

МАШИННОЕ ОБУЧЕНИЕ В КАЧЕСТВЕ ИНСТРУМЕНТА ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР

П.К. Куценогий¹, кандидат физико-математических наук,
В.К. Каличкин¹, доктор сельскохозяйственных наук,
А.Л. Пакуль², научный сотрудник, С.П. Куценогий¹, младший научный сотрудник

¹Сибирский федеральный научный центр агробиотехнологий Российской академии наук,
633501, Новосибирская обл., р.п. Краснообск, ул. Центральная
E-mail: peter@kutsenogiy.ru

²Кемеровский научно-исследовательский институт сельского хозяйства –
филиал Сибирского федерального научного центра агробиотехнологий Российской академии наук
650510, Кемеровская обл., п. Новостройка, ул. Центральная, 47

Исследования проводили с целью определения возможности использования машинного обучения для оценки комплекса воздействия погоды и агротехнических факторов на урожайность сельскохозяйственных культур для ее прогнозирования. Численные эксперименты осуществляли на материалах длительных полевых опытов, выполненных в лесостепной зоне Кемеровской области. Для тренировки модели использовали данные непрерывных наблюдений за 2013-2018 гг. по пшенице и ячменю. Для расчётов использовали алгоритм машинного обучения Random Forest Classifier. Показатель accuracy определялся как отношение числа правильных прогнозов на тестовой выборке к общему числу тестовых примеров. В случае тренировки модели с использованием в качестве входных данных информации по текущим агроприемам и метеословения предыдущего года (среднемесячные температуры и осадки) показатель accuracy для пшеницы был равен 0,81, для ячменя – 0,87, в среднем по культурам – 0,84. Для изучения вопроса о вкладе информации, содержащейся в данных о колебании среднемесячных температур за предыдущий год, и степени влияния агрономических мероприятий текущего года на точность прогноза урожайности были проведены две альтернативные тренировки модели с различными входными данными. В одном случае учитывали только образ погоды предыдущего года на основе среднемесячных значений температуры и осадков. Во втором, учитывали используемые агроприемы, а данные по погоде уменьшили до одного значения – среднегодовой температуры. Величина показателя accuracy для культуры (без различия ячмень или пшеница) при учете только погодных факторов предыдущего года была равна 0,7, в случае учета преимущественно агроприемов с минимальным учетом погодных факторов – 0,73. По результатам исследований сделан вывод о сравнимом вкладе каждой из групп учитываемых факторов (погода за предыдущий период и планируемые агроприемы) в ожидаемую точность прогноза урожайности культур.

MACHINE LEARNING AS A TOOL FOR CROP YIELD FORECAST

Kutsenogiy P.K.¹, Kalichkin V.K.¹, Pakul A.L.², Kutsenogiy S.P.¹

¹Siberian Federal Scientific Center for Agro-BioTechnologies of the Russian Academy of Sciences
633501, Novosibirskaya obl., r.p. Krasnoobsk, ul. Tsentral'naya
E-mail: peter@kutsenogiy.ru

²Kemerovo Research Institute of Agriculture –
Branch of the Siberian Federal Scientific Center for Agro-BioTechnologies of the Russian Academy of Sciences,
650510, Kemerovskaya obl., p. Novostroika, ul. Tsentral'naya, 47

The possibilities of using of machine learning for estimation of the influence of the complex of weather and agrotechnical factors on the yield of agricultural crops have been investigated in order to predict it. Numerical experiments were carried out on the materials of longterm field experiments of the Kemerovo Research Institute of Agriculture, a branch of the SFSCA RAS, located in the foreststeppe zone of the Kemerovo region. To train the model, we used continuous observation data for the years 2013-2018 for the main crops: wheat and barley. The Random Forest Classifier machine learning algorithm was used for calculations. The "accuracy" indicator was defined as the ratio of the number of correct predictions on the test sample to the total number of test cases. In the case of training the model using information on current agricultural practices and weather conditions of the previous year as input data using average monthly temperatures and precipitation, the "accuracy" indicator for wheat was 0.81, for barley - 0.87, on average for crops - 0.84. It was studied, how much influence on the accuracy of the yield forecast is contained in the weather data of previous year and how important is accounting of the agronomic factors. For this, two alternative training of the model was carried out with modified input data. In one case, the applied agricultural practices were not taken into account, but only the weather image of the previous year was taken into account based on the monthly average values of temperature and precipitation. In the second case, the weather data was reduced to 1 value - the average annual temperature. The value of the accuracy indicator for the crop (without distinction between barley or wheat) in the case of taking into account only weather factors of the previous year, was 0.7. In the case of accounting mainly for agricultural practices, with minimal consideration of weather factors, this indicator was 0.73. From which it was concluded that each of the groups of factors considered (weather for the previous period and planned agricultural practices) made a comparable contribution to the expected accuracy of the yield forecast.

Ключевые слова: машинное обучение, прогнозирование, урожайность культур

Key words: machine learning, forecast, crop yield

Одна из наиболее актуальных задач в области оптимизации и интенсификации сельскохозяйственного производства – прогнозирование урожайности выращиваемых культур [1-3]. Адекватная оценка природных условий в предстоящем году может оказать значительное влияние на принятие управленческих решений в части выбора возделываемых культур, их размещения в севооборотах, применяемых агротехнических приемов и др., а также на долгосрочное планирование использования

земельных ресурсов [4-7]. Все это – необходимые элементы интеллектуальных экспертных систем поддержки принятия решений сельхозпроизводителя [8-12].

На сегодняшний день подходы к прогнозированию сельскохозяйственной деятельности во многом строятся по принципу экономического моделирования, которое, в частности, сводится к «экономическому» прогнозированию, то есть научному предвидению направлений развития экономики и отдельных ее элементов или поиску

оптимальных способов достижения поставленных целей [13-15]. Такой подход служит признанием невозможности учета всех факторов, влияющих на достижение ожидаемого результата в будущем, и сопровождается построением «сценарных» моделей, из которых предлагается выбирать по мере фактического развития ситуации.

Прогноз благоприятных условий для возделывания той или иной культуры на предстоящий сезон невозможен без корректного учета погодных факторов, влияющих на развитие биологических объектов, к которым относятся сельскохозяйственные растения. С другой стороны, очевидна невозможность точного прогноза погоды за горизонтом в несколько суток. Отсюда возникает необходимость планирования в условиях неопределенности существенных факторов и граничных условий. Поскольку описать аналитически все имеющиеся природные взаимосвязи и влияние одних факторов на другие невозможно, в последние годы все шире распространяются идеи использования возможностей искусственного интеллекта (AI) и машинного обучения (ML) применительно к прогнозу поведения сложных, в том числе биологических систем на основе анализа больших массивов данных [16-19].

Еще одно важное достоинство ML-подхода – более полный учет многофакторности воздействия на результат. Так, многолетние стационарные опыты служат для выявления факторов, влияющих на урожайность культур в тех или иных климатических зонах. Однако каждый последующий эксперимент проводится в новых внешних (граничных) условиях, так как не существует двух одинаковых лет из-за флуктуации агрометеорологических ресурсов. Возникает вопрос: каким образом учитывать результаты многолетних экспериментов по возделыванию культур и использованию различных агроприемов с тем, чтобы создавать адекватные рекомендации по агротехнологиям на предстоящий год, исходя из накопленного опыта и мониторинга погодной ситуации за предыдущий период.

Критическая особенность метода ML – необходимость достаточного числа обучающих примеров, сформированных на основе качественных данных наблюдений за сложной системой – объектом изучения. Точность обучения интеллектуальной системы напрямую зависит, как от объема имеющейся в распоряжении информации, так и от ее достоверности.

Как было показано в работе [20] достоверность данных общедоступной сельскохозяйственной статистики вызывает сомнения. В отрасли наблюдается

движение в поддержку открытых систем, но его оппоненты утверждают, что данные имеют стоимость и должны рассматриваться как ценные активы. Другими словами, если сельхозпроизводители не будут хранить свои данные в тайне, они могут потерять потенциал их ценности для использования на рынке [21].

Цель исследований – изучить возможности машинного обучения для прогнозирования урожайности сельскохозяйственных культур.

Методика. Исследования проводили на материалах длительных полевых опытов, проведенных в лесостепной зоне Кемеровской области. Для тренировки модели использовали данные непрерывных наблюдений за 6 лет (2013–2018 гг.) по двум основным культурам: пшенице и ячменю. Входные данные представляли собой набор используемых агроприемов и результаты возделывания в виде средней урожайности культуры в центнерах на 1 га. В связи со структурой данных и большим количеством используемых признаков для обучения был выбран алгоритм машинного обучения Random Forest Classifier [22]. Целевой показатель тренировки – урожайность культуры, которую оценивали как лучшую или худшую, в сравнении со средней за период наблюдений. Изначально массив данных делили в соотношении 80/20, а именно 80 % примеров использовали в качестве обучающих, 20 % – в качестве тестовых для проверки результата тренировки модели. Для наиболее полного использования данных, имеющихся в распоряжении, обучение и проверку проводили 5 раз со сдвигом выборки тестовых примеров таким образом, чтобы все 100 % данных были задействованы при обучении. Показатель ассигасу (англ. точность) определяли как отношение числа правильных прогнозов на тестовой выборке к общему числу тестовых примеров. По завершении вычислений ассигасу усредняли по 5 значениям (рис. 1).

Очевидно, что непосредственным влиянием на урожайность будет обладать совокупность таких факторов, как погода и используемые агротехнические приемы при возделывании культуры. Причем для более полного учета влияния агроприемов на итоговый результат, необходимо также принимать во внимание их воздействие в текущем вегетационном периоде. Обосновано это тем, что, если есть понимание положительного влияния тех или иных воздействий на урожайность культуры, то у сельхозпроизводителя имеется возможность воспроизвести выбранные агроприемы в дальнейшем.

Иная ситуация с данными о метеоусловиях. Спрогнозировать погоду на будущее невозможно, в связи с

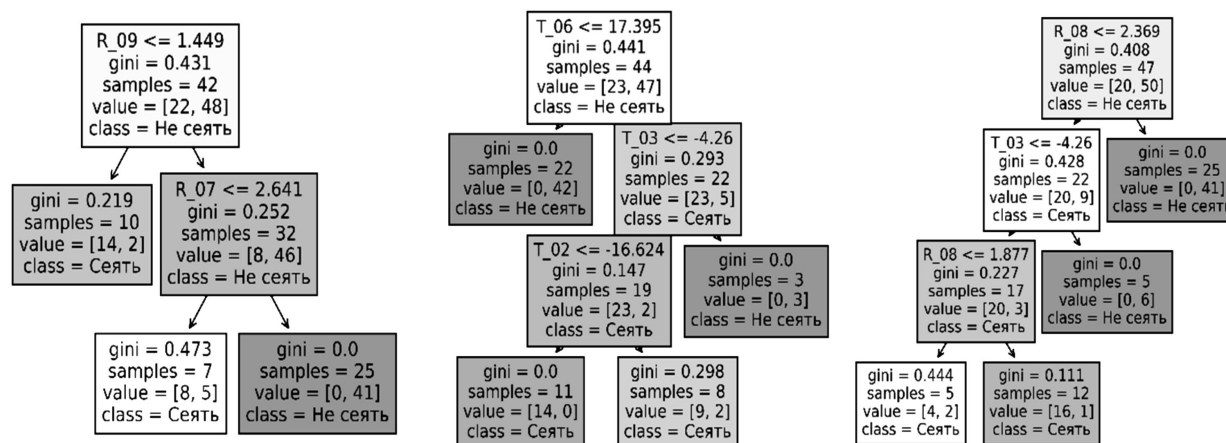


Рис. 1. Графическое представление дерева решений.

чем использовали данные предыдущего сельскохозяйственного года. При этом допускали, что метеоусловия предыдущего года должны влиять на результаты текущего года как минимум по двум причинам:

накопление влаги и прогрев почвы перед началом нового вегетационного периода;

стабильность долгосрочных климатических параметров, соответствующих норме определенного региона, из-за которой погодные аномалии одного вегетационного периода будут с высокой долей вероятности сглажены в следующем.

В исследованиях использовали сведения Кемеровского поста метеонаблюдений в Кемеровской области. Источником данных был web-ресурс «Погода и климат» (<http://www.pogodaiklimat.ru/>).

Результаты и обсуждение. Использование методов машинного обучения позволяет оценить весь комплекс воздействий на урожайность, включая как погодные, так и агротехнические факторы. Поэтому первой задачей исследования была демонстрация их применимости к анализу рядов данных урожайности, на которую оказывали управляющее воздействие в виде проводимых агротехнических мероприятий и внешних погодных факторов.

Данные по используемым агроприемам и культурам – категориальные признаки, поэтому к ним перед тренировкой модели применили классический метод преобразования (One-hot encoding) в результате чего входной вектор для тренировки модели выглядел как представлено в табл. 1.

Для определения влияния метеоусловий предыдущего года на урожайность текущего года использовали вектор, составленный из среднемесячных величин температур и осадков. Как было отмечено в работе [20], размерность вектора данных погодных значений, влияющих на точность прогноза, весьма ограничена. Увеличение размерности на данных статистики не приводило к росту точности прогноза, но снижало устойчивость

Табл. 1. Входной вектор агроприемов для тренировки модели

Агротехнический прием	Значение
Система обработки почвы _Без обработки	1/0
Система обработки почвы _Безотвальная (10-12 см)	1/0
Система обработки почвы _Безотвальная (20-22 см)	1/0
Система обработки почвы _Комбинированная глубокая	1/0
Система обработки почвы _Комбинированная минимальная	1/0
Система обработки почвы _Мульчирующая минимальная	1/0
Система обработки почвы _Нулевая	1/0
Система обработки почвы _Отвальная (20-22 см)	1/0
Система обработки почвы _Отвальная глубокая	1/0
Система обработки почвы _Отвальная минимальная	1/0
Предшественник_Пар	1/0
Предшественник_Сидеральный пар (донник)	1/0
Предшественник_Сидеральный пар (рапс)	1/0
Предшественник_Ямень без подсева донника	1/0
Предшественник_Ямень с подсевом донника	1/0
Высеваемая культура_Пшеница	1/0
Высеваемая культура_Ячмень	1/0
Посев_ПК Кузбасс-4,8	1/0
Посев_ПК Томь-5,1	1/0
Посев_СЗП-3,6	1/0

модели. За целевой показатель был взят бинарный индикатор (0/1), означающий то, что урожайность по культуре выше, чем в среднем за все годы наблюдений.

Результаты тренировки модели выглядели следующим образом. Средняя (по 5 значениям) величина показателя ассигасы для пшеницы составила 0,81; для ячменя – 0,87; в среднем по культурам – 0,84. Ввиду малого количества данных было принято решение в дальнейшем тренировать не отдельные модели для каждой культуры, а общую модель, где одним из входных параметров будет культура.

Полученный результат показывает достаточно высокую способность моделей прогнозировать ожидаемую благоприятность года по урожайности культур, основываясь на образе погоды предыдущего года с учетом применяемых или планируемых к использованию агроприемов. Однако наиболее важным остается вопрос о том, насколько много информации, влияющей на точность прогноза, содержится в данных о колебаниях среднемесячных температур за предыдущий год, в сравнении с влиянием факторов обработки почвы, предшественников и иных агрономических мероприятий текущего периода.

Для этого было проведено две альтернативные тренировки модели с измененными векторами входных данных. В одном случае учитывали только образ погоды предыдущего периода на основе среднемесячных значений температуры и осадков. Во втором учитывали применяемые агроприемы, а данные по погоде уменьшили до одного значения – среднегодовой температуры. По результатам тренировки модели с новыми величинами входных векторов показатель ассигасы для культуры (без различия ячмень или пшеница) в случае учета только погодных факторов предыдущего года был равен 0,7, при их минимальном учете (на уровне среднегодовых значений) – 0,73 (табл. 2).

Результат решения задачи с использованием входных векторов погоды за предыдущий год, представленный в графическом виде (рис. 2), отражает на плоскости распределение решений при кластеризации их в многомерном векторном пространстве. Точки, для которых цвет не соответствует цвету области, в которой они расположены, отражают неверный прогноз модели. Из рисунка видно, что в целом модель способна тренироваться на адекватный прогноз. Однако, очевидно, существует кластер решений, где она натренирована неверно. Ожидается, что подобных результатов можно будет избежать при наличии большего числа входных данных для тренировки.

Результаты исследования указывают на то, что влияние метеоусловий предыдущего года на текущую урожайность сравнимо с суммарным влиянием агроприемов, используемых в текущем году. Следовательно, при прогнозировании урожайности культур на ближайший сельскохозяйственный период одинаково важно учитывать, как планируемые к использованию агротехнические приемы, так и погодный профиль предыдущего года.

Таким образом, проведенные вычисления показали относительно высокую степень возможной достоверности предлагаемой системы прогнозирования. При

Табл. 2. Значения показателя ассигасы по результатам тренировки модели Random Forest Classifier в зависимости от структуры входных векторов

Структура вектора входных данных			агроприемы текущего года
погода за предыдущий год + агро-приемы текущего года	погода за предыдущий год		
пшеница	ячмень	0,7	0,73
0,81	0,87		

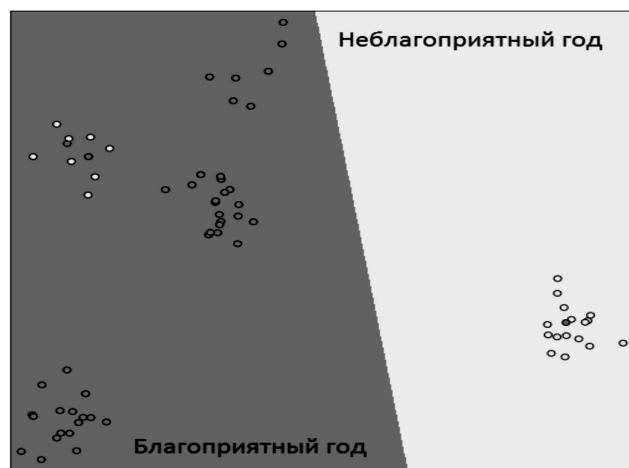


Рис. 2. Двумерный образ пространства решений прогноза урожайности при тренировке модели *Random Forest Classifier* с использованием данных о погоде предшествующего года.

достаточном количестве обучающих примеров (длительности постановки экспериментов и наблюдений) будет также возможным выявление степени воздействия каждого отдельного фактора и количественная оценка этого воздействия на ожидаемую урожайность.

Результаты тренировки модели показывают, что потенциальная урожайность культуры может быть с достаточной степенью значимости спрогнозирована, на основе агрометеорологических показателей предыдущего года и планируемых к использованию агротехнических приемов в текущем вегетационном периоде. Установлено, что даже небольшое количество данных по погоде за предыдущий период дает значимую информацию, которая может быть эффективно учтена при прогнозировании урожайности культур. Это открывает широкие перспективы для использования методов ML для учета погодных условий при сельскохозяйственном планировании, даже при отсутствии возможности составления достоверных долгосрочных прогнозов погоды. Необходимое условие решения такой задачи – корректный и полный набор данных по урожайности культур на исследуемых территориях вместе с длительными рядами данных о метеоусловиях.

Литература

1. Буховец А.Г., Семин Е.А., Бирючинская Т.Я. Современные подходы и методы в прогнозировании урожайности отдельных видов зерновых культур. Воронеж: Воронежский ГАУ им. Императора Петра I, 2016. 214 с.
2. Асалханов П.Г., Иванько Я.М., Полковская М.Н. Модели прогнозирования урожайности сельскохозяйственных культур в задачах параметрического программирования // Вестник Иркутского государственного технического университета. 2017. Т. 21. № 2 (121). С. 57–66. doi: 10.21285/1814-3520-2017-2-57-66.
3. Неверов А.А. Альтернативные модели долгосрочного прогнозирования урожайности зерновых культур для степной зоны Оренбуржья // Бюллетень Оренбургского научного центра УрО РАН. 2018. № 1. С. 1–9. doi: 10.24411/2304-9081-2018-11002.
4. Волкова Е.С., Мельник М.А., Фузелла Т.Ш. К оценке природных опасностей для сферы аграрного природопользования южной тайги Западной Сибири // Фундаментальные исследования. 2014. № 12 (часть 1). С. 153–157.

5. Агротехнологические основы технологий возделывания сельскохозяйственных культур / А.Н. Арефьев, С.В. Богомазов, В.А. Гуцина и др. Пенза: Пензенский ГАУ, 2018. 267 с.
6. Multi-objective land use allocation modeling for prioritizing climate-smart agricultural interventions / A. Dunnett, P.B. Shirsath, P.K. Aggarwal, et al. // Ecological Modelling. 2018. Vol. 381. P. 23–35. doi: 10.1016/j.ecolmodel.2018.04.008.
7. Keating B.A., Thorburn P.J. Modelling crops and cropping systems – Evolving purpose, practice and prospects // European Journal of Agronomy. 2018. Vol. 100. P. 163–176. doi: 10.1016/j.eja.2018.04.007
8. Альт В.В., Боброва Т.Н., Гурова Т.А. Компьютерные информационные системы в агропромышленном комплексе. Новосибирск: СибФТИ, 2008. 220 с.
9. Переход сельского хозяйства к цифровым, интеллектуальным и роботизированным технологиям / Е.А. Скворцова, И.С. Санду и др. // Экономика региона. 2018. Т. 14. Вып. 3. С. 1014–1028. doi: 10.17059/2018-3-23.
10. Аксенов А.Г. Анализ интеллектуальных систем поддержки принятия решений в сельском хозяйстве // Электротехнологии и электрооборудование в АПК. 2019. № 3. С. 46–51.
11. Jakku E., Thorburn P.J. A conceptual framework for guiding the participatory development of agricultural decision support systems // Agricultural systems. 2010. Vol. 103. No. 9. P. 675–682. doi: 10.1016/j.agsy.2010.08.007.
12. Lundström C., Lindblom J. Considering farmers' situated knowledge of using agricultural decision support systems (AgriDSS) to Foster farming practices: The case of CropSAT // Agricultural Systems. 2018. Vol. 159. P. 9–20. doi: 10.1016/j.agsy.2017.10.004.
13. Адамадзе К.Р., Касимова Т.М. Методы прогнозирования развития сельского хозяйства // Фундаментальные исследования. 2014. № 5-1. С. 122–126.
14. Косенко Т.Г. Оценка эколого-экономической эффективности сельскохозяйственного производства // Вестник Донского государственного аграрного университета. 2014. № 4–3. С. 12–17.
15. Сопряженное прогнозирование развития экономики сельского хозяйства / А.Н. Байдаков, О.Н. Бабкина, О.С. Звягинцева и др. // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2015. № 114. С. 1–16.
16. Model predictive control and its application in agriculture: A review / Y. Ding, L. Wang, Y. Li, et al. // Computers and Electronics in Agriculture. 2018. Vol. 151. P. 104–117. doi: 10.1016/j.compag.2018.06.004.
17. Chlingaryan A., Sukkarieh S., Whelan B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review // Computers and Electronics in Agriculture. 2018. Vol. 151. P. 61–69. doi: 10.1016/j.compag.2018.05.012.
18. Chavan T.R., Nandedkar A.V. AgroAVNET for crops and weeds classification: A step forward in automatic farming // Computers and Electronics in Agriculture. 2018. Vol. 154. P. 361–372. doi: 10.1016/j.compag.2018.09.021.
19. The Next Decade of Big Data in Ecosystem Science / S.L. LaDeau, B.A. Han, E.J. Rosi-Marshall, et al. // Ecosystems. 2017. Vol. 20. № 2. P. 274–283. doi: 10.1007/s10021-016-0075-y.
20. Каличкин В.К., Корякин Р.А., Куценогий П.К. Интеллектуальные системы в сельском хозяйстве (теоретический анализ возможного приложения): монография. Новосибирск: СФНЦА РАН, 2020. 296 с.
21. Меткальф Р. Еда 3.0. Бананы из Исландии и другие истории о продуктах. М.: Издательская группа «Точка», 2020. 287 с.
22. Leo B. Random Forests // Machine Learning. 2001. Vol. 45. No. 1. P. 5–32.

Поступила в редакцию 24.11.2020
После доработки 28.12.2020
Принята к публикации 12.01.2021