. 118. Pp. 143-158.

10. Русанов И.А., Павлюк Н.Т., Ващенко Т.Г., Голева Г.Г. Нейронная сеть как способ классификации исходного материала озимой пшеницы // Вестник Воронежского государственного аграрного университета. 2010. № 3. С. 27-31.
11. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика: пер. с англ. Ю.А. Зуева, В.А. Точенова. 1992. 184 с.
12. Gerlai R. Phenomics: fiction or the future? // Trends in Neurosciences. Elsevier BV. 2002. V. 25. № 10. P. 506-509.
13. Бондаренко В.Ю. и др. Анализ фенотипа декоративных растений с использованием искусственных нейронных сетей: определение таксономических и физиологических характеристик // Журнал Белорусского государственного университета. Биология. № 1. 25-32.

## REFERENCES

1. Akdemir D., Sanchez J.I., Jannink J.-L. Optimization of genomic selection training populations with a genetic algorithm // Genet Sel Evol. 2015. V. 47. № 1.
2. Purbarani S. C., Wasito I., Kusuma I. Adaptive genetic algorithm for reliable training population in plant breeding genomic selection // 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS): IEEE, 2016.
3. Holland J.H. Adaptation in natural and artificial systems. An introductory analysis with application to biology, control, and artificial intelligence. London: Bradford book edition, 1994. 211 p.
4. Azimzadeh M. etc. Computer aided selection in breeding programs using genetic algorithm in MATLAB program // Span J Agric Res. 2010. V. 8. № 3. P. 672.
5. Grinberg N. F., Orhobor O. I., King R. D. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat // Mach Learn. 2019. V. 109. № 2. Pp. 251-277.
6. Khaki S., Khalilzadeh Z., Wang L. Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach // PLoS ONE. 2020. V. 15. № 5.
7. Gianola D. etc. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat // BMC Genet. 2011. V. 12. № 1. P. 87.
8. Xiong X. etc. Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization // Plant Methods. Springer Science and Business Media LLC. 2017. V. 13. № 1.
9. Lu H. etc. Fine-grained maize tassel trait characterization with multi-view representations // Computers and Electronics in Agriculture. Elsevier BV. 2015. V. 118. Pp. 143-158.
10. Rusanov I.A., Pavlyuk N.T., Vaschenko T.G., Goleva G.G. *Neyronnaya set' kak sposob klassifikatsii iskhodnogo materiala ozimoy pshenitsy* [Neural network as a way to classify the source material of winter wheat] // Bulletin of the Voronezh State Agrarian University. 2010. No. 3. Pp. 27-31.
11. Wasserman F. *Neyrokomp'yuternaya tekhnika: teoriya i praktika: per. s angl. YU.A. Zuyeva, V.A. Tochenova* [Neurocomputer technology: theory and practice: translated from English by Yu. A. Zuev, V. A. Tochenov]. 1992. 184 p.
12. Gerlai R. Phenomics: fiction or the future? // Trends in Neurosciences. Elsevier BV. 2002. V. 25. № 10. Pp. 506-509.
13. Bondarenko V.Yu. *Analiz fenotipa dekorativnykh rasteniy s ispol'zovaniyem iskusstvennykh neyronnykh setey: Opredeleniye taksonomicheskikh i fiziologicheskikh kharakteristik* [Analysis of the phenotype of ornamental plants using artificial neural networks: Determination of taxonomic and physiological characteristics] // Journal of the Belarusian State University. Biology. No. 1. Pp. 25-32.

## ARTIFICIAL INTELLIGENCE TECHNIQUES IN BREEDING\*

**M.I. ANCHEKOV, Z.I. BOGOTOVA**

FSBSE «Federal scientific center  
«Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences»  
360002, KBR, Nalchik, 2, Balkarova street  
E-mail: kbncran@mail.ru

*This paper reviews the artificial intelligence methods used in breeding. The papers in which the classical statistical methods and methods based on artificial intelligence were compared are considered. The main problems that hinder the introduction of methods based on artificial intelligence are identified and the ways to solve them are proposed.*

**Keyword:** artificial intelligence, machine learning, selection, genetics.

*Работа поступила 09.12.2020 г.*

### Сведения об авторах:

**Анчѳков Мурат Инусович**, н.с. лаборатории «Молекулярная селекция и биотехнология» Кабардино-Балкарского научного центра РАН.  
360000, КБР, г. Нальчик, ул. Кирова, 224.  
E-mail: murat.antchok@gmail.com

**Боготова Залина Ихсановна**, к.б.н, зав. лабораторией «Молекулярная селекция и биотехнология» Кабардино-Балкарского научного центра РАН.  
360000, КБР, г. Нальчик, ул. Кирова, 224.  
E-mail: zalina\_bogotova@mail.ru

### Information about the authors:

**Anchekov Murat Inusovich**, researcher of the laboratory "Molecular selection and biotechnology" of KBSC of the Russian Academy of Sciences.

360000, KBR, Nalchik, Kirov street, 224.  
E-mail: murat.antchok@gmail.com

**Bogotova Zalina Ikhsanovna**, Candidate of Biological Sciences, Head of the laboratory "Molecular selection and biotechnology" of KBSC of the Russian Academy of Sciences.

360000, KBR, Nalchik, Kirov street, 224.  
E-mail: zalina\_bogotova@mail.ru

---

\* The work was carried out with the financial support of the RFBR grants No. No. 18-01-00658, 19-01-00648