

должны учитывать, в том числе, и знания выбывающих специалистов (вышедших на пенсию, уволившихся и пр.) и хранить большие объемы данных и информации из различных источников [3].

Существуют различные подходы, модели и языки описания данных и знаний. Однако все большую популярность в последнее время приобретают онтологии. Онтология, по определению Грубера [Gruber, 1997], есть спецификация концептуализации, формализованное представление основных понятий и связей между ними. Онтология – это структурная спецификация некоторой предметной области, ее формализованное представление, которое включает словарь (или имена) указателей на термины предметной области и логические выражения и описывает, как они соотносятся друг с другом. Таким образом, онтологии обеспечивают словарь для представления и обмена знаниями о некоторой предметной области и множество связей между терминами в данном словаре.

Проектирование и разработка онтологий, т. е. онтологический инжиниринг, не является тривиальной задачей. Он требует от разработчиков профессионального владения технологиями инженерии знаний – от методов извлечения знаний до структурирования и формализации [4]. При явном интересе к онтологическому инжинирингу на сегодняшний день не существует промышленных систем проектирования онтологий. Тем не менее, по мнению авторов, разработка онтологии растениеводства может стать существенным шагом вперед в формализации знаний в области растениеводства.

Литература:

1. Якушев В. В. Информационно-технологические основы прецизионного производства растениеводческой продукции. Автореферат диссертации. СПб.: АФИ, 2013.
2. Гаврилова Т. А., Лещева И. А., Кудрявцев Д.В. Использование моделей инженерии знаний для подготовки специалистов в области информационных технологий // Системное программирование. СПб., СПбГУ. 2012. № 1.
3. Гаврилова Т. А. Онтологический инжиниринг, http://bigc.ru/publications/bigspb/km/ontolog_engeneering.php
4. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем: учебник для вузов. СПб.: Изд-во Питер, 2000.

МЕТОДОЛОГИЯ ПРИМЕНЕНИЯ БИНАРНОЙ РЕГРЕССИИ В ТОЧНОМ ЗЕМЛЕДЕЛИИ

В. М. Буре

Санкт-Петербургский Государственный университет

Методология точного земледелия предполагает широкое применение компьютерных технологий и автоматизированных систем анализа данных, непрерывного сбора и обработки статистической информации [1].

Во многих прикладных исследованиях зависимыми переменными являются дихотомические или бинарные величины, которые принимают значения единица и ноль. Для статистического анализа таких данных обычно используется бинарная регрессия или, иными словами, логит и пробит анализы [2–6].

Методы анализа данных на основе бинарной регрессии находят применение и в точном земледелии [7–10]. В работе предлагается основанный на бинарной регрессии подход к задаче оценки вероятности достижения заданного порогового уровня урожайности на конкретном поле в рамках методологии точного земледелия в условиях наличия довольно полной информации о самом поле, технологических однородных зонах на нем и условиях возделывания культуры.

Спрогнозировать числовое значение урожайности на конкретном поле чрезвычайно трудно даже при наличии исчерпывающего набора факторов. Гораздо легче поддается прогнозированию возможность превышения или невыполнения некоторого фиксированного порогового значения, определяющего хороший или допустимый уровень урожайности культуры для данного региона.

Предположим, что поле разбито на несколько однородных зон. Каждую из зон разобьем на элементарные участки равной площади, которые перенумеруем. Зададим некоторый пороговый уровень урожайности. Введем бинарную переменную y_i , (где i – номер элементарного участка внутри данной зоны однородности) по следующему правилу: если на данном участке урожайность оказалась выше порогового значения, то будем считать, что бинарная переменная принимает значение, равное единице, в противном случае бинарная переменная принимает значение ноль. Таким образом, для каждого элементарного участка i внутри однородной зоны определено значение бинарной переменной y_i , равное единице или нулю. Выделим набор почвенных и климатических характеристик x_1, x_2, \dots, x_k , определяющих, по мнению авторов, урожайность культуры. Любой набор факторов не может исчерпывающим образом описать все возможные взаимосвязи, существующие в природе и влияющие на урожайность культуры, поэтому любой прогноз в принципе носит только вероятностный характер, что обосновывает применение стохастической методологии на основе использования бинарной регрессии.

Выбор факторов, оказывающих значимое влияние на урожайность культуры, не полностью очевиден. Может оказаться, что среди выбранных факторов некоторые не оказывают существенного влияния на урожайность и поэтому могут быть исключены из рассмотрения, тем более это касается ситуации, когда измерение данных факторов вызывает затруднения. В связи с этим представляет интерес задача определения минимально достаточного набора факторов, по которым возможно удовлетворительное прогнозирование урожайности культуры.

Для каждого элементарного участка с номером i в результате проведения соответствующих измерений на них получаем сопоставление набора объясняющих переменных $x_{i1}, x_{i2}, \dots, x_{ik}$ и результирующей переменной y_i . Расположенные рядом элементарные участки вряд ли могут считаться статистически независимыми. Однако при наличии большого количества таких участков внутри зоны однородности можно произвести «прореживание» элементарных участков так, чтобы исключить наличие соседних участков в «прореженной» совокупности. В результате появляется массив данных, состоящий из наблюдений, которые можно с некоторым приближением считать статистически независимыми.

По полученному после «прореживания» массиву данных можно построить бинарную регрессию (логит или пробит регрессию), которая будет представлять собой регрессионную эмпирическую модель вероятности превышения порогового уровня урожайности при данных значениях объясняющих переменных, построенную для данной зоны однородности. Для действительно однородной зоны значения характеристик x_1, x_2, \dots, x_k должны меняться незначительно и, следовательно, при изменении характеристик внутри зоны однородности вероятность, рассчитываемая по построенной модели, также должна меняться мало. Однако на практике значения характеристик могут быть подвержены существенным изменениям. При сильном варьировании характеристик внутри зоны вариабельность вероятности может оказаться большой, что будет означать отсутствие однородности данной зоны.

Построенная эмпирическая модель будет характеризовать вероятность превышения порогового значения урожайности для данной зоны однородности в зависимости от значений объясняющих переменных. Зная пределы изменения объясняющих переменных внутри данной зоны однородности, можно будет изучить вариабельность эмпирической вероятностной модели внутри указанной зоны. Кроме того, можно будет получить некоторое усредненное значение для данной вероятности, характеризующее данную зону однородности в целом. Построенная эмпирическая модель даст возможность оценить степень однородности рассматриваемой зоны и вероятность получения заданной урожайности внутри зоны.

Подобное исследование должно быть проведено для каждой зоны однородности, после чего можно будет сравнить построенные эмпирические модели. Такое сравнение позволит оценить степень изменчивости вероятности урожайности между разными зонами однородности. В принципе, подобная модель может быть построена для всего поля в целом, однако точность ее может оказаться не достаточно высокой. Построение отдельных моделей для каждой из зон должно привести к повышению точности и достоверности результатов. Проверка адекватности построенных моделей проводится при помощи традиционных статистических методов[2–6].

Решение перечисленных задач имеет важный методологический смысл, так как позволяет с альтернативных позиций ответить на основные вопросы статистического анализа данных (об информативности выбранных факторов и их достаточности для адекватного математического описания), а также поставить и решить задачу прогнозирования продуктивности сельскохозяйственных культур в условиях точного земледелия.

Литература

1. Якушев В. П., Якушев В. В. Информационное обеспечение точного земледелия Рос. акад. с.-х. наук, Агрофиз. науч.-исслед. ин-т. СПб.: ПИЯФ РАН, 2007.
2. Greene W. H. *Econometric Analysis*, 5th edition. – New Jersey: Pearson Education, 2003.
3. Amemiya T. *Advanced Econometrics*. Cambridge: Harvard University Press, 1985.
4. Ben-Akiva M., Lerman S. *Discrete choice analysis* – The MIT Press, Cambridge Massachusetts, 1985.

5. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика. СПб.: Лань, 2013.
6. Буре В. М. Методология статистического анализа данных. СПб.: СПб. Гос. ун-т, 2007.
7. Stan G. Daberkow, William D. McBride Farm and operator characteristics affecting the awareness and adoption of precision agriculture technologies in US // Precision agriculture, 4, 163–177, 2003.
8. Robertson, M. J., Llewellyn R. S., Mandel R., Lawes R., Bramley R. G. V., Swift L., Metz N., O'Callaghan C. Adoption of variable rate fertilizer application in the Australian grain industry: status, issues and prospects // Precision agriculture, 13, 181–199, 2012.
9. J. Anita Dille, David A. Mortensen, Linda J. Young Predicting weed species occurrence based on site properties and previous year's weed presence // Precision agriculture, 3, 193–207, 2002.
10. Dale R. Walters, Anna Avrova, Ian J. Bingham, Fiona J. Burnett, James Fountain, Neil D. Havis, Stephen P. Hoad, Gareth Hughes, Mark Loosely, Simon J. P. Oxley, Alan Renwick, Cairistiona F. E. Topp, Adrian C. Newton Control of foliar diseases in barley: towards an integrated approach // Eur. J. Plant Pathol, 133, 33–73, 2012.

ПОСТРОЕНИЕ БАЗ ДАННЫХ ПО РЕЗУЛЬТАТАМ ПРЕЦИЗИОННЫХ ЭКСПЕРИМЕНТОВ

А. В. Конев

Агрофизический НИИ Россельхозакадемии, Санкт-Петербург

Современные технологии точного земледелия открывают большие возможности для совершенствования методики постановки полевых опытов в условиях (и с учетом) неоднородности почв. С появлением систем глобального позиционирования, средств почвенно-растительного мониторинга и геоинформационных систем стала возможной реализация в полностью автоматическом режиме полевых опытов по широкому спектру факторов плодородия.

При проведении полевых опытов с применением технологий точного земледелия важным источником информации, характеризующим уровень продуктивности поля, является электронная карта урожайности. Автоматическое получение и пространственная регистрация подобной информации является несомненным прорывом в области опытного дела, так как наличие электронной карты урожайности позволяет характеризовать поле не одним числом – средней урожайностью поля, а большим числовым массивом – урожайностью отдельных участков площадью 25 м². Существенно возрастает при этом возможность детального исследования характера и степени влияния на урожайность почвенно-климатических, агрохимических и других аспектов, выявленных на этапе мониторинга.

Разработанный в Агрофизическом Институте программный комплекс поддержки технологий точного земледелия включает в себя функционал, обеспечивающий информационную поддержку всей цепочки работ в рамках прецизионных экспериментов:

- регистрацию данных почвенно-растительного мониторинга с выделением однородных зон сельскохозяйственного поля;
- компьютерное планирование схемы прецизионного эксперимента;