

Д.т.н., проф. Е.Г. Жиляков, В.И. Ломазова,  
д.ф.-м.н., доц. В.А. Ломазов (Белгородский ГУ)

E.G. Zhilyakov, V.I. Lomazova, V.A. Lomazov

**КОМПЬЮТЕРНАЯ КЛАСТЕРИЗАЦИЯ СОВОКУПНОСТИ  
АДДИТИВНЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ  
ВЗАИМОСВЯЗАННЫХ ПРОЦЕССОВ**

**COMPUTER CLUSTERING OF ADDITIVE MATHEMATICAL  
MODELS OF INTERACTING PROCESSES**

*Рассмотрено модельное представление взаимодействующих процессов. Предложена методика моделирования, предусматривающая построение совокупности математических моделей, различающихся степенью учета взаимного влияния процессов, и последующую эволюционную кластеризацию этой совокупности на основе выбранной метрики.*

*Keywords: complex system, interacting processes, mathematical model, genetic algorithm, clustering, classifying*

При исследовании сложных систем [1], относящихся к различным предметным областям, возникает необходимость учета взаимного влияния протекающих в этих системах (социально-экономических, научно-образовательных, химико-технологических, производственно-экономических и т.д.) процессов.

Закон функционирования (функциональная модель)  $Z$  системы  $S$  представляет собой набор соотношений, описывающих процесс  $P$  - изменение состояния системы (набор  $I$  значений свойств  $z_j$  всех  $J$  элементов  $e_j$  системы;  $j=1,2,\dots,J$ ;  $i=1,2,\dots,I$ ) с течением времени  $t$ :

$$Z = Z(t) \quad (1)$$

Однако при функциональном описании сложной системы естественно выделить функционирование отдельных подсистем  $S_r$ :  $Z_r(t)$ ,  $r=1,2,\dots,R$  в виде взаимосвязанных процессов  $P_r$ , описываемых системой связанных между собой соотношений

$$Z_r(t) = F_r(Z_1, Z_2, \dots, Z_R, t), \quad r=1,2,\dots,R \quad (2)$$

Связь между отдельными соотношениями системы (2) проявляется в наличии характеристик процессов функционирования всех подсистем  $S$  в качестве аргументов в функционале  $F_r$ .

Ограничимся рассмотрением моделей взаимосвязанных про-

цессов, в которых воздействие на каждый процесс со стороны других процессов представляется аддитивным способом, соответствующим сепарабельности функционала  $F_r$ , т.е.

$$Z_r(t) = F_r(Z_r, t) + \sum_{k \neq r}^R a_{rk} F_{rk}(Z_k, t), \quad r, k = 1, 2, \dots, R \quad (3)$$

В соотношениях (3) коэффициенты  $a_{rk}$  учитывают степень воздействия процесса  $P_k$  на процесс  $P_r$ . В рассматриваемом здесь простейшем случае эти коэффициенты полагаются бинарными и принимающими значения из множества  $\{0$  (неучет влияния  $S_k$  на  $S_r$ ),  $1$  (учет влияния  $P_k$  на  $P_r$ ) $\}$ . Таким образом, в зависимости от набора значений коэффициентов  $a_{rk}$  система соотношений (3) описывает одну из моделей процесса  $P$ , нужным в рамках конкретного исследования образом учитывающую взаимосвязь между процессами.

Основной проблемой математического моделирования является построение (селекция из имеющихся альтернатив) наиболее удобной (учетом обеспечения необходимого уровня адекватности) для проведения дальнейших исследований модели объекта (явления, процесса) [2]. Один из возможных подходов к решению этой проблемы в рамках исследования сложных систем (и взаимосвязанных процессов) рассмотрен в [3]. Однако большая мощность множества моделей многих сложных систем приводит к большой вычислительной сложности предложенных в [3] алгоритмов селекции. Для уменьшения пространства поиска (сжатия данных) алгоритмов селекции целесообразно проведение классификации (кластеризации) совокупности моделей, что позволяет построить иерархическую процедуру селекции: сначала выбор класса, затем выбор модели в выбранном классе (кластере). Процедура классификации (кластеризации) также может носить иерархический характер (процедура таксономии), когда крупные кластеры дробятся на более мелкие, те в свою очередь дробятся ещё мельче, и т. д., результатом чего является древообразная иерархическая структура. При этом разбиение совокупности объектов на группы на основе схожести признаков для объектов одной группы и отличий между группами в рамках кластеризации (обучения без учителя) отличается от классификации (обучения с учителем) тем, что число групп, как правило, неизвестно заранее и устанавливается в соответствии с некото-

рым субъективным критерием. Результат кластеризации существенно зависит от метрики, выбор которой, как правило, также субъективен и определяется экспертом [4]. Процедура кластеризации, как правило, гораздо более трудоемка, чем процедура классификации, однако она в большей степени отражает структуру совокупности объектов.

Совокупность моделей можно рассматривать в качестве метрического пространства, задав расстояние (метрику) на множестве моделей, например, взяв за основу расстояние Хэмминга (Hamming distance):

$$D_H(\mathbf{M}^*, \mathbf{M}^{**}) = \frac{1}{R(R-1)} \sum_{r \neq k}^R \sum_{k=1}^R \text{abs}(a_{rk}^* - a_{rk}^{**})$$

Однако учет разных связей не одинаков по сложности и зависит от особенностей системы, вида функциональных зависимостей, типа решаемой задачи и используемого метода ее решения, а также требуемой точности решения. Поэтому в качестве расстояния между моделями  $\mathbf{M}^*$  и  $\mathbf{M}^{**}$  предлагается использовать взвешенную метрику

$$D(\mathbf{M}^*, \mathbf{M}^{**}) = \frac{1}{R(R-1)} \sum_{r \neq k}^R \sum_{k=1}^R \beta_{rk} \text{abs}(a_{rk}^* - a_{rk}^{**})$$

где весовые коэффициенты удовлетворяют условиям:

$$\sum_{r=1}^R \sum_{k=1}^R \beta_{rk} = 1; \quad \beta_{rk} \geq 0 \quad (r, k = 1, 2, \dots, R; r \neq k)$$

и определяются экспертами. Смысл метрики может быть связан, например, с близостью моделей по степени удобства использования или по затратам ресурсов (времени работы вычислительной техники, труда специалистов и т.д.) при решении конкретных задач. Определение весовых коэффициентов целесообразно производить на основе экспертных оценок с использованием методов командного ранжирования или парных сравнений.

Расстояние между моделью  $\mathbf{M}^*$  и группой моделей  $\mathbf{G}^{**}$  определяется по формуле

$$D(\mathbf{M}^*, \mathbf{G}^{**}) = \min_{\mathbf{M}^{**} \in \mathbf{G}^{**}} D(\mathbf{M}^*, \mathbf{M}^{**})$$

Расстояние между группой моделей  $\mathbf{G}^*$  и группой моделей

$G^{**}$  определяется по формуле

$$D(G^*, G^{**}) = \min_{M^{**} \in G^{**}} D(G^*, M^{**})$$

Количество возможных разбиений множества из  $l$  элементов на  $m$  классов составляет величину  $m^l$ . Поскольку множество моделей, в различной степени учитывающих взаимодействие процессов, содержит  $2^{R(R-1)}$  элементов, то, казалось бы, простая задача построения разбиений совокупностей моделей 3-х взаимосвязанных процессов на 2 класса, приводит к возникновению более  $10^{20}$  вариантов разбиений. Отсюда следует, что применение метода перебора для кластеризации совокупности моделей взаимосвязанных процессов нецелесообразно.

Для кластеризации совокупности моделей взаимосвязанных процессов предлагается (основанная на использовании матрицы расстояний [4] и применении стандартного генетического алгоритма [5]) процедура следующего вида:

- 1) кодирование моделей в виде бинарных хромосом, определяемых коэффициентами  $a_{rk}$  ( $r, k=1, 2, \dots, R; r \neq k$ );
- 2) построение бинарных хромосом, соответствующих парам сравниваемых моделей, путем конкатенации хромосом этих моделей и построение нормализованной функции приспособленности на основе расстояния между моделями  $D$ ;
- 3) построение начальной популяции пар моделей, случайным выбором из класса моделей, описываемых соотношениями (3): турнирный (или рулеточный) отбор родительских пар;
- 4) применение генетических операторов скрещивания и мутации для получения новой популяции пар моделей;
- 5) выделение нескольких пар моделей с наименьшим расстоянием друг от друга; группировка (объединение выделенных пар в группы) и переход на п.4 для пар групп моделей.

Процедура кластеризации прекращается после организации групп с заданным количеством моделей, после чего группы, содержащие более одной модели, объявляются классами и последующее разбиение совокупности моделей производится на основе процедуры классификации.

Как и любой эвристический метод случайного поиска, предлагаемая процедура генетической кластеризации моделей не гарантирует нахождение оптимального решения, но представляется более

эффективный, чем гарантирующий точное решение метод полного перебора возможных разбиений.

*Работа выполнена при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 гг. государственный контракт № 02.740.11.5128 от 09 марта 2010 г.*

### Литература

1. Бусленко Н.П. К теории сложных систем. – "Изв. АН СССР", Техническая кибернетика, 1963, № 5, с.3-11.
2. Самарский А.А. и Михайлов А.П. Математическое моделирование: Идеи. Методы. Примеры. 2-е изд. М., Физматлит, 2005. 320 с.
3. Ломазова В.И. и Ломазов В.А. Формализация выбора математических моделей связанных полей при автоматизации исследований. – "Информационные системы и технологии", 2010, № 3, с.101-106.
4. Журавлев Ю.И., Рязанов В.В. и Сенько О.В. Распознавание. Математические методы. Программная система. Практические применения. М., Фазис, 2006. 159 с.
5. Гладков Л.А. Курейчик В.В., Курейчик В.М. и Сороколетов В.П. Биоинспирированные методы в оптимизации. М.: , Физматлит, 2009. 384 с.

*Статья поступила 12 10 2010*