КЛАССИФИКАЦИЯ ЭЛЕМЕНТОВ СЛОЖНОЙ СИСТЕМЫ С ИНТЕРВАЛЬНО-ЗНАЧНЫМИ ХАРАКТЕРИСТИКАМИ

Д. А. Вятченин, А. В. Доморацкий, Д. И. Новиков,
А. В. Юодялис (Минск, Беларусь)

Одной из важнейших особенностей имитационного моделирования как метода исследования сложных систем является то обстоятельство, что оно обладает такими неоспоримыми преимуществами, как возможность исследования системы на различных уровнях ее детализации и оценка ее характеристик в определенные моменты времени. В целом процесс имитационного моделирования условно можно разделить на следующие шесть этапов [3]:

- содержательное описание и структуризация объекта моделирования; построение концептуальной модели сложной системы;
- постановка задачи имитационного моделирования, формирование предварительной структуры объекта, целей и критериев их достижения; разработка моделирующего алгоритма и построение имитационной модели;
- формализация и описание постановки задачи;
- отладка и корректировка модели;
- проведение имитационных экспериментов, анализ результатов и выбор наилучшей схемы функционирования объекта по имитационной модели;
- внедрение полученных результатов в практику деятельности объекта.

Первые два этапа процесса имитационного моделирования включают в себя анализ исследуемой системы, подразумевающий ее декомпозицию на элементы, допускающие их математическое описание, и определение связей между элементами. Множество n элементов, составляющих сложную систему, образует совокупность объектов $X = \{x_1, \ldots, x_n\}$, каждый из которых может характеризоваться m признаками, так что данные о них могут быть представлены в виде матрицы $\tilde{X}_{n\times m} = [\tilde{x}_i^t], i = 1, \ldots, n, t = 1, \ldots, m$, i-я строка которой представляет собой вектор $\tilde{x}_i = \{\tilde{x}_i^1, \ldots, \tilde{x}_i^m\}$, полностью описывающий i-й элемент системы $x_i \in X$, и может интерпретироваться как точка в признаковом пространстве $I^m(X)$. Матрица $\tilde{X}_{n\times m} = [\tilde{x}_i^t]$ в специальной литературе [4] именуется матрицей «объект–свойство» и традиционно используется в задачах классификации для представления данных об исследуемой совокупности объектов $X = \{x_1, \ldots, x_n\}$.

Необходимо, однако, указать, что характеристики отдельных элементов сложных систем могут быть описаны не точными значениями, а интервалами – в работе [1] подобного рода характеристики названы динамическими.

В качестве иллюстративного примера системы, элементы которой описываются динамическими признаками, можно рассмотреть систему массового обслуживания, схематично изображенную на рис. 1, состоящую из одного прибора обслуживания, включающего накопитель заявок и канал обслуживания, в свою очередь, включающего n агрегатов $x_1, \ldots, x_n$, выполняющих одну и ту же операцию, но в течение различных интервалов времени. Кроме того, каждый из агрегатов $x_i \in X$ допускает, что в очереди на обслуживание в произвольный момент времени может находиться не более $a^i_{j\max}$. Иначе говоря, символом a^i_j обозначается количество заявок, могущих находиться в очереди на обслуживание агрегатом $x_i \in X$ безотносительно к рассматриваемому моменту времени, и, как следствие, также колеблющемуся в некотором интервале $[a^i_{j\min} = 0, a^i_{j\max}]$. Совершенно естественно, что оптимальная работа всей линии, состоящей из последовательности подобного рода приборов, возможна при максимально быстрой обработке каждой заявки каналом обслуживания, что, в свою очередь, зависит от интервала времени $[\text{€}_i^{t_1\min}, \text{€}_i^{t_1\max}]$, затрачиваемого на обслуживание некоторой заявки соответствующим агрегатом $x_i \in X$.

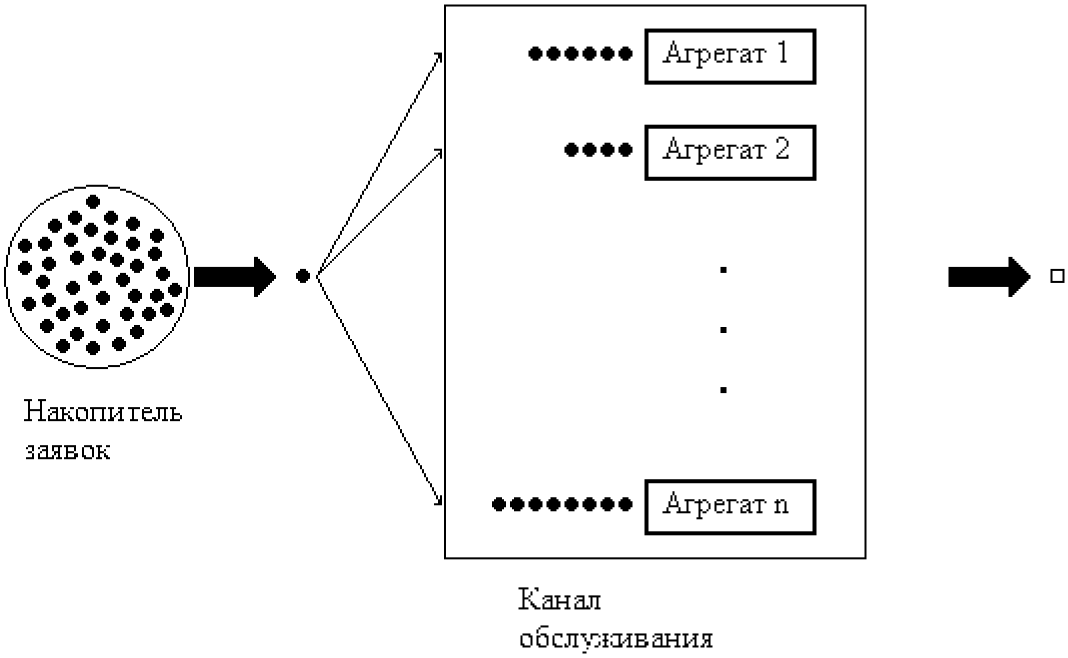

Рис. 1. Пример системы массового обслуживания

Так как каждый агрегат $x_i, i = 1,\dots,n$ канала обслуживания характеризуется интервально-значными признаками $[\text{€}_i^{t_1\min}, \text{€}_i^{t_1\max}]$и $[a^i_{j\min} = 0, a^i_{j\max}]$, то с целью определения агрегата, приоритетного для обслуживания поступающей в канал заявки, все агрегаты $X = \{x_1,\dots,x_n\}$ должны быть расклассифицированы по указанным признакам, и совершенно естественно, что подобного рода классификация должна быть упорядоченной. Таким образом, задача автоматической классификации элементов системы, описываемых динамическими признаками с целью формирования структуры объекта моделирования, относится к задачам динамической кластеризации.

В задачах динамической кластеризации признаки €^{t_1}, $t_1 = 1,\dots,m_1$ объектов $x_i \in X$ могут принимать значения в непрерывном интервале безотносительно к моменту измерения соответствующей характеристики объекта, так что каждый признак €^{t_1}, $t_1 = 1,\dots,m_1$ для объекта x_i, $i = 1,\dots,n$ представляет собой интервал значений $[\text{€}_i^{t_1\min}, \text{€}_i^{t_1\max}]$. Кластерная структура исследуемой совокупности, состоящей из подобных объектов, также является динамической и зависит от значений признаков в момент классификации. На содержательном уровне задача построения устойчивой кластерной структуры может быть сформулирована следующим образом: найти такое априори неизвестное число c областей признакового пространства $\Re^m$, в которых отображаются кластеры, при различных значениях принимаемых объектами исследуемой совокупности X признаков $\hat{x}^{t_1}$, $t_1 = 1,\ldots,m_1$, варьирующихся в интервале $[\hat{x}_i^{t_1 \min}, \hat{x}_i^{t_1 \max}]$.

Для решения задач динамической кластеризации удобным средством является эвристический D-AFC(c)-алгоритм возможностной кластеризации [8], строящий так называемое распределение по априори заданному числу c нечетких α-кластеров, представляющее собой частный случай возможностного разбиения, а в случае априорной неизвестности значения c можно воспользоваться его модификацией, получившей обозначение D-PAFC-алгоритма [10]. Отличительной особенностью указанных кластер-процедур является то, что матрицей входных данных для них является матрица нечеткой толерантности, причем независимо от типа соответствующего нечеткого отношения [1]. Таким образом, главная задача заключается в предварительной обработке матрицы вида «объект–признак», в которой значения признаков классифицируемых объектов описываются интервалами, с целью построения матрицы нечеткой толерантности. Для данной цели оказывается удобным прибегнуть к аппарату интервально-значных нечетких множеств [5].

Если X – некоторый универсум, то определенное на X нечеткое множество A с функцией принадлежности, принимающей значения в интервале, задается двумя функциями принадлежности: $\underline{\mu}_A(x_i)$, определяющей нижнее значение интервала значений принадлежности $x_i \in X$, и $\overline{\mu}_A(x_i)$, задающей верхнее значение, так что $0 \le \underline{\mu}_A(x_i) \le \overline{\mu}_A(x_i) \le 1$, и интервально-значное нечеткое множество A определяется как $A = \{x_i, \mu_A(x_i) = [\underline{\mu}_A(x_i), \overline{\mu}_A(x_i)] \mid x_i \in X, \underline{\mu}_A(x_i), \overline{\mu}_A(x_i) \in [0,1]\}$. Очевидно, что каждое обычное нечеткое множество A может быть представлено в виде интервально-значного нечеткого множества с совпадающими для каждого элемента $x_i \in X$ нижним и верхним значениями интервала значений принадлежности, то есть $\underline{\mu}_A(x_i) = \overline{\mu}_A(x_i)$, $\forall x_i \in X$.

Обозначая объекты исследуемой совокупности символами x_i, $i = 1,\ldots,n$, а признаки – соответственно символами $\hat{x}^{t_1}$, $t_1 = 1,\ldots,m_1$, матрица «объект–признак» $X_{n \times m_1} = [\hat{x}_i^{t_{1\min}(t_{1\max})}]$, где $\hat{x}_i^{t_{1\min}(t_{1\max})} = [\hat{x}_i^{t_1 \min}, \hat{x}_i^{t_1 \max}]$, может быть обработана с помощью обобщенной нормализации

$$x_i^{t_{1\min}(t_{1\max})} = \frac{\hat{x}_i^{t_{1\min}(t_{1\max})}}{\max\limits_{i, t_{1\max}} \hat{x}_i^{t_{1\min}(t_{1\max})}} \tag{1}$$

или обобщенной унитаризации

$$x_i^{t_{1\min}(t_{1\max})} = \frac{\hat{x}_i^{t_{1\min}(t_{1\max})} - \min\limits_{i, t_{1\max}} \hat{x}_i^{t_{1\min}(t_{1\max})}}{\max\limits_{i, t_{1\max}} \hat{x}_i^{t_{1\min}(t_{1\max})} - \min\limits_{i, t_{1\max}} \hat{x}_i^{t_{1\min}(t_{1\max})}}, \tag{2}$$

где $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$, предложенных в [1], вследствие чего каждый объект x_i может интерпретироваться как интервально-значное нечеткое множество на универсуме признаков, с функцией принадлежности $\mu_{x_i}(x^{t_1}) = [\underline{\mu}_{x_i}(x^{t_1}), \overline{\mu}_{x_i}(x^{t_1})]$, $i = 1,\ldots,n$, где $\underline{\mu}_{x_i}(x^{t_1}) = \mu_{x_i}(x^{t_1 \min})$ и $\overline{\mu}_{x_i}(x^{t_1}) = \mu_{x_i}(x^{t_1 \max})$.

Для интервально-значных нечетких множеств Х. Юу и Х. Юаном в [5] был определен ряд мер близости. В рассматриваемом случае, при представлении объектов исследуемой совокупности $x_i, x_j \in X$ как интервально-значных нечетких множеств x_i и x_j, $i, j = 1, \ldots, n$, определенных на универсуме признаков, одна из мер сходства, введенных в [5], примет вид

$$s_{JY(IVFS)(1)}(x_i, x_j) = 1 - \frac{1}{\sqrt[\lambda]{m_1}} \sqrt[\lambda]{\sum_{t_1=1}^{m_1} \left| \frac{\underline{\mu}_{x_i}(x^{t_1}) + \overline{\mu}_{x_i}(x^{t_1})}{2} - \frac{\underline{\mu}_{x_j}(x^{t_1}) + \overline{\mu}_{x_j}(x^{t_1})}{2} \right|^{\lambda}}, \qquad (3)$$

где $i, j = 1, \ldots, n$, $t_1 = 1, \ldots, m_1$, и λ – параметр, такой, что $1 \le \lambda < \infty$. Таким образом, значения коэффициентов близости $s_{JY(IVFS)(1)}(x_i, x_j)$, полученные с помощью выражения (3), будут представлять собой элементы матрицы нечеткой толерантности $T_{n \times n} = [\mu_T(x_i, x_j)]$, являющуюся, как указывалось выше, матрицей исходных данных для D-AFC(c)-алгоритма.

В свою очередь, нормализованное расстояние Хемминга между интервально-значными нечеткими множествами x_i и x_j, предложенное П. Гжегожевским [6], в рассматриваемом случае принимает вид

$$l_{N(H)(IVFS)}(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} \max \left\{ \left| \underline{\mu}_{x_i}(x^{t_1}) - \underline{\mu}_{x_j}(x^{t_1}) \right|, \left| \overline{\mu}_{x_i}(x^{t_1}) - \overline{\mu}_{x_j}(x^{t_1}) \right| \right\}, \qquad (4)$$

применение которого к интервально-значным нечетким множествам x_i, $i = 1, \ldots, n$ позволяет построить матрицу нечеткого отношения несходства $I_{n \times n} = [\mu_I(x_i, x_j)]$. В свою очередь, операция дополнения

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), \ \forall x_i, x_j, \ i, j = 1, \ldots, n, \qquad (5)$$

примененная к матрице $I_{n \times n} = [\mu_I(x_i, x_j)]$, также дает в результате матрицу нечеткой толерантности $T_{n \times n} = [\mu_T(x_i, x_j)]$,

С другой стороны, так как интервально-значные нечеткие множества представляют собой частный случай нечетких множеств типа 2, то оказывается возможным для решения указанной выше задачи воспользоваться подходом к предварительной обработке данных, изложенным в [1], [9]. Вычислительные эксперименты, проведенные с тестовыми интервально-значными данными М. Сато-Илик и Л. Джейна [7], показали преимущество подхода, основанного на методе, предложенном в [9], по сравнению с выражениями (3) и (4), в силу более высокой содержательной осмысленности результатов, так как при использовании функций расстояния между нечеткими множествами типа 2 [9] полученные нечеткие кластеры представляют собой субнормальные нечеткие множества [2].

В качестве предварительного вывода проведенного исследования можно отметить, что в условиях, когда характеристики элементов системы описываются интервалами, для проведения классификации соответствующих элементов с целью формирования предварительной структуры объекта имитационного моделирования возможным является использование обобщений расстояний между нечеткими множествами для случая нечетких множеств типа 2.

Литература

1. **Вятченин Д. А.** Динамические признаки в задачах классификации моделей // Вторая всероссийская научно-практическая конференция по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование. Теория и практика. ИММОД-2005». Сб. докладов. СПб.: ФГУП ЦНИИ технологии судостроения, 2005. Т. 1. С. 96–100.
2. **Вятченин Д. А**. Виды нечетких α-кластеров // Вести Института современных знаний. 2008. № 4. С. 95–101.
3. **Кобелев Н. Б.** Введение в общую теорию имитационного моделирования. Пособие для разработчиков имитационных моделей и их пользователей. М.: Принт-Сервис, 2007. 126 с.
4. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / **С. А. Айвазян, В. М. Бухштабер, И. С. Енюков, Л. Д. Мешалкин;** Под ред. С. А. Айвазяна. М.: Финансы и статистика, 1989. 607 с.
5. **Ju H., Yan X.** Similarity measures on interval-valued fuzzy sets and application to pattern recognition // Fuzzy Information and Engineering / Ed. by D.Y. Cao. Berlin, Springer-Verlag, 2007. P. 875–883.
6. **Grzegorzewski P.** Distances between intuitionistic fuzzy sets and/or on interval-valued fuzzy sets based on Hausdorff metric // Fuzzy Sets and Systems. 2004. Vol. 148. P. 319–328.
7. **Sato-Ilic M., Jain L. C.** Innovations in Fuzzy Clustering: Theory and Applications. Heidelberg: Springer-Verlag, 2006. 152+IX p.
8. **Viattchenin D. A.** A new heuristic algorithm of fuzzy clustering // Control & Cybernetics. 2004. Vol. 33. P. 323–340.
9. **Viattchenin D. A.** An outline for a heuristic approach to possibilistic clustering of the three-way data // Journal of Uncertain Systems. 2009. Vol. 3. P. 64–80.
10. **Viattchenin D. A.** An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality // Journal of Information and Organizational Sciences. 2009. Vol. 33. P. 205–217.