

ОЦЕНИВАНИЕ ПРОГРАММНЫХ СРЕДСТВ ПО МНОЖЕСТВУ ПРИЗНАКОВ

А. С. БУРЫЙ¹, Е. В. МОРИН²

¹*Всероссийский научно-исследовательский институт классификации, терминологии и информации по стандартизации и качеству („СТАНДАРТИНФОРМ“), 117418, Москва, Россия
E-mail: a.s.burij@gostinfo.ru*

²*„Ростест-Москва“, 117418, Москва, Россия
E-mail: morin.evg.vas76@mail.ru*

Рассмотрена задача оценивания программных средств по результатам тестирования в ходе их сертификации. Сформулированный подход основывается на взаимосвязи критериев, факторов, метрик и оцениваемых признаков программных средств и может быть использован при формировании баз данных результатов испытаний разнотипного программного обеспечения. На предварительном этапе целесообразно проводить анализ формируемых множеств, обеспечивая проверку структурных свойств (наблюдаемости, декомпозируемости, агрегируемости) измерительного комплекса испытательного центра (лаборатории). Применяемые оценки функции полезности могут быть использованы для экспресс-анализа программной продукции, а также формирования соответствующих баз данных испытательных подразделений.

Ключевые слова: *оценивание программного средства, множество признаков, шкалы признаков, функция полезности*

Введение. Оценке состояния объектов исследования практически всегда предшествует этап определения параметров, поддающихся измерению. Контролируемыми параметрами программных средств (ПС) могут выступать время выполнения вычислений (отдельных процедур), потребляемые или обеспечиваемые ресурсы (оперативная или долговременная память) и т.п. Арсенал средств измерений и процедур переработки результатов измерений постоянно расширяется. С традиционными методами статистического анализа все чаще комбинируют методы экспертного оценивания, методы интеллектуального анализа данных, методы машинного обучения и ряд других [1].

Как известно, качество программного обеспечения — пригодность программного продукта для обеспечения эффективности, производительности, безопасности и удовлетворенности конкретного пользователя*.

В соответствии с международным стандартом ISO 9126:1993 выделены шесть основных характеристик качества программ: функциональная пригодность, надежность, применимость, эффективность, сопровождаемость и переносимость**. В стандарте понятие надежности

* ГОСТ Р ИСО/МЭК 25040-2014. Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Процесс оценки. М.: Стандартинформ, 2014.

** ГОСТ Р ИСО/МЭК 9126-93. Информационная технология. Оценка программной продукции. Характеристики качества и руководство по применению. М.: ИПК Издательство стандартов, 2003.

вводится как обеспечиваемая комплексом программ достаточно низкая вероятность отказа в процессе функционирования.

При оценке комплексных показателей, каким является, например, надежность программных средств, требуется знание множества статистических данных, ряд которых не всегда доступен исследователю либо требует большого числа дополнительных испытаний. Это зачастую не позволяет адекватно проводить оценку надежности различными методами, чтобы обосновать выбор программного продукта как в ходе разработки и применения программных комплексов [2], так и в ходе эксплуатации в составе аппаратно-программных средств сложных динамических систем различного назначения [3, 4].

При проведении сертификации программных средств не менее дорогостоящей операцией, наряду с затратами на создание самого программного обеспечения, является разработка комплексных измерительных методик тестирования и шкал измерения показателей качества с учетом специфики целевого применения программного обеспечения. Необходимость контроля конечного продукта по набору признаков (свойств) фактически подводит к необходимости применения методов многокритериального принятия решений [5].

Постановка задачи. Формально задачу исследования можно представить как формирование методики проведения испытаний в ходе сертификации программных средств с целью составления паспорта (образа) качества программного продукта.

Будем характеризовать объект испытания (программное средство) кортежем множеств, образующих некоторую систему:

$$S = \langle T, X, W, U, Q \rangle, \quad (1)$$

где $T = \{T_1, T_2, \dots, T_h\}$ — множество типов тестируемых программных средств (прикладные ПС, ПС обслуживания, операционные системы и др.); $X = \{x_1, \dots, x_m\}$ — множество состояний ПС; $W = \{w_1, \dots, w_n\}$ — множество контролируемых признаков (показателей качества ПС); $U = \{u_1, \dots, u_r\}$ — множество моделей тестирования, применяемых на различных этапах испытания программных средств; $Q = \{q_1, \dots, q_n\}$ — множество правил отображения состояний в признаки, т.е. $Q: X \rightarrow 2^{|W|}$, где $2^{|W|}$ — число возможных подмножеств множества W мощностью $|W|$.

Множество признаков для k -го типа ПС образует вектор вида

$$\phi_w(T_k) = \{\phi_{w_j}(T_k)\}, \quad j = \overline{1, n}, \quad k = \overline{1, h}. \quad (2)$$

С другой стороны, значение w_j определяется измеренными параметрами k -го типа программного средства, т.е. существует функция принадлежности $\phi_{w_j}(T_k)$, которая отражает множество вещественных значений числовой оси во множество значений признаков качества в соответствующей шкале измерений.

В зависимости от количественного представления признаков можно записать

$$w_j = \phi(y_j) \in [y_{jн}, y_{jв}], \quad j = \overline{1, n},$$

где $y_{jн}$ и $y_{jв}$ — соответственно нижняя и верхняя границы интервала значений j -го признака для вещественных чисел.

При качественном задании признаков переменные, например, „Высокий“, „Средний“, „Низкий“, применяемые в экспертном оценивании, могут быть отнесены к базовому интервалу $[0, 1]$ либо характеризоваться двухэлементным множеством, в соответствии с ответами „Нет“ и „Да“, т.е.

$$\phi_{w_j}(T_k) \rightarrow [0, 1] \oplus \{0, 1\}, \quad j = \overline{1, n}, \quad k = \overline{1, h}, \quad (3)$$

где символ \oplus означает строгую дизъюнкцию.

Модель измерений признаков (свойств) тестируемых программных средств зададим в виде их отображения в пространство измеренных параметров $U_{\text{и}}$:

$$\Psi: T_k \rightarrow U_{\text{и}} \left(\left\{ u_{ijk}(w_{\text{и}}, w_{\text{э}}) \right\} \right), \quad i = \overline{1, r}; \quad j = \overline{1, n}; \quad k = \overline{1, h},$$

где элемент пространства измерений $u_{ijk}(w_{\text{и}}, w_{\text{э}}) \in U_{\text{и}}$ есть функция, характеризующая i -ю модель тестирования j -го признака качества для k -го типа ПС, в виде соответствия измеренного признака $w_{\text{и}}$ его эталонному (требуемому) значению $w_{\text{э}}$. В общем виде совокупность элементов $u_{ijk}(w_{\text{и}}, w_{\text{э}})$ представляется $[r \times n]$ -матрицей, строки которой соответствуют выбранным моделям тестирования, а столбцы — оцениваемым признакам качества.

В большинстве случаев отдельные признаки измеряются как в абсолютных шкалах, так и в шкалах отношений с различными основаниями [6], что приводит к необходимости сравнения векторов признаков вида (2) по отдельным компонентам. При сравнении измеренного j -го признака $w_{j\text{и}}$ с соответствующим эталонным значением $w_{j\text{э}}$ предпочтительность первого записывается как $w_{j\text{и}} \succ w_{j\text{э}}$, а для числовых данных этому соответствует запись $w_{j\text{и}} > w_{j\text{э}}$. Числовой характеристикой предпочтения в этом случае выступает отношение $v_j = w_{j\text{и}} / w_{j\text{э}}$.

В результате покомпонентного сравнения при экспертном оценивании каждого измеренного признака с его эталонным значением получим функцию полезности [7]:

$$\Phi(v, \beta) = \prod_{j=1}^n v_j^{\beta_j}, \quad j = \overline{1, n}, \quad (4)$$

где β_j — параметр предпочтения эксперта по результатам тестирования j -го признака, при ограничениях $\beta_j > 0$, $\sum_{j=1}^n \beta_j = 1$ для всех $j = \overline{1, n}$.

Функция $\Phi(v, \beta)$ вида (4) представляет собой среднее геометрическое взвешенное результатов сравнений признаков при их последовательном тестировании. Она характеризуется свойствами мультипликативности, монотонности и однородности по вектору признаков (2), а также инвариантна для шкал контролируемых признаков.

Таким образом, в результате последовательного тестирования программных средств заданного типа делается вывод о соответствии или несоответствии контролируемого ПС требованиям испытательной документации, а функция полезности вида (4) рассматривается как обобщенная характеристика качества ПС при фиксируемых предпочтениях.

Для организации контроля партии ПС заданного типа и объема F будем определять обобщенный показатель качества для всех элементов партии, выявляя образец, характеризуемый максимальным уровнем полезности:

$$T^* : \arg \max_f \left(\Phi_f(v_j, \beta_j) \mid j = \overline{1, n}; f = \overline{1, F}; \sum_{j=1}^n \beta_j = 1 \right). \quad (5)$$

Обоснование алгоритма оценивания объектов по множеству признаков. При подготовке и планировании тестирования ПС разрабатывается множество $U_{\text{и}} \subset U$ моделей тестирования, нормализуются шкалы признаков к виду кусочно-линейных монотонных функций полезности. Анализ технико-топологических и информационных структур при оценке возможностей испытательной лаборатории по проведению комплексного тестирования программных средств в соответствии с техническим заданием позволяет сделать предварительные

выводы о состоятельности и достоверности результатов планируемых работ. Одним из структурных показателей на этом этапе выступает наблюдаемость системы, т.е. оценка способности измерительной системы определить текущее состояние объекта по полученным измерительным данным [3, 8].

Для оценки структурных свойств рассматриваемой задачи измерения зададим матрицу отклонений M_δ , число строк в которой определяется количеством применяемых моделей тестирования, а число столбцов — количеством измеряемых признаков качества ПС.

Если ранг матрицы отклонений M_δ полностью определяется количеством линейно-независимых строк, т.е. моделей тестирования, то

$$\text{rank } M_\delta = r. \quad (6)$$

Под отклонением значений признаков будем понимать $[1 \times n]$ -вектор δ следующего вида:

$$\delta = \left\{ (w_{1и} - w_{1э}), \dots, (w_{jи} - w_{jэ}), \dots, (w_{ни} - w_{нэ}) \right\}^T, \quad j = \overline{1, n},$$

причем физически превышение измеренного значения признака $w_{jи}$ над фактическим — $w_{jэ}$ или их равенство, соответствует высказыванию, что „оцениваемое ПС по j -му признаку качества не хуже заданных требований“.

Количество моделей тестирования ограничено числом наблюдаемых признаков ПС, т.е. $r \leq n$. Следует уточнить, что величина n соответствует общему числу признаков качества, характерных для всех типов ПС, тогда как для конкретно выбранного μ -го типа ПС количество признаков $n_\mu \leq n$, т.е. не превышает максимально возможного их числа. При этом столбец матрицы M_δ можно рассматривать как совокупность измерений по признаку j .

По существу, формирование градаций шкал признаков актуально при выявлении нового типа программного средства. Для повторного тестирования, например, когда по определенной выборке тестируемых средств из общей партии однотипных объектов применим упрощенный (сокращенный алгоритм испытаний), могут быть использованы составные шкалы критериев за счет выявления агрегированных свойств объектов испытаний [9]. При этом эксперт каждый раз должен искать компромисс между временными ресурсами решения задачи, сокращая процедуры тестирования, и точностными факторами, т.е. достоверностью принятия окончательного решения о соответствии испытываемого образца требованиям на данный тип ПС [10].

Для различных типов ПС, представленных в табл. 1 („1“ — операционные системы, „2“ — ПС управления базами данных, „3“ — сервисные, а также интерфейсные программы, „4“ — ПС для научных и инженерных расчетов, „5“ — прикладные ПС для проектирования, например в составе САПР, „6“ — прикладные ПС управления технологическими процессами; „+“ — применимость данного показателя; „-“ — неприменимость; „+/-“ — определяется пользователем), рекомендуются определенные показатели качества, которым данное ПС должно удовлетворять стандарту ISO 9126:1993.

Таблица 1

Примеры показателей качества (ГОСТ 28195-99)	Применимость показателя по типу ПС					
	1	2	3	4	5	6
Устойчивость функционирования	+	+	+	-	+	+
Работоспособность	+	+	+	+	+	+
Структурность	+/-	+/-	+/-	+/-	+/-	+/-
Наглядность	+/-	+/-	+/-	-	+/-	+/-
Легкость освоения	+/-	+/-	+	+/-	+	+/-
Удобство обслуживания	+	+	+	-	+	+
Полнота реализации	+	+	+	+	+	+
Защищенность	+	+	+/-	+/-	+/-	+/-

В табл. 2 для фактора эффективности представлены весовые коэффициенты показателей качества (табл. 1) ПС. Фазы жизненного цикла (ЖЦ) обозначены буквами: для процесса разработки изделия (А — анализ; П — проектирование; Р — реализация; Т — тестирование; И — изготовление), а для этапа применения — фаза О — обслуживание/сопровождение.

Таблица 2

Критерий	Фаза ЖЦ	Весовой коэффициент показателя					
		1	2	3	4	5	6
Уровень автоматизации	А	0,25	0,25	0,25	–	0,3	0,3
	П	0,3	0,3	0,3	–	0,5	0,5
	Р, Т, И, О	0,25	0,25	0,25	–	0,3	0,3
Временная эффективность	А	0,25	0,25	0,25	1,0	0,35	0,35
	Р, Т, И, О	0,25	0,25	0,25	1,0	0,35	0,35
Ресурсоемкость	А	0,5	0,50	0,5	–	0,35	0,35
	П	0,7	0,7	0,7	–	0,5	0,5
	Р, Т, И, О	0,5	0,5	0,5	–	0,35	0,35

Весовой коэффициент 0,25 характеризует уровень автоматизации и временной эффективности, а 0,5 — ресурсоемкость.

Численный анализ изменения функции полезности вида (4) проводился по пяти признакам качества с соответствующими весовыми коэффициентами β_j . Для примера рассматривались результаты экспертного оценивания фактора „Надежность ПС“, оценки которых находятся в интервале $[0, 1]$, а в качестве признаков (j) выступали: 1) полнота обработки ошибочных ситуаций; 2) наличие средств контроля непротиворечивости входных данных; 3) наличие повторного старта с точки останова; 4) наличие средств, обеспечивающих сокращенный объем вычислений в случае ошибок; 5) вероятность безотказной работы (метод оценки расчетный) как $P = 1 - N_0/N_3$.

Результаты вычисления функции полезности в соответствии с (4) представлены в табл. 3. При этом все признаки нормировались относительно соответствующих эталонных признаков, т.е. $v_j = w_{ji}/w_{jэ}$.

Для „Оценок 1“ в соответствии с выражением (4) вычисляем соответствующую функцию полезности:

$$\Phi_1(v, \beta) = \prod_{j=1}^5 \left(w_{ji}/w_{jэ} \right)^{\beta_j} = 1,0 \times 1,03 \times 1,01 \times 1,05 \times 1,0 = 1,098.$$

Оценки 1 и 2 соответствуют тестируемым образцам ПС, у которых признаки качества незначительно превышают эталонные параметры в среднем более чем на 10 % (для „Оценок 1“), и около 7 % (для „Оценок 2“). Оценки 3 и 4 отражают незначительные колебания функции полезности, вызванные отклонением от требований отдельных показателей. Оценка 5 соответствует факту, когда не выполнены требования ни по одному из анализируемых показателей.

Для задач сравнения объектов по признакам качества иногда прибегают к сравнению относительно максимально возможных признаков [6], предварительно ранжируя все измерения по каждому признаку $\max w_{ji}$. Тогда нормирование признаков осуществляется в соответствии с $v'_j = w_{ji}/\max_j w_{ji}$, а выражение для функции полезности преобразуется к виду

$$\Phi_H(v, \beta) = \prod_{j=1}^n \left(v'_j \right)^{\beta_j}, \quad j = \overline{1, 5}.$$

Результаты вычислений помещены в табл. 3. Данный подход позволяет ранжировать все оценки по полезности, но дает менее наглядные результаты применительно к решаемой задаче многопризнаковой оценки ПС.

Таблица 3

Оценка	j					$\Phi(v, \beta)$	$\Phi_n(v, \beta)$
	1	2	3	4	5		
1	1,00	1,03	1,01	1,05	1,00	1,098	1,000
2	1,00	1,03	1,00	1,02	1,00	1,045	0,952
3	0,99	1,00	1,00	1,00	0,98	0,972	0,886
4	0,99	0,98	0,98	1,00	0,93	0,886	0,807
5	0,99	0,95	0,98	0,92	0,90	0,770	0,702

Выводы. Предложен подход к оценке качества многопризнаковых объектов на примере программных средств с использованием функции полезности. На предварительном этапе целесообразно проводить анализ формируемых множеств, обеспечивая проверку структурных свойств (наблюдаемости, декомпозируемости, агрегируемости) измерительного комплекса испытательного центра (лаборатории). Применяемые оценки функции полезности могут быть использованы для экспресс-анализа программной продукции, а также формирования соответствующих баз данных испытательных подразделений.

СПИСОК ЛИТЕРАТУРЫ

1. Михайлов Ю. Б. Научно-методические основы практических приложений количественных оценок характеристик систем. М.: Горячая линия – Телеком, 2018.
2. Иыуду К. А. Надежность, контроль и диагностика вычислительных машин и систем. М.: Высш. шк., 1989.
3. Бурый А. С. Отказоустойчивые распределенные системы переработки информации. М.: Горячая линия – Телеком, 2016.
4. Ломакин М. И., Миронов А. Н., Шестопалова О. Л. Многомодельная обработка измерительной информации в интеллектуальных системах прогнозирования надежности космических средств // Измерительная техника. 2014. № 1. С. 8—13.
5. Ногин В. Д. Принятие решений в многокритериальной среде: количественный подход. М.: Физматлит, 2005.
6. Буравлев А. И., Брезгин В. С. Оценка качества объектов по неметризуемому вектору характеристик // Вооружение и экономика. 2009. № 1. С. 26—29.
7. Фишберн П. Теория полезности для принятия решения. М.: Наука, 1978.
8. Месарович М., Такахара И. Общая теория систем. М.: Мир, 1978.
9. Шаракшанэ А. С., Халецкий А. К., Морозов И. А. Оценка характеристик сложных автоматизированных систем. М.: Машиностроение, 1993.
10. Липаев В. В. Сертификация программных средств. М.: СИНТЕГ, 2010.

Сведения об авторах**Алексей Сергеевич Бурый**

— д-р техн. наук; Всероссийский научно-исследовательский институт классификации, терминологии и информации по стандартизации и качеству („СТАНДАРТИНФОРМ“); директор департамента;
E-mail: a.s.burij@gostinfo.ru

Евгений Васильевич Морин

— канд. техн. наук; „Ростест-Москва“; первый заместитель генерально-го директора; E-mail: morin.evg.vas76@mail.ru

Поступила в редакцию
05.07.19 г.

Ссылка для цитирования: Бурый А. С., Морин Е. В. Оценка программных средств по множеству признаков // Изв. вузов. Приборостроение. 2019. Т. 62, № 10. С. 907—913.

SOFTWARE ASSESSMENT BY A SET OF INDICATORS

A. S. Buryy¹, E. V. Morin²

¹All-Russian Research Institute of Classification, Terminology and Information on Standardization and Quality (STANDARTINFORM), 117418, Moscow, Russia
E-mail: a.s.burij@gostinfo.ru

²Rostest-Moscow, 117418, Moscow, Russia
E-mail: morin.evg.vas76@mail.ru

The problem of software evaluation based on results of tests performed during its certification is considered. The formulated approach is based on relationship between criteria, factors, metrics and evaluated features of software and can be used in creation of a database of test results for different types of software. At the preliminary stage, it is advisable to analyze the formed sets, providing a check of the structural properties (observability, decomposability, aggregation) of the measuring complex of the test center (laboratory). The applied estimates of utility function can be used in express analysis of software products, as well as for formation of relevant databases of test units.

Keywords: assessment of software mean, set of indicators, features scales, utility function

REFERENCES

1. Mikhaylov Yu.B. *Nauchno-metodicheskiye osnovy prakticheskikh prilozheniy kolichestvennykh otsenok kharakteristik sistem* (Scientific and Methodological Foundations of Practical Applications of Quantitative Assessments of System Characteristics), Moscow, 2018. (in Russ.)
2. Iyudu K.A. *Nadezhnost', kontrol' i diagnostika vychislitel'nykh mashin i sistem* (Reliability, Control and Diagnostics of Computers and Systems), Moscow, 1989. (in Russ.)
3. Buryy A.S. *Otkazoustoychivyye raspredelennyye sistemy pererabotki informatsii* (Fault-Tolerant Distributed Information Processing Systems), Moscow, 2016. (in Russ.)
4. Lomakin M.I., Mironov A.N., Shestopalova O.L. *Measurement Techniques*, 2014, no. 1(57), pp. 8–15.
5. Nogin V.D. *Prinyatiye resheniy v mnogokriterial'noy srede: kolichestvennyy podkhod* (Decision Making in a Multi-criteria Environment: a Quantitative Approach), Moscow, 2005. (in Russ.)
6. Buravlev A.I., Brezgin V.S. *Vooruzheniye i ekonomika* (Armament and Economics), 2009, no. 1, pp. 26–29. (in Russ.)
7. Fishburn P.C. *Utilitu Theory for Decision Making*, NY, 1970.
8. Mesarović M., Takahara Y. *General systems theory: mathematical foundations*, Academic Press, 1975.
9. Sharakshane A.S., Khaletskiy A.K., Morozov I.A. *Otsenka kharakteristik slozhnykh avtomatizirovannykh sistem* (Assessment of the Characteristics of Complex Automated Systems), Moscow, 1993. (in Russ.)
10. Lipayev V.V. *Sertifikatsiya programmnykh sredstv* (Software Certification), Moscow, 2010. (in Russ.)

Data on authors

- Alexey S. Buryy** — Dr. Sci.; All-Russian Research Institute of Classification, Terminology and Information on Standardization and Quality (STANDARTINFORM); Head of Department; E-mail: a.s.burij@gostinfo.ru
- Evgeny V. Morin** — PhD; Rostest-Moscow; First Deputy General Director; E-mail: morin.evg.vas76@mail.ru

For citation: Buryy A. S., Morin E. V. Software assessment by a set of indicators. *Journal of Instrument Engineering*. 2019. Vol. 62, N 10. P. 907–913 (in Russian).

DOI: 10.17586/0021-3454-2019-62-10-907-913