АЛГОРИТМ ЭКОНОМИКО-МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ СЕЛЬСКОХОЗЯЙСТВЕННОЙ ИНФОРМАЦИИ

В. А. КОЛЕМАЕВ

(Москва)

1. Для сельскохозяйственного производства характерно сейчас (и это сохранится еще длительное время) получение в пределах географического региона одного вида продукции многими (а не одним!) в известном смысле однотипными по способам производства и специализации хозяйственными объектами. При моделировании каждой группы таких объектов * целесообразно строить для них единые функции затраты-выпуск и другие соотношения между экономическими факторами. Такие соотношения, полученные статистическими методами (например, с помощью аппарата регрессионного анализа), естественно использовать при формировании моделей математического программирования (см., [1]).

Рассматриваемый нами алгоритм является формализованным выражением вышеупомянутой идеи. В соответствии с этим он включает в себя следующие последовательно выполняемые этапы.

Первый этап — транспонирование основной информации — вытекает из самой сущности статистической части алгоритма и является общим для всех задач данного класса, решаемых с одной и той же основной информацией.

Второй этап — преобразование транспонированной информации в каноническую — специфический для каждой задачи и заключается в преобразовании столбцов-показателей к нужному виду и в расположении их в требуемом порядке.

Третий этап — первичная статистическая обработка и контроль — состоит в получении различных статистических характеристик столбцов-показателей (среднее, дисперсия, асимметрия, эксцесс, парные коэффициенты корреляции и т. п.) для использования их с целью исключения грубых ошибок случайного характера с помощью соответствующих статистических критериев. Этап по существу является общим для всех задач.

Четвертый этап — обработка и преобразование проверенной информации — рассчитан на подготовку информации к получению уравнений регрессии и на окончательное преобразование информации, не используемой для получения уравнений регрессии (например, агрегирование ресурсов, принадлежащих отдельным объектам, в соответствии с информацией об агрегации (ИА)). Если рассматривается статистическая задача, то на этом этапе производится осреднение информации по времени.

Пятый этап — расчет уравнений регрессии — кроме получения уравнений и их статистических характеристик предназначен также для анализа уравнений и их коррекции. Если рассматриваемая совокупность хозяйственных объектов неоднородна, то для каждой группы однородных объектов рассчитываются свои функции затраты-выпуск.

Шестой этап — формирование задачи математического программирования. В отличие от предыдущих он не отработан еще достаточно хорошо. Основная идея, которая используется в настоящее время для стандартизованного формирования матрицы задачи линейного программирования, состоит в следующем: матрица разбивается на непересекающиеся и далее уже неделимые подматрицы, причем это разбиение определяется экономической и структурной автономией подматриц. Если подматрицы уже сформированы и находятся во внешней памяти (ВМ), то для формирования матрицы задачи достаточно соответствующим образом задать информацию о ее формировании (ИФ). Алгоритмы же для формирования подматриц придется писать для каждого класса задач одного типа (т. е. отличающихся только размерами матриц, но имеющих одинаковое строение), хотя некоторая стандартизация выработана и в этой части: нулевые подматрицы вообще не рассматриваются, для образования единичных подматриц достаточно задать их размеры. Предусмотрено также составление диагональных подматриц с неравными элементами и не-задач данного типа позволит сделать выводы об общих типах встречающихся в них подматриц и тем самым довести уровень стандартизации при их формировании до уровня других этапов.

Седьмой этап — решение задачи линейного программирования — стандартный для всех задач данного класса, различия могут быть только в том, что для каждой

* Примером такой группы объектов является совокупность всех (или почти всех) хозяйств области, если область однородна по своим природно-экономическим условиям, или совокупность хозяйств любой природно-экономической зоны области — в противном случае. При этом каждая зона характеризуется своими функциями затраты-выпуск.

структуры такой задачи может оказаться более подходящим соответствующий алгоритм математического программирования.

Восмой этап — анализ решения и печать результатов. Оптимальное решение сравнивается с принятым; печатаются двойственные оценки, характеризующие дефицитность ресурсов; оптимальное решение печатается в соответствии со структурой переменных (например, вначале печатаются площади разных видов почв под разными культурами, затем затраты ресурсов под разные культуры и т. п.).

2. К настоящему времени во Всесоюзном научно-исследовательском институте кибернетики МСХ СССР (ВНИИК) отлажена система программ на ЭВМ «Минск-22» по конкретному варианту алгоритма. Перечислим основные его особенности (в дальнейшем имеется в виду составление программ по более сложным вариантам алгоритма, например с разными функциями затраты-выпуск): отсев ошибочных данных производится по правилу, аналогичному правилу «трех сигм»; информация усредняется по времени, т. е. рассматривается статистический вариант; затратные факторы делятся на прямые, включаемые в уравнения регрессии, и косвенные, для которых определяется статистическая связь с прямыми; учитывается получение продукции с земель разного качества, при этом набор соответствующих производственных функций строится для каждой почвы в отдельности (в связи с чем в алгоритм добавляются части, связанные с переработкой данных о почвах); используется блочный алгоритм Данцига — Вулфа для решения задачи линейного программирования, причем блоки характеризуются одинаковыми функциями затраты-выпуск и отличаются только правыми частями (этим условиям, например, удовлетворяют задачи по распределению ресурсов и заданий между объектами, различающимися по наличию разных видов земель); формирование матрицы коэффициентов блоков и связующих условий производится автоматически по заданной информации о формировании матриц (ИФ) и заранее подготовленным подматрицам, расчет которых (за исключением диагональных подматриц) нестандартизован; при решении задачи линейного программирования по алгоритму Данцига — Вулфа выполняется только одна итерация, что определяется возможностями применяемой готовой программы (см. [2]), в дальнейшем имеется в виду дополнить эту программу таким образом, чтобы проводились все необходимые итерации для получения оптимального решения.

В качестве критерия выбран максимум валовой продукции, причем при необходимости с помощью изменения сопровождающей информации можно перейти к любому другому линейному критерию.

Система программ включает в себя как составные части следующие самостоятельные программы: вывода чисел на широкую печать*; преобразований из программы регрессионного анализа [3]; регрессионного анализа**, компактного мультипликативного алгоритма симплекс-метода для задач малой размерности (см. [4]); аппроксимационного алгоритма по принципу разложения Данцига — Вулфа (см. [2]); мультипликативного алгоритма симплекс-метода для задач большой размерности***.

Все они перед включением в систему были подвергнуты соответствующей переработке. Кроме того, была разработана серия новых программ, которые вместе с вышеуказанными полностью обеспечивают выполнение алгоритма. Отметим также, что система математического обеспечения алгоритма включает в себя ряд программ, позволяющих автоматизировать отладку при его видоизменениях. Общее количество команд в системе составляет около 16 тыс.

Необходимо остановиться на кратком описании принятого алгоритма, потому что предыдущее изложение дает только общее представление о нем. Для удобства интерпретации будем применять терминологию, соответствующую задаче об оптимизации (по некоторому критерию) производства растениеводческой продукции.

Предполагается, что основная информация подразделяется на ряд массивов (информация за один год состоит из одного или нескольких массивов), расположенных в плотную друг за другом в ВП. Каждый массив транспонируется по отдельности. Если массив целиком не помещается в один блок магнитного оперативного запоминающего устройства (МОЗУ), то транспонирование производится по частям. В сопровождающей информации (универсальных константах, УК) для каждого массива отведена одна ячейка, имеющая структуру ****):

* Составлена во ВНИИК Л. Л. Хлюпиной.

** Составлена во Всесоюзном научно-исследовательском институте экономики сельского хозяйства (ВНИИЭСХ) А. И. Сидоровой.

*** Составлена во ВНИИК Б. Н. Михалёвым.

**** Далее будем придерживаться следующих обозначений при занесении нескольких целых чисел в одну ячейку: $(m)^0 - (m)_s$ единиц 4-го (старшего) восьмеричного разряда, $(m)_s \leq (7777)_s$; $(m)' - (m)_s$ единиц 8-го восьмеричного разряда; $(m)'' - (m)_s$ единиц последнего (самого младшего) восьмеричного разряда.

$$(\hat{m}_s)^0 \vee (\tilde{n}_s)' \vee (m_s)''^* \quad (\kappa_s - 1)\tilde{m}_s \pm \hat{m}_s = m_s,$$

$$0 < \hat{m}_s < \tilde{m}_s; \quad (\kappa_s - 1)\tilde{n}_s + \hat{n}_s = n, \quad 0 < \hat{n}_s < \tilde{n}_s,$$

где s — номер массива; m_s — число показателей в нем; κ_s — число частей; $\tilde{m}_s, \tilde{n}_s$ — показатели, характеризующие размеры частей. Они выбираются так, что $\tilde{n}_s m_s$ чисел первичной информации должны разместиться в первом блоке МОЗУ, а $\tilde{m}_s n$ чисел транспонированной информации — во втором блоке. После транспонирования информация также располагается в ВП массивами, следующими вплотную один за другим.

Преобразование транспонированной информации в каноническую происходит порциями (информация за год разбивается на одну или несколько порций), причем размер порции определяется емкостью второго блока МОЗУ (в первом размещаются программы). Для каждой порции в сопровождающей информации (в информации о канонических показателях, ИК) задается информация о вызываемых показателях и их преобразованиях. Ячейки вызова (порядок их следования определяет порядок канонических показателей) имеют следующую структуру: $(\kappa_\sigma)^0 \vee (s)' \vee (\mu_\sigma)''$, где σ — номер группы вызываемых показателей транспонированной информации; κ_σ — число показателей в σ -й группе (порядок следования которых совпадает с требуемым порядком для канонической информации); μ_σ — номер начального показателя σ -й группы в s массиве.

Вызванная во второй блок МОЗУ информация подвергается преобразованию на основе списка преобразований, каждая строчка которого представляет собой команду трехадресной ЭВМ, смоделированной на базе команд ЭВМ «Минск-22» (см. [3]).

Приведем таблицу некоторых команд этой гипотетической машины

- +01 $N_1 N_2 N_3$ — сложение;
- +02 $N_1 N_2 N_3$ — вычитание;
- +03 $N_1 N_2 N_3$ — умножение;
- +04 $N_1 N_2 N_3$ — деление;
- +06 $N_1 00 N_3$ — перенос;
- +10 $N_1 n_2 N_3$ — сложение с константой;
- +11 $N_1 n_2 N_3$ — умножение на константу;
- +17 $N_1 n_2 00$ — сумма элементов столбца,

где N_1 — первый адрес моделируемой машины (занимает индексную часть); N_2 — второй адрес (два старших восьмеричных разряда адресной части); N_3 — третий адрес (два следующих восьмеричных разряда адресной части); n_2 — номер константы.

На основе этих команд производится операции со всеми n элементами столбцов-показателей с номерами N_1, N_2 , а результат занимает место столбца с номером N_3 второго блока МОЗУ. Константы помещаются в определенное место в первом блоке МОЗУ.

Порции канонической информации, последовательно формируемые в оперативной памяти, записываются затем в ВП.

Каноническая информация по своему назначению делится на две части: общую и специальную. Общие данные представляют собой сведения о ресурсах объектов и используются для формирования правых частей в задаче линейного программирования. Общие данные перерабатываются следующим образом: вначале осредняются по времени, затем фактические затраты экономических ресурсов (труд, удобрения и т. п.) пересчитываются в нормативные по найденным коэффициентам связи между фактическими затратами и величинами, определяющими потенциальные значения затрат (например, трудоспособные — труд). Ресурсы различных видов земель определяют с помощью данных об удельных весах этих видов и общей площади пашни в каждом объекте.

В заключение, в соответствии с информацией об агрегации (ИА) производится агрегирование, т. е. ресурсы объектов, образующих блок в задаче линейного программирования, складываются. Для агрегации нескольких объектов в один блок отводится 5 ячеек, рассматриваемых как одна сплошная 185-разрядная ячейка, в которой расставлены единицы в разрядах по номерам агрегируемых объектов.

Специальные данные представляют собой информацию о выращивании и первичной переработке отдельных видов продукции растениеводства (культур — лен, зерно, силос и т. п.). Остановимся, например, на описании обработки информации о выращивании одного из них. Каноническая информация о выращивании имеет следующий вид (индекс культуры опускаем)

$$S^t, U_1^t, \dots, U_v^t, \dots, U_N^t, U^t, Y^t, \quad t = 1, \dots, T,$$

* Знак $\vee$ применяется для соединения отдельных элементов псевдочисла.

где t — номер года; S^t — вектор-столбец площадей данной культуры в году t ; U_v^t — вектор-столбец затрат v -го вида (в денежном или натуральном выражении) при выращивании рассматриваемой культуры, $v = 1, \dots, N$; U^t — вектор-столбец общих затрат (в денежном выражении); Y^t — вектор-столбец выхода продукции (в натуральном выражении).

Вначале информация преобразуется к виду $S^t, U_1^t, \dots, U_v^t, \dots, U_N^t, \dots, V^t, Y^t, t = 1, \dots, T$, где V^t — вектор косвенных затрат. Затем все данные (кроме первого вектора S^t) подвергаются единообразной обработке. Покажем ее на примере вектора V^t :

- 1) получение относительных показателей: $v_j^t = V_j^t / S_j^t, j = 1, \dots, n, t = 1, \dots, T$, где j — номер объекта;
- 2) получение статистических характеристик каждого столбца

$$\bar{v}_t = \frac{1}{n} \sum_{j=1}^n v_j^t, \quad \hat{\sigma}_v^t = \sqrt{\frac{1}{n} \sum_{j=1}^n (v_j^t - \bar{v}_t)^2};$$

$t = 1, \dots, T;$

- 3) регистрация в каждом столбце элементов, отклоняющихся от среднего более чем на $\lambda \hat{\sigma}$ (λ задан в УК);
- 4) осреднение показателей во времени

$$v_j = \begin{cases} \frac{1}{T_j} \sum_{t \in \tau_j} v_j^t, & T_j > 0, \\ -1, & T_j = 0; \end{cases} \quad \tau_j = \{t: |v_j^t - \bar{v}_v| \leq \lambda \hat{\sigma}_v\},$$

где T_j — количество элементов множества τ_j ;

- 5) отсев тех объектов, для которых хотя бы один показатель не определен (т. е. $T_j = 0$), в результате чего число используемых объектов, вообще говоря, уменьшится ($\bar{n} \leq n$);
- 6) нахождение коэффициентов статистической связи между прямыми и косвенными затратами;
- 7) окончательное формирование информации для регрессионного анализа в следующем виде: $p_1, \dots, p_\mu, \dots, p_M, \Delta u_1, \dots, \Delta u_v, \dots, \Delta u_N, y$, где p_μ — вектор удельных весов μ -го вида земли* при производстве данной культуры,

$$\Delta u_v = u_v - \frac{1}{\bar{n}} \sum_{j=1}^{\bar{n}} u_{vj},$$

$\Delta u_v, v = 1, \dots, N$, — векторы центрированных прямых затрат.

Уравнение регрессии рассчитывается (см. [5]) в форме

$$y = \sum_{\mu=1}^M \alpha_\mu p_\mu + \sum_{v=1}^N \beta_v \Delta u_v + \varepsilon,$$

где ε — статистическая погрешность; α_μ — выход продукции с 1 га μ -й земли при средних затратах; β_v — приращение продукции при увеличении v затрат на единицу (от среднего уровня).

З а м е ч а н и е. Выбор затратных факторов происходит на предварительной стадии исследования, до обращения к программе. Предполагается, что факторы слабо коррелированы. Если факторы экономических затрат коррелированы в значительной степени, то при предварительных исследованиях берется набор производных слабо коррелированных факторов (см. [6]), что должно быть отражено в ИК.

После расчета параметров каждого уравнения регрессии производится их анализ, в результате которого незначимым оценкам присваиваются нулевые значения. Если же оказывается незначимым все уравнение в целом, то нулевые значения присваиваются всем оценкам β_v . Результаты регрессионного анализа, а также полученные ранее различные статистические коэффициенты запоминаются в ВП.

При формировании матрицы задачи линейного программирования исходим из того, что технологические коэффициенты у всех объектов одинаковы и задаются набором найденных ранее статистических коэффициентов, поэтому достаточно сформировать матрицу только одного блока; для всех остальных блоков они будут такими же.

* В связи с недостатком места в данном изложении не будем касаться вопроса о переработке информации о видах земель.

же (правые части для разных блоков были сформированы при обработке общих канонических данных). Сформированные матрицы блоков запоминаются в ВП. Затем формируется и запоминается матрица связующей задачи, ограничениями которой служат общие задания и общие подвижные ресурсы для всей системы рассматриваемых объектов. Различия в местоположении объектов принимаются во внимание путем учета транспортных издержек. В качестве критерия выбран валовой доход всей совокупности объектов.

При решении задачи линейного программирования делается только одна итерация по алгоритму Данцига — Вулфа, т. е. вначале находятся системы решений для каждого блока на основе учета только закрепленных за ними неподвижных ресурсов, а затем с помощью решения связующей задачи подбираются такие их линейные комбинации, для которых функционал принимает наибольшее значение (с учетом всех ограничений). Полученное квазиоптимальное решение затем печатается на АЦПУ с учетом блочного строения задачи и структуры производственных процессов в каждом блоке. Кроме того, в тех же графах печатаются для сравнения фактические значения переменных. На этом процесс решения заканчивается.

Следует отметить, что с того момента, когда основная и сопровождающая информация занесены в ВП, весь процесс решения происходит полностью автоматически (т. е. при этом не требуется работа за пультом, за исключением случаев сбоя ЭВМ).

В настоящее время к данной программе сформирована сопровождающая информация для задачи прогноза на 2—3 года распределения плана заготовок между хозяйствами природно-экономического района области в соответствии с их природно-экономическими ресурсами. Общий подход к такому типу задач содержится в [7].

Изложенный в статье алгоритм переработки массовой сельскохозяйственной информации может быть положен в основу решения ряда задач автоматизированной системы управления сельским хозяйством.

ЛИТЕРАТУРА

1. G. Tintner. *Methodology of Mathematical Economics and Econometrics*. Univ. of Southern California, 1965.
2. Б. Н. Михалев. Программа определения оптимального варианта размещения и специализации сельскохозяйственного производства с учетом зональных особенностей. В сб. Программы решения экономических задач на ЭЦВМ. М., 1965 (НИИ ЦСУ СССР).
3. О. М. Дукарский, А. Г. Закурдаев. Статистический анализ и обработка наблюдений на ЭВМ «Минск-22». М., «Статистика», 1971.
4. Г. Я. Данелян, Б. Н. Михалев. Программа компактного мультипликативного алгоритма симплекс-метода для ЭВМ «Минск-22». В сб. Программы решения экономических задач на ЭЦВМ. Вып. 1 (7), М., 1967 (НИИ ЦСУ СССР).
5. В. А. Колемаев. Статическая корреляционная модель экономической оценки земли. В сб. Актуальные вопросы аграрной теории. Воронеж, 1968 (Воронеж. с.-х. ин-т им. К. Д. Глинки).
6. Д. Н. Лоули, А. Максвелл. Факторный анализ как статистический метод. М., «Мир», 1967.
7. В. А. Колемаев. Производственные функции и земельный кадастр. Экономика и матем. методы, 1969, т. V, вып. 2.

Поступила в редакцию
30.XII.1969