

4. Моисеев Н.Н. Математические задачи системного анализа.- М.: Наука, 1991. – 369 с.
5. Михалевич В.С., Волкович В.Л. Вычислительные методы исследования и проектирования сложных систем.- М.: Наука, 1982.- 286 с.

## УДК 681.3

*Шматко А.В., канд. техн. наук, доц., УГЗУ,  
Неронов А.А., зам. нач. курса, УГЗУ*

### **ИСПОЛЬЗОВАНИЕ ТЕОРИИ R-ФУНКЦИЙ В ЗАДАЧЕ ИДЕНТИФИКАЦИИ ИСТОЧНИКОВ ЗАГРЯЗНЕНИЯ**

(представлено д-ром техн. наук Куценко Л.Н.)

В работе предложен метод идентификации источников загрязнения, базирующийся на использовании структурно-аналитического (СА) подхода для машинного построения правил классификации в классе R-функций с адаптирующейся к реальным условиям структурой. Приведен алгоритм применения СА-метода для идентификации источника загрязнения по анализам проб воды, отобранных ниже места сброса

**Постановка проблемы.** Содержательная постановка данной проблемы сводится к следующему. Загрязняющие вещества попадают в речную воду со сточными водами промышленных предприятий. При этом возможны случаи, когда по тем или иным причинам на каком-либо из этих предприятий произойдет аварийное увеличение объема промышленных стоков. При известных химическом составе и концентрации основных ингредиентов (химический спектр) сточных вод каждого предприятия необходимо, по результатам химического анализа образца речной воды, отобранного на участке реки, расположенном ниже места сброса сточных вод, определить предприятие, виновное в техногенной чрезвычайной ситуации.

**Анализ последних исследований и публикаций.** Решению задач классификации посвящено большое количество работ отечественных [1-5] и зарубежных [6-12] ученых. В настоящее время сформированы общие дискриминанты [7, 8] и лингвисти-

ческий (структурный) подходы [10] к распознаванию образов. Однако проблема машинного узнавания еще далеко не решена. Основные трудности связаны с принципиальной сложностью моделирования механизма узнавания, разнотипностью признаков объектов, правильной оценкой объема таблицы эмпирических данных (ТЭД) и качества получаемого по ней правила классификации (ПК). Существующие сейчас методы классификационной обработки данных (КОД), за некоторым исключением [1,13], непригодны или мало эффективны для распознавания объектов с разнотипными признаками. При использовании дискриминантного подхода, для синтеза ПК, непосредственно не анализируются признаки объекта, игнорируется реальная структура данных и навязывается гипотетическая структура образов в пространстве признаков (компактность, нормальность распределения и т.д.). Лингвистический подход имеет пока ограниченные возможности для распознавания изображений и сцен. Адекватное отражение предметной области реального мира в структуре данных позволяет найти эффективную модель классификации, успешно решать задачи принятия решений. Если данным навязана жесткая структура, то можно прийти к отрицательному результату. Поэтому в данной работе рассматривается структурно-аналитический метод (СА) распознавания объектов с разнотипными признаками с применением R-отображений [14, 15]. Ранее этот метод успешно применялся для решения задач идентификации сложных производственных объектов [14], распознавания производственных ситуаций [16], технической диагностики и контроля качества работы технических систем. Для идентификации источников загрязнения СА-метод ранее не использовался.

**Постановка задачи и ее решение.** Идея СА-метода распознавания состоит в системном комбинированном использовании принципов дискриминантного и лингвистического подходов к проблеме распознавания с применением R-отображений.

Рассмотрим более подробно алгоритм решения задачи идентификации источника загрязнения речной воды, в котором реализуется СА-подход к распознаванию образов.

На первом шаге определяются все источники загрязнения, у которых расстояние до устья ( $L_i$ ) меньше  $LL$ . Для каждого из отобранных предприятий строятся структурные образы источников загрязнений (СОИЗ) ( $OP_i, i=1...h$ ). При построении СОИЗ используется информация из паспорта предприятия, сельскохозяйствен-

ного объекта и других источников информации о  $C_{ij}$  - концентрации загрязняющих веществ в сточных водах источника загрязнения ( $i$  – номер источника загрязнения,  $j$  – номер ингредиента в сточных водах источника).

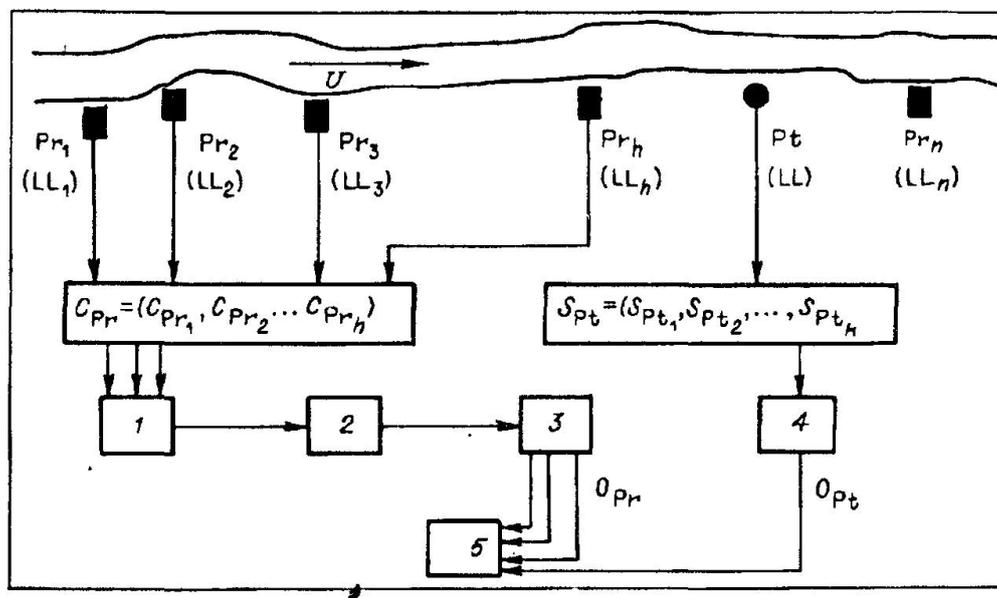


Рис. 1 – Иллюстрация к алгоритму идентификации источников загрязнения, где  $S_{ij}$  ( $i$  – номер пробы,  $j$  – номер ингредиента) - проба речной воды, пункт контроля -  $P_t$ ,  $LL$  - расстояние от устья реки,  $P_r$  - предприятие (или участок реки)

Алгоритм построения СОИЗ заключается в следующем:

В блоке 1 для каждого источника загрязнения составляется система уравнений, моделирующих процесс разбавления отдельных компонентов сточных вод при прохождении ими пути от места сброса сточных вод до места забора речной воды на анализ. Методика построения таких моделей описана в работе [17]. При этом учитываются гидрологические характеристики реки (глубина, скорость течения и др.) а также метеорологические условия.

В блоке 2 описывается процесс разложения ингредиента в речной воде (без учета метеоданных) описывается уравнением

$$S_i = S_{oi} \exp(k_i x / V_{cp}),$$

где  $S_{oi}$  - начальная концентрация  $i$ -го ингредиента в речной воде,  $k_i$  - скорость распада  $i$ -го вещества в единицу времени в данных условиях [4],  $x$  - расстояние от места забора до предприятия источника загрязнения,  $V_{cp}$  - средняя скорость течения реки.

В блоке 3 проводится формирование СОИЗ.

В блоке 4 формируется «образ пробы» в тех же координатах, что были использованы и для СОИЗ.

В блоке 5 выполняется соответственно сама процедура распознавания образов: по образу пробы  $OP_t$  проводится распознавание класса из множества существующих классов  $OP_r$ .

При практической реализации решение поставленной задачи СА-методом производится в варианте КОД с обучением на основе построения СА-моделей распознавания, которые представляются двузначными R-функциональными регулярными грамматиками  $G_R^{(2)}$  [14,15], порождающими язык описания образов  $L_R(G_R^{(2)})$  в терминах R-функций, заданных на решающих прадеревьях. Внутренние вершины прадеревя  $G_R^{(2)} = \Phi(x)$  соответствуют терминальным свойствам-предикатам (С-П)  $f_i$  ( $1, \tilde{M}; \tilde{M} \leq M$ ), внешние — листьям с названиями распознаваемых классов или их фрагментов, а дуги — бинарным отношениям в виде R-конъюнкций [16] в отличие от традиционной конкатенации [10]. Путь, соединяющий корневую вершину  $f_1$ , грамматического прадеревя  $G_R^{(2)}$  с любым его листом, образует конъюнктивную цепочку (ветвь) соответствующих С-П, эквивалентную предложению в языке  $L_R(G_R^{(2)})$ . Конечное множество предложений, определенных по входной ТЭД, соответствует синтаксическому образцу, по которому непосредственно восстанавливается неизвестная  $G_R^{(2)}$  с помощью обучающихся или самообучающихся алгоритмов. Грамматика  $G_R^{(2)}$  порождает предложения синтаксического образца и позволяет предсказывать другие предложения той же природы, что и в образце. В качестве рабочей гипотезы принимается, что в любой задаче КОД, как правило, найдется словарь терминальных С-П из небольшого числа слов, для которого каждая реализация  $x \in Z_0$  удовлетворяет лишь одному предложению языка  $L_R(G_R^{(2)})$ , а вероятность ошибки классификации не превышает допустимую величину. При этом между объектом распознавания  $\bar{x}$  и предложением

из ПК  $\Phi(x)$  установлено отношение истинности вместо традиционной перекодировки [1, 10]. Истинность предложения  $\Phi(\bar{x}) \in L_R(G_R^{(2)})$  для данного  $\bar{x}$  означает, что объект  $\bar{x}$  обладает всеми свойствами, которые определены входящими в предложение терминальными С-П с точностью до минимума эмпирического риска и с учетом структурной сложности  $\Phi(\bar{x})$ . Классификация объектов на  $s$  классов производится путем подстановки реализаций  $\bar{x}$  в С-П корневой вершины  $f_1 \in G_R^{(2)}$  и дальнейшего самоанализирующегося (без громоздкой синтаксической процедуры анализа [10]) процесса грамматического разбора (распознавания). Результат классификации определяется номером  $x_{0k}$ , ( $k = \overline{1, s}$ ) листа грамматической ветви, для которой  $\Phi(\bar{x})$  оказалось истинным после подачи распознаваемого  $\bar{x}$  объекта на вход  $G_R^{(2)}$ .

Приведем основные эмпирические положения, которые применяются в процедуре восстановления грамматик:

- элементы цепочки синтаксического описания какой-либо смеси веществ с одинаковыми номерами могут меняться местами друг с другом ввиду случайных изменений концентраций ингредиентов в речной воде;
- в синтаксическом описании спектра пробы речной воды некоторые элементы, характерные для описаний спектров сточных вод предприятий, могут отсутствовать. Это соответствует результатам химических реакций веществ и образованию новых соединений, которые могут разлагаться или выпадать в осадок;
- наиболее информативными элементами цепочки синтаксического описания спектра пробы речной воды являются те, которые имеют наибольшие номера интервалов. Это объясняется малой вероятностью исчезновения загрязняющих веществ с наибольшей концентрацией в сточных водах предприятий. Соотношения данных концентраций более устойчивы к случайным возмущениям, чем соотношения концентраций веществ, представленных в промышленных стоках малыми концентрациями.

**Выводы.** В работе предложен алгоритм идентификации источников загрязнения, базирующийся на использовании СА-метода и теории R-функций. Представление описанных ПК R-функциями создает единую формальную основу для логической структуры идентифицируемых источников загрязнения и их аналитического описания. Применение операторов  $\aleph$ -формализации

к грамматикам  $G^{(2)}$  позволяет получить R-функциональную грамматику  $G_R^{(2)}$ , которая описывается R-дизъюнкцией и R-конъюнкцией терминальных С-П. При этом СА-модель идентифицируемого источника загрязнения можно представить значной  $R_s$  функцией.

## ЛИТЕРАТУРА

1. Александров В.В., Горский Н.Д. Алгоритмы и программы структурного метода обработки данных. – Л.: Наука, 1970. – 383 с.
2. Айвазян С.А., Бежаева З.И., Староверов О.В., Классификация многомерных наблюдений. – М.: Статистика, 1974. – 239 с.
3. Бонгард М.М. Проблемы узнавания. – М.: Наука, 1967. – 320 с.
4. Вапник В.Н., Червоненкис А.Я., Теория распознавания образов. – М.: Наука, 1974. – 415 с.
5. Цыпкин Я.З. Основы теории обучающих систем. – М.: Наука, 1970. – 251 с.
6. Бенерджи Р. Теория решения задач. – М.: Мир, 1972. – 224 с.
7. Дуда Р., Харт П., Распознавание образов и анализ сцен. – М.: Мир, 1976. – 511 с.
8. Нильсон Н. Обучающие машины. – М.: Мир, 1967. – 180 с.
9. Ту Дж. Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 411 с.
10. Фу К.С. Структурные методы в распознавании образов. – М.: Мир, 1977. – 319 с.
11. Хант Э., Марин Дж., Стоун Ф. Моделирование процесса формирования понятий на вычислительной машине.. – М.: Мир, 1970. – 303 с.
12. Хант Э. Искусственный интеллект. – М.: Мир, 1978. – 558 с.
13. Лбов Г.С. Методы обработки разнотипных экспериментальных данных. – Новосибирск : Наука., 1981. – 158 с.
14. Сироджа И.Б. Салыга В.И. Идентификация динамических объектов с использованием R-функций. Изв. АН КазССР. сер.физ.-мат., 1972, №3, с. 51-55
15. Сироджа И.Б. Теоретические основы синтеза структурно-аналитических моделей и алгоритмов распознавания образов для автоматизации классификационной обработки данных

- (КОД). – Мат. Методы анализа динам. Систем, 1978, вып. 2, с.79-102
16. Стоян Ю.Г., Проценко В.С., Манько Г.П. и др. Теория R-функций и актуальные проблемы прикладной математики . – Киев: Наук. Думка, 1986. – 264 с.
17. Методические основы оценки и регламентирования антропогенного влияния на качество поверхностных вод / Под ред. А.В. Караушева . – Л: Гидрометеиздат, 1987. – 287с.