

УДК 004.942

КРИТЕРИИ АДЕКВАТНОСТИ МАТЕМАТИЧЕСКОЙ МОДЕЛИ ПРЕДМЕТНОЙ ОБЛАСТИ

Е. В. Малахов

Кандидат технических наук, доцент, заведующий
кафедрой
Кафедра информационных систем в менеджменте*
Контактный тел.: (0482)734-84-17
E-mail: opmev@mail.ru

Г. Н. Востров

Кандидат технических наук, доцент, заведующий
кафедрой
Кафедра прикладной математики и информационных
технологий в бизнесе*
Контактный тел.: (0482)7348478
E-mail: vostrov@gmail.com

*Одесский национальный политехнический университет
пр. Шевченко, 1, г. Одесса, Украина, 65044

Розглянуто ймовірності появи в предметних областях сутностей, які описуються випадковими величинами і процесами, номінальними або змішаними атрибутами. Запропоновано критерії адекватності моделей предметних областей, що залежать від належності атрибутів об'єктів до відповідних класів

Ключові слова: модель, предметна область, адекватність, атрибути об'єктів

Рассмотрены вероятности появления в предметных областях сущностей, описываемых случайными величинами и процессами, номинальными или смешанными атрибутами. Предложены критерии адекватности моделей предметных областей, зависящие от принадлежности атрибутов объектов к соответствующим классам

Ключевые слова: модель, предметная область, адекватность, атрибуты объектов

Entities appearance probabilities in the subject domains which are described by stochastic quantities and processes, nominal or mixed attributes are considered. The subject domains models adequacy criteria depending on a membership of object's attributes to corresponding classes are offered

Keywords: model, subject domain, adequacy, attributes of objects

Введение

В последнее время в литературе, в том числе, учебной [1], достаточно большое внимание уделяется вопросам идентификации объектов в условиях априорной и апостериорной неопределённости статистических характеристик их свойств. При этом в основном рассматриваются модели динамических систем и методы их параметрической идентификации, в частности, робастные (устойчивые к наличию априорной неопределённости), но мало затрагиваются проблемы снижения самой неопределённости.

Одним из средств устранения (снижения) неопределённости являются базы данных, в которых накапливается информация, отражающая динамику изменения свойств предметных областей, их объектов и характера взаимодействия между ними. Основной проблемой их создания является обеспечение максимальной адекватности соответствующим предметным областям ещё на первых этапах проектирования, зависящей от полноты описания предметных областей формальными средствами [2].

Оценка адекватности моделей сущности в зависимости от шкалы измерения её свойств

В работе [3] отмечено, что оценка адекватности модели ПрО должна начинаться с оценки адекватности моделей универсальной сущности и её свойств. Для такой оценки может потребоваться определять характеристики свойств сущностей и на их основании оценивать и ранжировать свойства по степени их важности (информативности). Проверка адекватности полученных характеристик выполняется путём прогнозирования значений свойств универсальных сущностей или их проекций на ПрО [3, 4] и сравнения значений атрибутов спрогнозированного экземпляра сущности со значениями атрибутов экземпляра, который будет наблюдаться при (K+1)-м испытании [3].

Для каждого свойства универсальной сущности, приближённо описываемой конечным множеством свойств, строится метод измерения степени его проявления в одной из шкал: интервальной, балльной, порядковой или номинальной. Очевидно, что в пер-

вых трёх шкалах измеряются свойства, которые могут быть представлены случайными величинами и случайными процессами.

От принадлежности свойств к этим шкалам, степени их проявления и используемых мер информационной ценности свойств [4] будет зависеть критерий адекватности как g -й универсальной сущности, так и ПрО, которой принадлежит проекция этой сущности.

Если свойства сущности представляются *случайными величинами и случайными процессами*, то критерий адекватности приближённой математической модели ПрО $D(t)$ имеет вид:

$$C(M(D(t)), D(t)) = \sum_{r=1}^M W_r(E_r(t)) \sum_{S=1}^n \sum_{j=1}^{m_r} I(A_{jr}(t), \bar{A}_{jr}(t)) (A_{jr}^{K+S}(t) - \bar{A}_{jr}^{K+S}(t))^2, \quad (1)$$

где M — количество проекций универсальных сущностей на D , включенных в модель; n — количество последовательных этапов проверки степени адекватности математической модели ПрО её действительной реальности; $W_r(E_r(t))$ — функция степени важности g -й сущности как элемента моделирования ПрО; m_r — количество атрибутов, выделенных для приближенного описания g -й сущности с необходимой точностью; K — количество экземпляров g -й сущности (кортежей в реляционном отношении, представляющем g -ю сущность); $I(A_{jr}(t), \bar{A}_{jr}(t))$ — информационная мера атрибута $A_{jr}(t)$ (для реальных и спрогнозированных значений) из множества наиболее информативных атрибутов g -й сущности [4]; переменная S определяет номер шага прогнозирования.

Прогнозирование осуществляется для такого количества статистических испытаний, на основе которого можно оценить значение критерия адекватности, путем наблюдения последовательно получаемых экземпляров g -й сущности после K -го испытания, которым завершилось построение математической модели ПрО.

Критерий адекватности для всех универсальных сущностей, которые описываются *только номинальными атрибутами*, имеет вид

$$C(M_H(D(t)), D(t)) = \frac{1}{M_H} \sum_{r=1}^{M_H} W(E_r(t)) \cdot I(\bar{B}_r^F, \bar{B}_r) \cdot \sum_{j=1}^{L+L_1} |\bar{P}_F(\bar{B}_j^F) - \bar{P}(\bar{B}_j)| \quad (2)$$

где $W(E_r(t))$ — весовая функция оценки важности g -й сущности $E_r(t)$ со множеством наиболее существенных свойств, $I(\bar{B}_r^F, \bar{B}_r)$ — мера количества информации реальной классификации экземпляров g -й сущности в сравнении с системой классов, полученных в результате прогнозирования.

Рассмотрим случай g -й сущности, которая описывается *смешанным множеством атрибутов*. Это означает, что некоторое подмножество атрибутов $A_i = \{A_{1r}, \dots, A_{qr}\}$ представляет собой дискретные или непрерывные случайные процессы. Его дополнение относительно выделенного множества наиболее информативных атрибутов представляет множество

атрибутов, которые измеряются в номинальной шкале $A_n = \{A_{(q+1)r}, \dots, A_{mr}\}$. Очевидно, что

$$A_i \cap A_n = \emptyset \text{ и}$$

$$A_i \cup A_n = \{A_{1r}(t), A_{2r}(t), \dots, A_{qr}(t), \dots, A_{mr}(t)\}. \quad (3)$$

В этом случае математическую модель g -й сущности можно представлять в виде объединения двух моделей. Одна строится над множеством атрибутов, моделируемых с помощью случайных процессов, а другая математическая модель строится над множеством номинальных атрибутов. В любом случае будем оценивать качество адекватности математических моделей на основе оценки эффективности прогнозирования по каждому из этих двух подмножеств в виде суммы адекватности моделирования по каждому из них. В этом случае адекватность математического моделирования ПрО по совокупности универсальных сущностей с приведенной структурой атрибутов может оцениваться с помощью выражения

$$C(M_{Hi}(D(t)), D(t)) = \frac{1}{M_{Hi}} \sum_{r=1}^{M_{Hi}} W(E_r(t)) \sum_{S=1}^n \sum_{j=1}^{m_r} I(A_{jr}^{K+S}(t) - \bar{A}_{jr}^{K+S}(t))^2 + I(\bar{B}_r^F, \bar{B}_r) \sum_j^{L+L_1} |\bar{P}_F(\bar{B}_j^F) - \bar{P}(\bar{B}_j)|, \quad (4)$$

где M_{Hi} — количество важных универсальных сущностей, которые описываются смешанным множеством номинальных атрибутов и атрибутов, представляемых случайными величинами и случайными процессами.

Обобщённый критерий оценки адекватности моделирования ПрО

На основании критериев адекватности приближенной математической модели ПрО для универсальных сущностей, описываемых атрибутами, которые представляются случайными величинами и случайными процессами (1), сущностей, описываемых только номинальными атрибутами (2), и сущностей, описываемых смешанными атрибутами (4), создан общий критерий оценки адекватности моделирования ПрО через систему универсальных сущностей ее ядра и множества наиболее информативных свойств каждой из них, который может быть представлен выражением

$$C(M(D(t)), D(t)) = \frac{1}{M_i + M_H + M_{Hi}} \times (M_i C(M_i(D(t)), D(t)) + M_H C(M_H(D(t)), D(t)) + M_{Hi} C(M_{Hi}(D(t)), D(t))) \quad (5)$$

По сути, вероятности появления универсальных сущностей, которые описываются атрибутами, представляемыми случайными величинами и случайными процессами ($E_i(t)$), номинальными ($E_H(t)$) и смешан-

ными $(E_{Hi}(t))$ атрибутами, могут быть оценены с помощью выражений

$$\begin{aligned}\bar{P}(E_i(t)) &= \frac{M_i}{M_i + M_H + M_{Hi}}, \\ \bar{P}(E_H(t)) &= \frac{M_H}{M_i + M_H + M_{Hi}}, \\ \bar{P}(E_{Hi}(t)) &= \frac{M_{Hi}}{M_i + M_H + M_{Hi}},\end{aligned}\quad (6)$$

где $M_i + M_H + M_{Hi} \gg K$.

Согласно закону больших чисел или усиленному закону больших чисел [6] будут справедливы соотношения

$$\begin{aligned}\lim_{M_i \rightarrow \infty} P(|P(E_i(t)) - \bar{P}(E_i(t))| < \varepsilon) &\rightarrow 1, \\ \lim_{M_H \rightarrow \infty} P(|P(E_H(t)) - \bar{P}(E_H(t))| < \varepsilon) &\rightarrow 1, \\ \lim_{M_{Hi} \rightarrow \infty} P(|P(E_{Hi}(t)) - \bar{P}(E_{Hi}(t))| < \varepsilon) &\rightarrow 1,\end{aligned}\quad (7)$$

где $M = M_i + M_H + M_{Hi}$ — общее количество испытаний по всем универсальным сущностям, вероятность $P(E_i(t))$ получается благодаря испытаниям, а вероятность $\bar{P}(E_i(t))$ рассчитывается с помощью (6).

Аналогично, для любых $\varepsilon > 0$ и $\eta > 0$ существует M_0 такое, что для любого S при $M \in [M_0, M_0 + S]$, со-

гласно Утверждению 3 и выражению (10), отмеченным в [4], справедливы соотношения

$$\begin{aligned}P\left(\left|P_{M_i > N_{\varepsilon}}(E_i(t)) - \bar{P}(E_i(t))\right| < \varepsilon\right) &\rightarrow 1 - \eta, \\ P\left(\left|P_{M_{Hi} > N_{\eta}}(E_H(t)) - \bar{P}(E_H(t))\right| < \varepsilon\right) &\rightarrow 1 - \eta, \\ P\left(\left|P_{M_{Hi} > N_{\eta}}(E_{Hi}(t)) - \bar{P}(E_{Hi}(t))\right| < \varepsilon\right) &\rightarrow 1 - \eta.\end{aligned}\quad (8)$$

Таким образом, при недостаточной представительности совокупности выборочных данных будет сохраняться неопределенность в представлениях о законе распределения каждого атрибута универсальной сущности, но в соответствии со шкалой его измерения.

Выводы

Как итог необходимо отметить, что эффективность моделирования ПрО может быть существенным образом улучшена, если более точно учитывать особенности шкал измерения с одновременным усовершенствованием работы экспертов, лиц, принимающих решения, и пользователей, которые выбирают наилучшие варианты индивидуальных задач. Кроме того, из приведенного критерия адекватности и метода его оценки следует, что отсутствие полноты системы БД и наличие неопределенностей в представлении динамических закономерностей свойств универсальных сущностей могут обуславливать недостаточный уровень представительности совокупности накопленных в них данных.

Литература

1. Семенов А.Д. Идентификация объектов управления: Учебн. пособие. / А.Д. Семенов, Д.В. Артамонов, А.В. Брюхачев. Пенза: Изд-во Пенз. гос. ун-та, 2003. 211 с.
2. Малахов Е.В. Оценка степени адекватности баз данных как информационных моделей предметных областей [текст] // Тр. Одес. политехн. ун-та. – 2004. – Вып. 1(21). – С. 82-86.
3. Малахов Е.В. Определение и оценка характеристик свойств универсальных сущностей предметных областей // Е.В. Малахов, Г.Н. Востров / Электромашиностроение и электрооборудование. – К.: «Техника», 2010 – № 76. – с. 43-49.
4. Малахов Е.В. Методы определения степени важности свойств сущностей предметных областей [текст] // Е.В. Малахов, Г.Н. Востров, М.Г. Микулинская / Холодильная техника и технология. – Одесса, 2010. – Вып. 4(126). – С. 73-77.
5. Малахов Е.В. Манипулирование метамоделями предметных областей [текст] // Восточно-европейский журнал передовых технологий. – Харьков, 2007. – Вып. 5/3(29). – С. 6-10.
6. Гнеденко Б.В. Курс теории вероятностей 7-е изд., испр. М: УРСС. 2009. 318 с.