

О классификации, распознавании образов и моделировании в задачах прогноза урожайности яровой пшеницы в условиях Алтайского края

Л.А. Хворова, Н.В. Гавриловская
АлтГУ, г. Барнаул

В задачах прогноза урожайности сельскохозяйственных культур важную роль играют методы классификации эмпирических данных. Общая же теория классификации сформировалась на основе работ по распознаванию образов. Поэтому классификация данных и распознавание образов, как правило, в любых научных исследованиях выступают в качестве двух взаимосвязанных этапов.

Классификация – это разбиение множества объектов на классы по степени их сходства (в нашем случае без информации о принадлежности объектов к каким-либо классам), так что объекты, отнесенные к одному и тому же классу, оказываются сходными в каком-то отношении, а отнесенные к разным классам – не сходными или менее сходными. *Распознавание же образов* предполагает отнесение нового объекта к определенному классу по некоторому решающему правилу.

С помощью методов классификации и распознавания образов авторами решается задача определения года-аналога с целью осуществления прогноза урожайности яровой пшеницы в условиях Алтайского края.

При решении задачи прогноза важно установить, какие исходные данные в большей степени определяют будущий урожай. С этой целью была проведена оценка влияния различных метеорологических ситуаций на урожай яровой пшеницы. Анализ обработки экспериментальных данных показал (позже было найдено подтверждение в литературе), что урожайность, являясь биологическим свойством, несет в себе информацию о соответствии между оптимальными и фактическими значениями множества факторов по периодам (фазам) вегетационного цикла.

В результате обработки агрометеорологических данных дисперсионным анализом по значимости и силе их влияния на урожайность из всех вариантов расчета агроклиматических показателей были отобраны следующие: сумма эффективных температур, сумма осадков, количество дней с осадками, дефицит влажности насыщения.

Процедуру кластеризации (разбиение на классы или кластеры) проводили несколько раз с помощью программы статистической обработ-

ки данных Statistica 6.0 при различных значениях числа кластеров (3, 4, 5, и т.д.), после чего выбиралась лучшая группировка в смысле критерия минимума отношений средних внутри кластерных и меж кластерных расстояний:

$$F = \frac{d_w / f_w}{d_b / f_b}.$$

Лучший вариант – разбиение на 13 кластеров. Так, например, годами-аналогами для 1997 года по результатам кластеризации стали: 1964, 1965, 1973, 1991 гг.

Используя данные по годам-аналогам, можно осуществлять **предварительный прогноз** урожайности ранней весной по эмпирической модели с помощью функций отклика:

$$Y_{j+1} = \begin{cases} Y_{\min} + (Y_j - Y_{\min}) \cdot \prod_{i=1}^k H_i(x), & \text{àññè } Y_j \geq \bar{Y}; \\ Y_{\min} + (Y_{\max} - Y_j) \cdot \prod_{i=1}^k H_i(x), & \text{àññè } Y_j < \bar{Y}, \end{cases} \quad (1)$$

где Y_{j+1} – урожайность текущего года (ожидаемая); Y_j – урожайность предыдущего года; Y_{\max} – максимальная; Y_{\min} – минимальная; \bar{Y} – средняя урожайности по всему временному ряду урожайностей; $H_i(x)$ – нормированные функции отклика, k – количество рассматриваемых факторов, влияющих на урожайность.

Общий вид нормированных функций отклика $H_i(x)$ установлен на основе экспериментальных и теоретических исследований. Совместное влияние нормированных функций отклика считается мультипликативным.

$$H_i(x) = \begin{cases} 1 - \frac{(x - x_2)^2}{(x_1 - x_2)^2}, & x_1 < x < x_2; \\ 1 - \frac{(x - x_3)^2}{(x_4 - x_3)^2}, & x_3 < x < x_4; \\ 1, & x_2 \leq x \leq x_3; \\ 0, & x \leq x_1, x \geq x_4, \end{cases}$$

где x_1 – минимальное; x_4 – максимальные значения i -го фактора, за пределами которых формирование урожая не происходит; x_2, x_3 – биологически оптимальные границы i -го фактора.

В общем случае задание параметрической формы частных функций отклика определяется нашими знаниями о качественном поведении рассматриваемого объекта. На основе выбранных частных функций отклика формируется обобщенная функция отклика, имеющая обычно мультипликативную или аддитивную формы. Определение параметров функций отклика производится процедурой совместной идентификации параметров всех рассматриваемых частных функций отклика.

Определение пороговых значений параметров функций отклика осуществляется из условия минимизации функционала

$$I = \sum_{i=1}^N \|y_i - x_i\|^2, \text{ где } x_i - \text{расчетная урожайность по модели, } y_i - \text{фак-$$

тическая урожайность. В связи с этим рассматривалась следующая задача оптимального управления: минимизировать функционал

$$I = \sum_{i=1}^N \|y_i - x_i\|^2 \rightarrow \min$$

при условиях $x_{i+1} = F_i(x_i, u_i), i = 0, 1, \dots, N-1, x_0 = a;$

где $[u_i] = (u_0, \dots, u_{N-1}), u_i = (u_i^1, \dots, u_i^r) -$ параметры управления;

$x_i = (x_i^1, \dots, x_i^n) -$ искомые параметры;

$F_i = (F_i^1, \dots, F_i^n) -$ заданные функции; $[y_i] = (y_0, \dots, y_N),$

$y_i = (y_i^1, \dots, y_i^n) -$ данные экспериментальных исследований.

Исследования при решении задачи прогноза показали, что в условиях Алтайского края существенное влияние на будущий урожай оказывают первые фазы развития яровой пшеницы: посев-всходы-кущение. Поэтому был осуществлен **уточняющий прогноз** урожайности по модели (1) после двух-трех недель вегетационного периода (когда сценарий погоды уже известен).

Модель прогноза урожайности зерновых культур идентифицирована на основе данных, представленных Алтайским центром по гидрометеорологии. Пороговые значения параметров функций отклика найдены в результате проведения численных экспериментов с моделью по многолетним временным рядам экспериментальных данных. Результаты численных расчетов ожидаемой урожайности на 1989 и 1997 гг. представлены в таблице 1.

Таблица 1. Прогнозируемая величина урожайности, ц/га

Год	Урожайность фактическая	Урожайность расчетная	Средняя абсолютная процентная ошибка
1989	25,3	23,97	5,2%
1997	12,8	12,6	1,5%

Процент отклонения предсказанных данных от фактических в среднем составляет 3,4%. Как видно из таблицы 1 и приведенных оценок, результаты численных расчетов, проведенных по эмпирической модели урожайности, хорошо согласуются с экспериментальными данными. Результаты исследований обобщены в многочисленных таблицах и представлены на графиках, отражающих связь между исследуемыми элементами метеорологической ситуации и урожайностью.

Проведенную классификацию и прогноз урожайности яровой пшеницы следует рассматривать как начальный этап работы по оценке урожайности зерновых культур в условиях Алтайского края. Однако он показывает, что данная методика имеет вполне определенный смысл и достаточно хорошо позволяет осуществлять прогноз. Практическое приложение разработанной методики по прогнозу урожайности состоит в возможности заблаговременно, еще до начала сева, оценить вероятную урожайность. Данные оценки позволили бы рационально маневрировать структурой посевных площадей, целесообразно распределять производственные ресурсы, эффективно планировать внешнеторговые операции, более дифференцированно использовать арсенал технологических приемов воздействия на урожайность.