

УДК 311:681.3.06

Л.А. Хворова, Н.В. Гавриловская

Прогнозирование урожайности зерновых культур: методы и расчеты

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 [1].

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 определении года-аналога и 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 [2]. Анализ обработки экспериментальных данных показал, что урожайность, являясь биологическим свойством, несет в себе информацию о соответствии между оптимальными и фактическими значениями множества факторов по периодам (фазам) вегетационного цикла. В условиях Алтайского края существенное влияние на будущий урожай оказывают первые фазы развития яровой пшеницы: посев–всходы–кущение. 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.

Проведем процедуру кластеризации (разбиение на классы или кластеры) с помощью программы статистической обработки данных Statistica 6.0 несколько раз при различных значениях числа кластеров, после чего необходимо выбрать лучшую группировку

в смысле критерия минимума отношений средних внутрикластерных и межкластерных расстояний:

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

Для сравнения нескольких типизаций и выбора наиболее оптимальной из них необходим критерий, численная мера качества классификации.

В качестве достаточно общего критерия успешности можно взять следующий показатель [3]

$$J = J_1 - J_2 \text{ (или } J = J_1 / J_2 \text{),}$$

где

$$J_1 = \frac{2}{m(m-1)} \sum_{i=1}^m \sum_{j=i+1}^m \bar{D}_{ij}, \quad J_2 = \frac{1}{m} \sum_{i=1}^m \bar{D}_{ii}$$

Здесь \bar{D}_{ii} – среднее расстояние между точками внутри i -го класса; \bar{D}_{ij} – среднее расстояние между парами точек i -го и j -го классов.

Данные расстояния рассчитываются в модуле Cluster Analysis. Так, среднее расстояние между парами точек i -го и j -го классов можно подсчитать используя опцию «Кластерные усреднения & евклидова расстояния» (Cluster Means&Euclidean Distances).

Среднее расстояние между точками внутри i -го класса можно найти используя опцию «Элементы каждого кластера & расстояния» (Member of each cluster&distances). Подсчитав сумму строки расстояний каждого элемента до центра кластера, получим сумму расстояний между точками внутри i -го класса. Останется только просуммировать расстояния для каждого класса. Результаты расчетов приведены в таблице 1.

В общей форме критерий качества классификации представляется в виде условного математического ожидания (или среднего риска)

$$J(C) = \int_X Q(X, C) p(X) dX,$$

где $Q(X, C)$ – функция вектора $C = (c_1, c_2, \dots, c_n)$, зависящая также и от вектора X (точнее, от его расстояния до соответствующего эталона), c_1, c_2, \dots, c_n – параметры классификации, $p(X)$ – функция плотности вероятности вектора X .

Для графической интерпретации используем критерий «каменистой осыпи». Обычно для выбора размерности какого-либо пространства используют

Расчет численного показателя мера качества классификации

Кластеры	Dij	m	Dii	J1	J2	J
2	0,70	2	1,25	0,70	0,62	1,13
3	0,61	3	1,26	0,20	0,42	0,49
4	0,57	4	1,26	0,09	0,31	0,30
5	0,47	5	1,50	0,05	0,30	0,16
6	0,43	6	1,45	0,03	0,24	0,12
7	0,36	7	1,52	0,02	0,22	0,08
8	0,34	8	1,48	0,01	0,18	0,07
9	0,31	9	1,45	0,01	0,16	0,05
10	0,28	10	1,41	0,01	0,14	0,04
11	0,26	11	1,35	0,00	0,12	0,04

график зависимости стресса от размерности (график каменистой осыпи). Этот критерий впервые был предложен Кэттелом (Cattell, 1966) в контексте решения задачи снижения размерности в факторном анализе.

Кэттел предложил найти такую абсциссу на графике, в которой график стресса начинает визуально сглаживаться в направлении правой, пологой его части, и, таким образом, уменьшение стресса максимально замедляется. Образно говоря, линия на рисунке напоминает скалистый обрыв, а черные точки на графике напоминают камни, которые ранее упали вниз. Таким образом, внизу наблюдается как бы каменистая осыпь из таких точек. Справа от выбранной точки на оси абсцисс лежит только «факторная осыпь».

Выбор критерия качества классификации существенно зависит от цели, для которой она производится, и оставляет исследователю значительную свободу.

Лучший вариант – разбиение на 6 кластеров (рис. 1, 2). Так, например, годами-аналогами для 1997 г. по результатам кластеризации стали: 1971, 1973, 1974, 1976, 1981, 1988 гг.



Рис. 1. Диаграмма численной меры качества классификации

Из рисунка 2 видно, что в первый кластер попали годы с низкими показателями суммы эффективных температур, количеством осадков и дней с осадка-

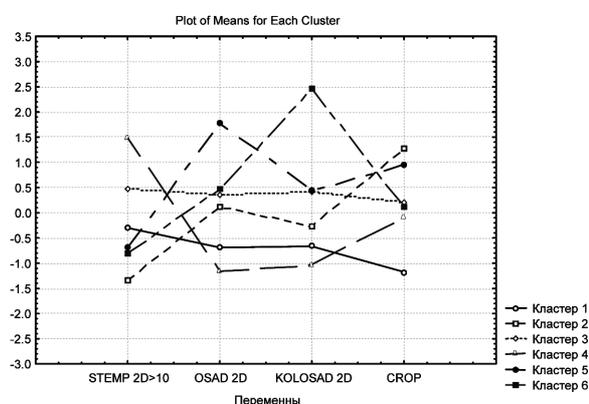


Рис. 2. Значения стандартизованных данных по кластерам: STEMP2D > 10 – сумма эффективных температур; OSAD2D – количество осадков; KOLOSAD2D – количество дней с осадками; CROP – урожайность яровой пшеницы

ми за первые две декады вегетационного периода и с минимальной урожайностью яровой пшеницы. Во второй кластер – годы с минимальным показателем суммы эффективных температур, средними значениями показателей количества осадков и дней с осадками и максимальной урожайностью. Третий кластер содержит годы со средними значениями всех факторов. Четвертый кластер характеризуется высоким значением суммы эффективных температур, минимальными значениями количества осадков и дней с осадками и средней урожайностью. В пятый кластер попали годы с достаточно низкой суммой эффективных температур, большим количеством осадков, средним количеством дней с осадками и высокой урожайностью. Шестой кластер характеризуется следующим образом: достаточно низкое значение сумм эффективных температур, среднее значение количества осадков и большим числом дней с осадками, а урожайность яровой пшеницы имеет среднее значение.

Характеристики каждого класса достаточно точно подтверждают влияние комплекса соответствующих факторов на урожайность.

Используя данные по годам-аналогам, можно осуществлять **предварительный прогноз** урожайности после двух-трех недель вегетационного периода (когда сценарий погоды уже известен) по эмпирической модели с помощью функций отклика:

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

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 – максимальное 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 [4].

Определение пороговых значений параметров функций отклика осуществляется из условия минимизации функционала $I = \sum_{i=1}^N \|y_i - x_i\|^2$, где x_i –

расчетная урожайность по модели; y_i – фактическая 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 $x_{i+1} = (x_i, u_i)$, $i = 0, 1, \dots, N-1$, $x_0 = a$;

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.

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 2.

Таблица 2
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.

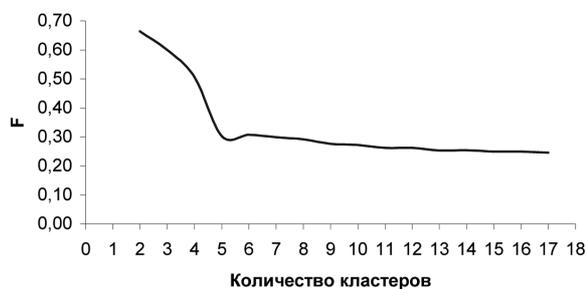


Рис. 3. Зависимость отношения средних внутрикластерных и межкластерных расстояний от количества кластеров

Проведенная классификация и результаты расчетов позволяют взять погодный сценарий всего ве-

гетационного периода года-аналога и осуществить уточняющий прогноз урожайности. Для этого была проведена повторная классификация для уточнения года-аналога. Лучший вариант – разбиение на 13 кластеров (рис. 3). Так, например, годами-аналогами для 1997 г. по результатам кластеризации стали: 1964, 1965, 1973, 1991 гг.

Сравнительный анализ различных разбиений показал, что годом-аналогом для 1997 г. является 1973 г. Величина урожайности по погодному сценарию 1973 по формуле (1) составила 13,2 ц/га.

Проведенную классификацию и прогноз урожайности яровой пшеницы следует рассматривать как начальный этап работы по оценке урожайно-

сти 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, оценивать 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.

Библиографический список

1. Сонечкин, Д.М. Математическая теория классификации и ее применение в метеорологии / Д.М. Сонечкин // Метеорология и гидрология. – 1969. – №12.

2. Хворова, Л.А. Применение информационных технологий, математических методов и моделей для обработки и ана-

лиза многомерных данных / Л.А. Хворова, Н.В. Гавриловская, Н.Н. Лопатин // Известия АлтГУ. – 2006. – №1 (49).

3. Груза, Г.В. О принципах автоматической классификации метеорологических объектов / Г.В. Груза, Э.Я. Ранькова // Метеорология и гидрология. – 1970. – №32.