

Методы адаптивного структурного прогнозирования¹

Ю.А. Дорофеюк

Аннотация. Рассматривается задача анализа и прогнозирования в многопараметрической системе управления. Предполагается, что структура объектов исследуемой системы может меняться со временем. Разработан адаптивный алгоритм отслеживания таких структурных изменений. Построение такого интегрального описания объектов и адаптивной прогнозной модели позволяет существенно повысить эффективность результатов принимаемых управленческих решений.

Ключевые слова: классификационные методы анализа данных, адаптивный алгоритм автоматической классификации, адаптивная прогнозная модель, оптимизация выбора числа классов.

1. Постановка задачи

В работе рассматривается крупномасштабная система управления, состоящая из большого числа формально не структурированных объектов. Пусть исследуемая система состоит из n объектов, каждый из которых характеризуется набором из k параметров. Изучается поведение этого множества объектов в последовательные дискретные моменты времени. В момент времени t объект $x_j(t)$ можно представить точкой в k -мерном пространстве параметров X . Точки $x_j(t_1), \dots, x_j(t_m)$ представляют известную часть траектории j -го объекта.

В большинстве приложений для принятия управленческого решения в момент времени t_m используется совокупная информация об известных траекториях каждого объекта и прогноза значений $x_j(t_m + 1)$, $j = 1, \dots, n$. При этом, как правило, информация по каждому объекту рассматривается независимо от остальных [1]. Однако для многих прикладных задач требуется прогнозировать не точные значения параметров объекта в момент t_{m+1} , а лишь класс,

к которому будет принадлежать этот объект в соответствующие моменты времени в рамках некоторой структуры множества объектов изучаемой системы. В работе для выявления такой структуры использовался комплексный алгоритм автоматической классификации [2].

Подобная схема структуризации хорошо работает в условиях стационарного функционирования исследуемой системы управления, когда структура объектов в пространстве параметров X меняется незначительно. В условиях же существенной динамики структуры исследуемых объектов (например, когда изменяется число классов такой структуры) необходимо разрабатывать более адекватные схемы структуризации и прогнозирования.

2. Методы адаптивной структуризации

Пусть исследуемая система состоит из n объектов, каждый из которых характеризуется набором из k параметров. Вводится в рассмотрение k -мерное пространство параметров X , в котором каждый объект представляется точкой

¹ Работа выполнена при частичной поддержке РФФИ, проекты 08-07-00349-а, 10-07-00210-а.

$x_j = (x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(k)})$, $j = 1, \dots, n$. Предполагается, что вектор значений параметров x_j достаточно полно характеризует состояние j -го объекта на момент сбора информации, то есть взаиморасположение множества точек x_1, \dots, x_n в пространстве X отражает реальную структуру исследуемых объектов на этот момент времени. Для выявления такой структуры используется комплекс алгоритмов структурно-классификационного анализа, специально разработанный для решения таких задач [2]. Этот комплекс включает алгоритмы: m -локальной оптимизации заданного критерия J , выбора информативных параметров, выбора начального разбиения, выбора «оптимального» числа классов. Так как существенная часть алгоритмов комплекса используется и в адаптивном варианте структурно-классификационного анализа, приведём краткое описание некоторых из них.

2.1. Алгоритм m -локальной оптимизации

Вначале опишем работу алгоритма 1-локальной оптимизации. Для простоты, без ограничения общности, рассматривается случай двух классов $r = 2$. Пусть задано начальное разбиение R_0 всех точек классифицируемой выборки x_1, \dots, x_n . Обозначим через $x_j \in A_1$ точки, относящиеся к первому классу, а через $x_j \in A_2$ – ко второму. Алгоритм итерационный - на каждом шаге рассматривается одна точка из последовательности $x_1, \dots, x_n, x_1, \dots, x_n, x_1, \dots$ («зацикленная» исходная последовательность). Отнесение точки к одному из двух классов обозначается с помощью индекса $\rho(x_j) = \begin{cases} 1, & \text{если } x_j \in A_1 \\ -1, & \text{если } x_j \in A_2 \end{cases}$.

Тогда алгоритм 1-локальной оптимизации определяется следующим образом: $\rho(x_j) = \text{sign}[J(x_j \in A_1) - J(x_j \in A_2)]$.

В итоге точка x_j относится к тому классу, для которого значение критерия J будет больше (если эти значения равны, то для определённости точка относится к классу с меньшим номером). Алгоритм заканчивает работу, если на некотором цикле среди точек x_1, \dots, x_n не будет сделано ни одной «переброски» точки из класса в класс.

Алгоритм m -локальной оптимизации – это поэтапное применение к выборке алгоритмов s -локальной оптимизации, $s = 1 \div m$. На s -ом этапе алгоритм работает по той же схеме, только на каждом его шаге происходит пробная «переброска» из класса в класс не одной, а s точек. Подсчитывается значение критерия J до и после «переброски», Принадлежность каждой из s точек к классу либо остаётся неизменной (J до «переброски» больше, чем после), либо меняется на другой класс - в противном случае. В данном случае цикл – это число шагов, равное числу всевозможных различных наборов, в каждый из которых входит s точек, выбранных из n точек исходной выборки. Доказана сходимость алгоритма за конечное число шагов к локальному максимуму критерия J .

Разработан эвристический алгоритм сокращённого перебора, который на каждом шаге для пробной «переброски» использует s точек в определённом смысле ближайших к границе между классами.

При моделировании и в приложениях в качестве критерия J использовался функционал J_1 средней близости точек в классах, определяемый через потенциальную функцию [3, 4] близости точек x и y :

$$K(x, y) = 1 / \{1 + \alpha R^p(x, y)\}, \quad (1)$$

где α и p - настраиваемые параметры алгоритма. Средняя близость точек в классе определяется как:

$$K(A_i, A_i) = \frac{2}{n_i(n_i - 1)} \sum_{i=1}^{n_i} \sum_{j>i} K(x_i, x_j), \quad (2)$$

где $K(x_i, x_j)$ определяется формулой (1), n_i – число точек в классе A_i . Тогда критерий J_1 находится из выражения:

$$J_1 = \sum_{i=1}^r \frac{n_i}{n} K(A_i, A_i). \quad (3)$$

2.2. Алгоритм выбора числа классов

Для выбора числа классов используется специальная экспертно-компьютерная процедура, которая работает следующим образом. Сначала эксперт-пользователь оценивает диапазон (r_{\min}, r_{\max}) , в пределах которого заведомо находится искомое число классов. Далее, используя любой алгоритм автоматической клас-

сификации (в настоящей работе применялся алгоритм m -локальной оптимизации), проводится разбиение анализируемого множества объектов на $r_{\min}, r_{\min} + 1, \dots, r_{\max}$ классов. Качество каждой из полученных классификаций оценивалось с помощью критерия:

$$J_3 = J_1 - qJ_2, \quad (4)$$

где J_1 вычисляется по формуле (3), а величина средней близости между классами J_2 – по формуле (5):

$$J_2 = \frac{1}{r-1} \sum_{i=1}^r \sum_{j>i}^r \frac{n_i + n_j}{n} K(A_i, A_j), \quad (5)$$

где $K(A_i, A_j) = \frac{1}{n_i n_j} \sum_{x_i \in A_i} \sum_{x_p \in A_j} K(x_i, x_p)$ – мера

близости классов A_i, A_j . Величина потенциальной функции $K(x_i, x_j)$ определяется формулой (1). Величина q в (4) и величины α и p в (1) – это свободные (настраиваемые) параметры алгоритма. Фактически, параметр q является масштабирующим параметром, приводящим к одному масштабу средние значения функционалов J_1 и J_2 ; на практике величина q имеет значение порядка 2-7 (обычно во столько раз отличается средняя близость внутри классов от средней близости между самими классами). Более подробно процедуры выбора значений свободных параметров рассмотрены в [5].

Формально в качестве «оптимального» можно выбрать такое число классов r_{opt} , которое соответствует максимальному значению $J_3(r_j)$, т.е. $r_{opt} = r_j$, для которого $\max J_3(r_j)$, $r_j = r_{\min}, \dots, r_{\max}$. Однако существование значимой, но неиспользованной при классификации информации, например, ввиду отсутствия данных, может привести к тому, что полученное таким способом r_{opt} не будет «истинно оптимальным».

Для компенсации этого недостатка предлагается использовать следующую экспертную процедуру. Экспертам - специалистам в соответствующей предметной области представляются значения $J_3(r_j)$, $r_j = r_{\min}, \dots, r_{\max}$, показанные для удобства в виде графика, на

котором отмечается значение r_{opt} (оно соответствует максимальной точке на графике $J_3(r_j)$).

Используя эту информацию, эксперты могут корректировать выбираемое число классов. В подавляющем числе случаев экспертное число классов либо совпадает с r_{opt} , либо незначительно (± 1) отличается от него.

При классификации многомерных объектов во время такой экспертизы анализируется также классификация каждого объекта. Для этой цели экспертам сообщается информация о мере близости $K(x_i, c_j)$ каждой точки x_i к центрам классов c_j $j=1, \dots, r_{opt}$ в оптимальной классификации, т.е. матрица близости $\|K(x_i, c_j)\|$, $i=1, \dots, n$, $j=1, \dots, r_{opt}$. Перенесение точки (объекта) x_i из j -го класса в l -й считается допустимым, если величины $K(x_i, c_j)$ и $K(x_i, c_l)$ отличаются незначительно. Другими словами, содержательно обоснованное перенесение допустимо для точек, расположенных вблизи границы между соответствующими классами.

2.3. Адаптивный алгоритм структуризации объектов

Описанный выше алгоритм выбора «оптимального» числа классов хорошо работает в стабильных условиях развития исследуемой системы. Для отслеживания изменений структуры объектов в условиях существенной динамики предлагается следующая схема.

В момент времени t_1 с помощью описанного выше алгоритма m -локальной оптимизации [2] производится структуризация n точек в пространстве X на r классов, каждый из которых и характеризует определённый тип объекта. Число классов $r = r_{opt}(t_1)$ определяется с помощью описанного в разделе 2.2 алгоритма выбора «оптимального» числа классов из диапазона (r_{\min}, r_{\max}) , определяемого экспертами. Вводится понятие модели (эталона) класса $a_i(t)$, $i=1, \dots, r$ (чаще всего - это центр класса) [6]. Для каждого объекта кроме принадлежности к классу вычисляются расстояния до этало-

нов всех классов $R_{ij}(t)$, $i = 1, \dots, r$; $j = 1, \dots, n$. Для отслеживания существенных изменений структуры исследуемой системы объектов во времени все результаты обработки данных на момент времени t_1 , включая результаты классификации на $r \in (r_{\min}, r_{\min} + 1, \dots, r_{\max})$ классов и расстояния до эталонов $R_{ij}(t_1)$, сохраняются.

Заметим, что на практике структуризация объектов чрезвычайно редко проводится в пространстве исходных признаков, обычно сначала производится выделение набора информативных параметров. В работе для этой цели используются алгоритмы экстремальной группировки параметров, входящие в комплексный алгоритм [2].

Данные об исследуемых объектах, собранные в момент времени t_2 , распределяются по классам каждой классификации на r_l классов, $r_l \in (r_{\min}, r_{\max})$. Для этой цели используется алгоритм распознавания образов с учителем метода потенциальных функций [3, 4]. После этого подсчитываются значения критерия $J_3(r_l, t_2)$ (4) и в качестве оптимального (на момент времени t_2) выбирается такое число классов $r_{opt}(t_2)$, которое соответствует максимальному значению этого критерия. В случае необходимости используется экспертная процедура коррекции $r_{opt}(t_2)$.

После того, как определена принадлежность всех точек к тому или иному классу в пределах каждой классификации $r_l \in (r_{\min}, r_{\max})$, производится пересчёт эталонов $a_{il}(t_2)$, $i = 1, \dots, r_l$, $r_l \in (r_{\min}, r_{\max})$. Для каждой точки с предыдущего шага пересчитываются, а для каждой новой точки вычисляются расстояния до новых эталонов $R(a_i(t_2), x_j(t_2))$, $i = 1, \dots, r_l$, $j = 1, \dots, n$, $r_l \in (r_{\min}, r_{\max})$. Такая процедура выполняется для всех m моментов времени в пределах диапазона T стационарности, выбираемого экспертным путём [7]. В первый же момент времени вне этого диапазона система «перезагружается» – все предыдущие результаты отправляются в архив, структуризация проводится заново, как для момента времени t_1 . Разработан вариант итерационного адаптивного

алгоритма, когда структуризация объектов производится для данных, собранных в моменты времени, находящихся в пределах скользящего окна ширины T .

3. Адаптивный алгоритм прогнозирования

В качестве прогнозной модели для каждого объекта используется марковская цепь с r состояниями (здесь r – число классов разбиения), то есть на каждом шаге рассчитываются элементы матрицы переходных вероятностей $P = \|p_{ji}\|$, $j = 1, \dots, n$, $i = 1, \dots, r$. Разработан алгоритм пересчёта на каждом шаге соответствующих переходных вероятностей p_{ji} с использованием информации о значениях расстояний до центров классов и условия нормировки $\sum_{i=1}^r p_{ji} = 1$ для всех $j = 1, \dots, n$ [8].

Адаптивный вариант алгоритма работает следующим образом. Пусть после первого шага для всех классификаций на r_l классов, где $r_l \in (r_{\min}, r_{\max})$, подсчитаны расстояния $R_{ji}^{(1)} = R(x_j(t_1), a_{il}(t_1))$, $i = 1, \dots, r_l$, $j = 1, \dots, n$, $r_l \in (r_{\min}, r_{\max})$ от точек $x_j(t_1)$ до эталонов $a_{il}(t_1)$. Индекс l в этой и последующих формулах обозначает, что соответствующая величина рассчитана для классификации на r_l классов. Тогда элементы матрицы переходных вероятностей $p_{ji}^{(1)} = p_{ji}(t_1)$ рассчитываются следующим образом:

$$p_{ji}^{(1)} = \frac{\alpha_{ij}^{(1)}}{R_{ji}^{(1)}}, \quad i = 1, \dots, r_l, j = 1, \dots, n, \quad r_l \in (r_{\min}, r_{\max}), \quad (6)$$

где нормирующий множитель $\alpha_{ij}^{(1)}$ определяется выражением (7):

$$\alpha_{ij}^{(1)} = \frac{\prod_{i=1}^{r_l} R_{ji}^{(1)}}{\sum_{i=1}^{r_l} \frac{1}{R_{ji}^{(1)}} \prod_{i=1}^{r_l} R_{ji}^{(1)}}, \quad r_l \in (r_{\min}, r_{\max}). \quad (7)$$

На s -ом шаге для каждой классификации $r_l \in (r_{\min}, r_{\max})$ элементы матрицы переходных

вероятностей (6) модифицируются при помощи следующей процедуры. Введём обозначения:

$$\Delta R_{lji}^{(s)} = R_{lji}^{(s-1)} - R_{lji}^{(s)}; \quad \Delta \hat{R}_{lji}^{(s)} = \frac{R_{lji}^{(s-1)} - R_{lji}^{(s)}}{R_{lji}^{(s-1)} + R_{lji}^{(s)}}. \quad \text{Если}$$

j -ая точка совпадает с эталоном i_0 -го класса ($x_j(t_s) = a_{i_0}(t_s)$), т.е. $R_{lji_0}^{(s)} = 0$, то

$$p_{lji}^{(s)} = \begin{cases} 1, & \text{если } i = i_0, \\ 0, & i = 1, \dots, r_l, i \neq i_0 \end{cases}$$

Для случая, когда $R_{lji_0}^{(s)} \neq 0$, происходит модификация всех переходных вероятностей по следующей схеме:

$$p_{lji}^{(s)} = \gamma \left[p_{lji}^{(s-1)} + \left(\frac{1 + \text{sign}(\Delta R_{lji}^{(s)})}{2} - p_{lji}^{(s-1)} \text{sign}(\Delta R_{lji}^{(s)}) \right) \Delta \hat{R}_{lji}^{(s)} \right], \quad (8)$$

где, как обычно: $\text{sign}(z) = \begin{cases} 1, & \text{если } z \geq 0, \\ -1, & \text{если } z < 0 \end{cases}$, а

γ - нормирующий множитель (9) (аналогичный (7)), который определяется условием нормировки переходных вероятностей $\sum_{i=1}^r p_{lji}^s = 1$:

$$\gamma = \frac{1}{1 + \left(\frac{1 + \text{sign}(\Delta R_{lji}^{(s)})}{2} - p_{lji}^{(s-1)} \text{sign}(\Delta R_{lji}^{(s)}) \right) \Delta \hat{R}_{lji}^{(s)}}. \quad (9)$$

Введение в (8) и (9) величины $\text{sign}(\Delta R_{lji}^{(s)})$ вызвано необходимостью производить различными способами модификацию переходных вероятностей для случаев увеличения и уменьшения расстояния от точки $x_j(t_s)$ до эталонов классов $a_{i_0}(t_s)$ на s -ом шаге. А именно: в случае уменьшения величины $R_{lji}^{(s)}$ по отношению к $R_{lji}^{(s-1)}$ (т.е. $\Delta R_{lji}^{(s)} < 0$) изменение соответствующей переходной вероятности происходит за счёт её увеличения на некоторую долю от $(1 - p_{lji}^{(s-1)})$; а в случае увеличения величины $R_{lji}^{(s)}$ по отношению к $R_{lji}^{(s-1)}$ (т.е. $\Delta R_{lji}^{(s)} > 0$) изменение соответствующей переходной вероятности происходит за счёт её уменьшения на некоторую долю от $p_{lji}^{(s-1)}$. Это необходимо для

выполнения условий нормировки для переходных вероятностей $0 < p_{lji}^{(s)} < 1, i = 1, \dots, r$.

Построенные при помощи описанного выше алгоритма матрицы переходных вероятностей P_l используются для прогнозирования принадлежности объекта тому или иному классу. При этом на каждом шаге используется только одна матрица переходных вероятностей, соответствующая классификации на «оптимальное» для данного шага число классов, то есть на s -ом шаге для прогнозирования используется матрица $P_{l(s)}$, где $r_{l(s)} = r_{opt}(t_s)$.

На практике обычно используется не рандомизированная, а байесовская схема, когда объект относится к тому классу i_0 , для которого $p_{ji_0} = \max_{i=1, \dots, r_{opt}(t_s)} p_{ji}$. В случае равенства переходных вероятностей p_{ji} для прогнозируемого объекта для двух или нескольких классов, он относится к классу с наименьшим номером.

Заключение

Разработанная методология использовалась при анализе и совершенствовании процедур принятия решений для нескольких крупномасштабных социально-экономических систем управления, в основном, регионального характера; в том числе – региональная система управления здравоохранением, пассажирскими автоперевозками, система анализа и прогнозирования социально-экономического развития субъектов РФ и др. Во всех приложениях, а также при машинном моделировании была подтверждена высокая эффективность разработанной методологии структурно-классификационного анализа и прогнозирования.

Оказалось, что для некоторых приложений (с достаточно высоким уровнем помех при измерении параметров) существенно более эффективным оказывается использование алгоритмов размытой классификации, в том числе с фоновым классом [6].

Литература

1. Статистическое моделирование и прогнозирование. Сборник под ред. Гранберга А.Г. – М.: Финансы и статистика, 1990. – 382 с.

2. Дорофеюк Ю.А. Комплексный алгоритм автоматической классификации и его использование в задачах анализа и принятия решений. / Таврический вестник информатики и математики. 2008. № 1. –С. 171-177.
3. Айзерман М.А., Браверман Э.М., Розоноэр Л.И. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970.
4. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. – М.: Наука, 1983.
5. Дорофеюк Ю.А. Моделирование и анализ эффективности комплексного алгоритма классификационного анализа сложно-организованных данных. / Управление развитием крупномасштабных систем (MLSD'2009): Труды Третьей международной конференции. / -М.: ИПУ РАН, 2009. –с.299-308.
6. Бауман Е.В., Дорофеюк А.А. Классификационный анализ данных. / Труды Международной конференции по проблемам управления. Том 1. – М.: СИНТЕГ, 1999. - С. 62-67.
7. Дорофеюк А.А., Покровская И.В., Чернявский А.Л. Экспертные методы анализа и совершенствования систем управления / Автоматика и телемеханика. 2004, №10. -С. 172 – 188.
8. Дорофеюк Ю.А. Структурные методы прогнозирования в крупномасштабных системах управления. / Управление развитием крупномасштабных систем (MLSD'2008). Материалы Второй международной конференции. Том 1. / -М.: ИПУ РАН, 2008. –С. 32-35.

Дорофеюк Юлия Александровна. Научный сотрудник Института проблем управления им. В.А. Трапезникова РАН (ИПУ РАН). Окончила Московский государственный институт электроники и математики, Технический университет (МГИЭМ ТУ) в 2007 году. Автор 61 печатной работы. Область научных интересов: интеллектуальный анализ информации, экспертно-классификационные методы анализа данных, распознавание образов, искусственный интеллект. E-mail: dorofeyuk_julia@mail.ru.