

УДК 004.05; 004.413.5; 519.226.3
DOI: 10.15827/0236-235X.125.034-041

Дата подачи статьи: 06.08.18
2019. Т. 32. № 1. С. 034–041

Использование формулы Байеса при оценивании качества программного обеспечения по стандарту ISO/IEC 9126

Д.П. Бураков¹, к.т.н., доцент кафедры «Математика и моделирование», burakovdmitry8@gmail.com
Г.И. Кожомбердиева¹, к.т.н., доцент кафедры «Информационные и вычислительные системы»,
kgi-liizht@yandex.ru

¹ Петербургский государственный университет путей сообщения Императора Александра I,
г. Санкт-Петербург, 190031, Россия

В статье обсуждается способ использования подхода, основанного на применении известной формулы Байеса, для оценивания качества программных продуктов в рамках моделей качества и процесса оценивания, предусмотренных стандартом ISO/IEC 9126 (ГОСТ Р ИСО/МЭК 9126-93). Кратко описываются модели качества ПО и процесса оценивания, предлагаемые стандартом ISO/IEC 9126 и заменившим его стандартом ISO/IEC 25010:2011, указывается место применения подхода при реализации процесса оценивания.

Оценку качества ПО предлагается представлять в виде распределения вероятностей на множестве гипотез о том, что качество оцениваемого программного продукта достигло одного из predetermined уровней ранжирования, предусмотренных моделью. С использованием формулы Байеса формируется апостериорное распределение вероятностей, базирующееся на пересмотренном и уточненном в ходе оценивания качества априорном распределении вероятностей, сформированном перед началом процедуры оценивания. В качестве исходных данных для получения вероятностей используются результаты измерения разнородных метрик произвольного набора атрибутов качества, определяемых моделью качества, причем подход позволяет использовать как метрики, измеренные непосредственно, так и получившие свои значения в результате экспертного оценивания.

Предлагаемый подход позволяет получать осмысленные оценки качества даже в случае наличия неполных, неточных и противоречивых результатов измерения метрик качества.

Ключевые слова: качество ПО, ISO/IEC 9126, метрики качества, методы оценки ПО, экспертное оценивание, аттестация ПО, формула Байеса, байесовский подход.

В области разработки ПО уверенности потребителя в высоком качестве продукта можно добиваться двумя путями [1, 2]:

– совершенствовать процессы производства ПО в соответствии с современными подходами (модели CMM/CMMI, стандарты ISO/IEC 9003:2004 и ISO/IEC 15504) с целью достижения высокого уровня зрелости и/или возможностей предприятий-разработчиков;

– контролировать соответствие характеристик качества разработанной программной продукции требованиям (заказчика, ГОСТ) на всех этапах ее создания (проектирование, реализация, внедрение и эксплуатация).

Эти направления обеспечения качества ПО взаимно дополняют друг друга. Проблематике оценивания качества производственных процессов согласно модели CMMI® посвящен ряд работ авторов, в частности [3]. В настоящей статье авторов интересует второе направление обеспечения качества ПО – оценивание качества программной продукции.

На данный момент предложено большое количество моделей оценки качества ПО, краткий обзор которых дается, например, в [4]. Общим моментом во всех предлагаемых моделях является попытка построить интегральный критерий оценки качества программного продукта через оценку некоторого набора характеристик качества (таких, как надежность, эффективность, мобильность и т.п.), выполняемую посредством измерения значений выбранных количе-

ственных признаков, называемых метриками, и сопоставления их с требуемыми (ожидаемыми) значениями.

Основная проблема всех известных моделей – именно способ получения обоснованной интегральной оценки, поскольку измеряемые метрики являются разнородными. Для получения значений выбранных метрик используются различные способы: экспертное оценивание, непосредственное измерение, в том числе в процессе тестирования или аттестации программного изделия [5]. Для непосредственного измерения значений характеристик ПО в рамках тестирования, аттестации и даже последующей эксплуатации могут применяться различные специализированные программные инструменты, такие как профилировщики и вспомогательные утилиты, фиксирующие степень использования системных ресурсов [6].

Кроме того, модели оценки качества не предлагают обоснованного способа агрегации результатов частных измерений в итоговую интегральную оценку качества, отдавая это на откуп лицам, выполняющим оценивание. Как правило, рекомендуются процедуры, сводящиеся к получению взвешенных средних, что дополнительно требует определения относительной важности показателей и приведения значений разнородных метрик к некоторой единой шкале [7].

Ясно, что результаты единичных измерений метрик (и тем более экспертные оценки) не обладают достаточной точностью, которая могла бы гарантировать высокую точность обобщенного показателя качества ПО. Для повышения надежности измерений прихо-

дится прибегать к процедурам их многократного повторения с последующим усреднением полученных результатов [8].

В связи с возникающей неопределенностью в результатах оценивания имеет смысл перейти к получению вероятностной оценки качества в виде распределения вероятностей на множестве гипотез о достижении качеством оцениваемого ПО определенного уровня. При формировании такой оценки должны использоваться имеющиеся результаты измерений, обязательно являющиеся точными, и субъективные экспертные оценки.

Для получения подобных оценок авторы предлагают использовать известную формулу Байеса [9]. Способ ее применения базируется на ранее сформулированном подходе к оцениванию качества управленческих решений в железнодорожной отрасли. Авторы показали возможность его эффективного использования при оценивании выполнения практик модели СМММ® [3] и далее предприняли попытку развить его в работах [10, 11].

Модели качества ПО и процесса оценивания качества по ISO/IEC 9126

В настоящей статье авторы следуют определению качества ПО, данному в ГОСТе Р ИСО/МЭК 9126-93.

Этот стандарт подготовлен на основе аутентичного текста международного стандарта ISO/IEC 9126:1991 «Software engineering – Product quality» [12], который впоследствии был расширен в систему из четырех взаимосвязанных стандартов:

- ISO/IEC 9126-1:2001. Part 1: Quality model;
- ISO/IEC TR 9126-2:2003. Part 2: External metrics;
- ISO/IEC TR 9126-3:2003. Part 3: Internal metrics;
- ISO/IEC TR 9126-4:2004. Part 4: Quality in use metrics.

Далее в статье указанное семейство стандартов будет упоминаться как стандарт ISO/IEC 9126. Согласно ему, различают следующие понятия качества ПО:

– *внутреннее качество (internal quality)*, связанное с характеристиками ПО как такового, на каждой фазе его разработки, без учета поведения конечного продукта;

– *внешнее качество (external quality)*, характеризующее ПО с точки зрения его поведения;

– *качество ПО при использовании (quality in use)* в различных контекстах, то есть качество, воспринимаемое пользователями в конкретных сценариях работы ПО.

Для описания внутреннего и внешнего качества в ISO/IEC 9126 введена трехуровневая модель: на первом уровне располагаются характеристики качества ПО, каждая из которых уточняется набором комплексных показателей (атрибутов), а каждый атрибут оценивается по совокупности метрик. В таблице 1 приведены определения характеристик качества ПО и ис-

пользуемых для их оценивания атрибутов по стандарту ISO/IEC 9126-1:2001.

Таблица 1

Характеристики и атрибуты качества ПО

Table 1

Software quality characteristics and attributes

Характеристика качества ПО	Определение характеристики	Используемые атрибуты
Функциональность	Способность ПО в определенных условиях решать задачи, соответствующие установленным или предполагаемым потребностям пользователя. Определяет, что именно выполняет ПО, какие задачи решает	Функциональная пригодность, точность (вычислений), способность к взаимодействию (с нужным набором других систем), соответствие стандартам и нормативным актам, защищенность
Надежность	Способность ПО поддерживать определенный уровень качества функционирования (работоспособность) в заданных условиях в установленный период времени	Зрелость, устойчивость к отказам, способность к восстановлению, соответствие стандартам надежности
Удобство использования	Способность ПО быть удобным при обучении и в использовании, а также привлекательным для пользователей	Понятность, удобство обучения, простота использования, привлекательность, соответствие стандартам практичности
Эффективность/производительность	Соотношение между уровнем качества функционирования ПО и объемом используемых ресурсов при установленных условиях. Можно определить как отношение получаемых с помощью ПО результатов к затрачиваемым ресурсам	Временная эффективность, эффективность использования ресурсов, соответствие стандартам производительности
Сопровождаемость	Удобство осуществления всех видов деятельности, связанных с сопровождением программ	Анализируемость, удобство внесения изменений, устойчивость, тестируемость, соответствие стандартам удобства сопровождения
Переносимость	Способность ПО сохранять работоспособность при переносе из одного окружения в другое с учетом организационных, аппаратных и программных аспектов окружения	Адаптируемость, удобство установки, способность к сосуществованию, взаимозаменяемость, соответствие стандартам переносимости

Для описания качества ПО при использовании стандарт ISO/IEC TR 9126-4:2004 предлагает другой, более узкий набор характеристик.

- *Эффективность*. Способность ПО предоставлять пользователям возможность решать нужные задачи с необходимой точностью при использовании в заданном контексте.

- *Продуктивность*. Способность ПО предоставлять пользователям определенные результаты в рамках ожидаемых затрат ресурсов.

- *Безопасность*. Способность ПО обеспечивать минимальный уровень риска нанесения ущерба жизни и здоровью людей, бизнесу, собственности или окружающей среде.

- *Удовлетворение пользователей*. Способность ПО приносить удовлетворение пользователям при использовании в заданном контексте.

Для оценки каждого атрибута качества в стандарте определены наборы метрик:

- *полнота реализации функций*, измеренная как процент реализованных функций по отношению к перечисленным в требованиях; используется для измерения функциональной пригодности;

- *корректность реализации функций*, оцененная через правильность их реализации по отношению к требованиям; используется для измерения функциональной пригодности;

- *наглядность и полнота документации* оцениваются экспертами и используются для оценки понятности.

Отметим, что в новом стандарте ISO/IEC 25010:2011 [13], заменившем ISO/IEC 9126-1:2001, в систему характеристик качества внесены следующие изменения.

- 31 атрибут качества разнесен по восьми характеристикам (вместо шести, введенных в ISO/IEC 9126).

- Характеристика *Функциональность* переименована в *Функциональную пригодность*, добавлен атрибут *Функциональная полнота*, а атрибуты *Способность к взаимодействию* и *Защищенность* перемещены в другую характеристику. Атрибут *Точность* переименован в *Функциональную корректность*, а *Функциональная пригодность* переименована в *Функциональную уместность*.

- Характеристика *Эффективность/производительность* переименована в *Эффективность работы*, добавлен атрибут *Емкость*.

- Добавлена характеристика *Совместимость*, характеризуемая атрибутами *Способность к сосу-*

ществованию, перенесенным из *Переносимости*, и *Способность к взаимодействию*, перенесенным из *Функциональности*.

- Для характеристики *Удобство использования* добавлены атрибуты *Защищенность от ошибок пользователя* и *Доступность* (для пользователей с различными наборами возможностей), атрибут *Понятность* переименован в *Простоту распознавания*, а атрибут *Привлекательность* – в *Эстетичность пользовательского интерфейса*.

- Для характеристики *Надежность* добавлен атрибут *Доступность*.

- Добавлена характеристика *Безопасность*, характеризуемая атрибутами *Конфиденциальность*, оценивающим доступность данных только авторизованным пользователям, *Целостность*, оценивающим невозможность неавторизованной модификации данных, *Неотрицаемость*, оценивающим возможность доказуемости совершения действий, *Подотчетность*, оценивающим возможность установления автора совершенных действий, и *Подлинность*, оценивающим возможность установления подлинности идентификации.

- Для характеристики *Сопровождаемость* добавлены атрибуты *Модульность* и *Повторное использование*; атрибуты *Удобство внесения изменений* и *Устойчивость* сведены к атрибуту *Модифицируемость*.

Процесс оценивания качества ПО в ISO/IEC 9126 (и заменившем его стандарте ISO/IEC 25010:2011) состоит из трех последовательных стадий (рис. 1).

1. *Определение требований* к качеству в терминах характеристик и атрибутов качества.

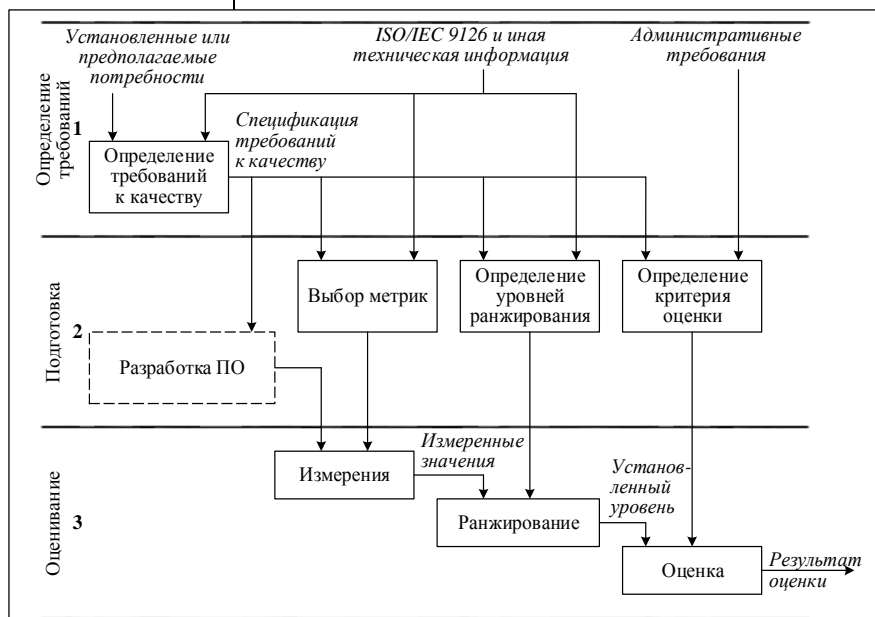


Рис. 1. Модель процесса оценивания качества ПО согласно ISO/IEC 9126

Fig. 1. A model of the software quality evaluation according to ISO/IEC 9126

2. Подготовка к оцениванию, включающая 3 этапа.

2.1. Выбор используемых метрик качества.

2.2. Определение уровней ранжирования для каждой метрики. Шкала измерений каждой метрики разделяется на диапазоны, соответствующие различным уровням удовлетворения поставленных требований. Уровни ранжирования определяются исходя из нанесения на шкалу значений метрики трех уровней, называемых *запланированным* уровнем, *установленным* уровнем и *худшего случая*. Установленный уровень определяется таким образом, чтобы оцениваемое ПО не было по данному атрибуту хуже уже существующего. Запланированный уровень устанавливает уровень качества, который считается достижимым при доступных разработчику ресурсах. Уровень худшего случая определяет минимально допустимый уровень качества, приемлемый для использования ПО. На базе отмеченных уровней стандарт определяет от двух до четырех уровней ранжирования качества, описанных в таблице 2, как показано на рисунке 2.

Таблица 2

Уровни ранжирования метрики

Table 2

Metric rating levels

Четыре уровня ранжирования		Два уровня ранжирования	
Уровень	Описание	Уровень	Описание
Низкий	Значение метрики не превышает уровень худшего случая	Неудовлетворительно	Значение метрики не превышает уровень худшего случая
Средний	Значение метрики достигает уровня худшего случая, но не превышает установленный уровень	Удовлетворительно	Значение метрики превышает уровень худшего случая
Хороший	Значение метрики достигает установленного уровня, но не превышает запланированный уровень		
Отличный	Значение метрики достигает или превышает запланированный уровень		

2.3. Определение критерия оценки. Результаты оценивания различных характеристик качества ПО должны быть подытожены. Для этого подготавливаются специальные процедуры (использование таблиц решений, определение среднего взвешенного и т.п.).

3. Оценивание, включающее в себя 3 этапа.

3.1. Измерение значений выбранных метрик для получения значения количественного признака.

3.2. Ранжирование измеренных значений.

3.3. Получение обобщенной (интегральной) оценки качества.

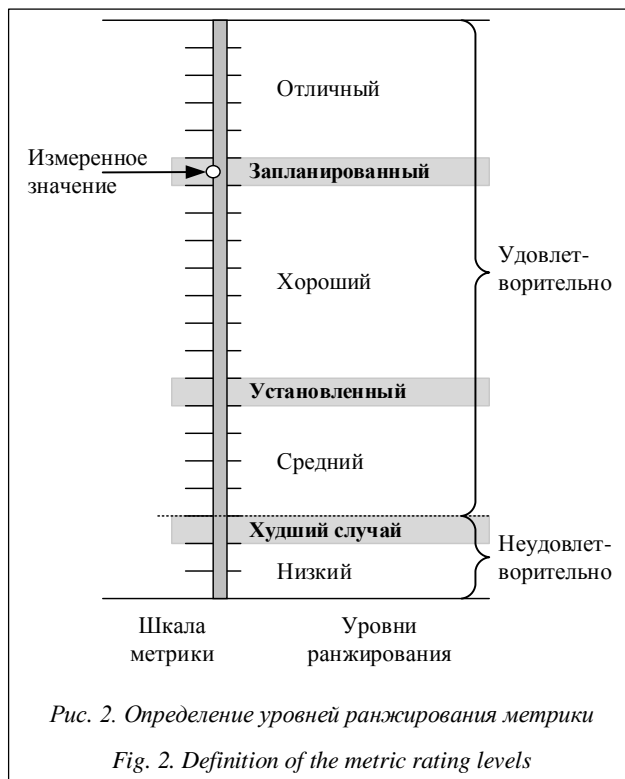


Рис. 2. Определение уровней ранжирования метрики

Fig. 2. Definition of the metric rating levels

Далее заключение о качестве ПО анализируется с учетом факторов времени и стоимости. Результатом является решение руководства о приемке/отбраковке или выпуске/отказе от выпуска программной продукции.

Байесовский подход к оцениванию качества ПО

Рекомендуемая в стандарте ISO/IEC 9126 модель трехэтапного процесса оценивания качества ПО может применяться в любой подходящей фазе жизненного цикла для каждого компонента программной продукции. Прямые рекомендации по выбору метода получения интегральной оценки качества, подытоживающей результаты оценивания различных характеристик, в стандарте отсутствуют. Имеется только замечание о необходимости подготовки специальных процедур.

С точки зрения авторов, в условиях отсутствия прямых рекомендаций при выборе метода получения интегральной оценки качества следует обратить особое внимание на тот факт, что высокая точность при оценивании в любом случае невозможна, да и не требуется в силу неопределенности исходных данных. Неопределенность обусловлена естественной неточностью измерений и субъективностью экспертных оценок, и в этом контексте привлекательным выглядит применение в процедуре оценивания качества ПО формулы Байеса. Как известно, байесовский подход давно и успешно используется в области принятия решений в условиях неопределенности [14–17].

В данной статье предлагается использовать формулу Байеса для формирования интегральной оценки качества ПО на основе неточных измерений разнородных метрик качества и субъективных экспертных оценок. Рассмотрим предлагаемый способ получения интегральной оценки качества программного продукта.

1. В модели процесса оценивания качества, рекомендуемой стандартом ISO/IEC 9126, каждое измеренное значение метрики соотносится с одним из k заданных уровней ранжирования L_1, \dots, L_k , определенных согласно таблице 2. Напомним, что стандарт определяет варианты с 2, 3 и 4 уровнями ранжирования. Будем использовать эти же уровни ранжирования применительно к итоговой интегральной оценке качества ПО. Обозначим через H_i гипотезу вида «Качество программного продукта в целом соответствует уровню L_i », где $i = 1, \dots, k$ – порядковый номер уровня ранжирования. Для оценивания качества необходимо выбрать набор характеристик и атрибутов качества (из перечня, предлагаемого стандартом), а также наборы метрик для оценки каждого атрибута. Обозначим отобранные для оценивания качества атрибуты через $A_j, j = 1, \dots, n$, где n – общее число отобранных атрибутов, а через M_{jl} – l -ю метрику, используемую для оценки атрибута $A_j, l = 1, \dots, s_j$, где s_j – число метрик, отобранных для оценки атрибута A_j так, что $s = s_1 + \dots + s_n$ – общее число метрик, привлеченное для оценивания качества ПО по всем n атрибутам.

2. Перед началом оценивания лицо, его выполняющее, формирует априорное распределение вероятностей $P(H_i)$ на множестве гипотез $H_i, i = 1, \dots, k$. Каждая вероятность $P(H_i)$ рассматривается как степень уверенности этого лица в справедливости i -й гипотезы об уровне качества оцениваемого программного продукта до начала оценивания, то есть до получения каких-либо результатов измерения выбранных метрик. Если априорная информация о качестве отсутствует, то все гипотезы можно считать равновероятными: $P(H_i) = 1/k$. При наличии достаточного основания допускается использование и неравномерного распределения априорных вероятностей на множестве гипотез. Например, крайние гипотезы H_1 и H_k в случае, если $k > 2$, представляются менее вероятными, чем все остальные, поэтому их априорные вероятности могут иметь более низкие значения. Кроме того, в качестве априорных вероятностей $P(H_i)$ могут быть использованы апостериорные байесовские вероятности $P(H_i | A_1, \dots, A_n)$, полученные в ходе предыдущей итерации оценивания качества ПО.

3. Условная вероятность $P(A_j | H_i)$ понимается как степень соответствия полученных результатов измерений множеству метрик, используемых для атрибута $A_j, j = 1, \dots, n$, предположению об истинности данной гипотезы $H_i, i = 1, \dots, k$. Такая вероятность в [15] называется правдоподобностью (*likelihood*). Значение этой условной вероятности определяется следующим обра-

$$\text{зом: } P(A_j | H_i) = \frac{\text{count}(M_{jl} \in L_i)}{r \cdot s_j}, \quad (1)$$

где M_{jl} – измеренные значения l -й метрики атрибута $A_j, l = 1, \dots, s_j; r \geq 1$ – число измерений значений метрик или число экспертов, привлеченных к оцениванию значения данной метрики; s_j – число метрик, используемых для оценивания атрибута качества $A_j, j = 1, \dots, n$. Согласно (1), каждая вероятность $P(A_j | H_i)$ определяется через отношение числа отметок о принадлежности результатов измерений метрик уровню ранжирования L_i к общему количеству произведенных измерений значений метрик данного атрибута. Назначенные уровни ранжирования показывают, какой (по результату измерения или субъективному мнению эксперта) уровень качества достигнут по выбранной метрике в каждом из случаев. При этом естественно возникающая в процессе измерения значений метрик неопределенность может быть отражена в назначении одному и тому же значению метрики M_{jl} двух смежных уровней качества L_i и $L_{i'}$, таких, что $|i - i'| = 1$.

4. Условная вероятность $P(H_i | A_1, \dots, A_n)$ понимается как степень уверенности лица, производящего оценивание, в справедливости i -й гипотезы об уровне интегрального качества ПО после получения оценок по всем атрибутам A_1, \dots, A_n . В соответствии с теоремой Байеса, при условии независимости всех измерений она вычисляется как апостериорная байесовская вероятность:

$$P(H_i | A_1, \dots, A_n) = \frac{P(H_i) \cdot P(A_1 | H_i) \cdot P(A_2 | H_i) \cdot \dots \cdot P(A_n | H_i)}{\sum_{h=1}^k P(H_h) \cdot P(A_1 | H_h) \cdot P(A_2 | H_h) \cdot \dots \cdot P(A_n | H_h)} \quad (2)$$

5. Полученное по формуле (2) апостериорное распределение вероятностей $P(H_i | A_1, \dots, A_n)$ на множестве гипотез $H_i, i = 1, \dots, k$, является итоговой интегральной оценкой качества ПО и показывает, насколько правдоподобными по завершении процедуры оценивания стали гипотезы о том, что интегральное качество программного продукта достигло каждого из уровней.

Пример байесовского оценивания качества

Пусть для оценки качества разрабатываемого ПО выбраны три атрибута – A_1, A_2 и A_3 ; в свою очередь, для оценки уровня качества ПО для атрибута A_1 используются три метрики – M_{11}, M_{12}, M_{13} ; аналогично для атрибута A_2 используются две метрики – M_{21} и M_{22} , а для атрибута A_3 четыре метрики – M_{31}, M_{32}, M_{33} и M_{34} соответственно. Также определен набор из четырех уровней ранжирования $L_i \in \{\text{Низкий, Средний, Хороший, Отличный}\}, i = 1, \dots, 4$.

В таблице 3 показаны результаты соотнесения измерений значений каждой из метрик с установленными уровнями ранжирования и вычисления на их основе условных вероятностей $P(A_j | H_i)$ по каждому из атрибутов. Выполнено по одному измерению значений каждой из метрик, при этом для моделирования

неопределенности в процессе измерения выполнено соотнесение значений некоторых из метрик сразу с двумя смежными уровнями качества.

Таблица 3
Оценка качества по выбранным атрибутам и метрикам

Table 3
Quality evaluation according to specified attributes and metrics

Ранжирование метрик атрибута	Уровень качества			
	Низкий, L_1	Средний, L_2	Хороший, L_3	Отличный, L_4
Атрибут A_1				
M_{11}	*	*		
M_{12}		*	*	
M_{13}				*
$P(A_1 H_i)$	1/3	2/3	1/3	1/3
Атрибут A_2				
M_{21}	*			
M_{22}	*	*		
$P(A_2 H_i)$	1	1/2	0	0
Атрибут A_3				
M_{31}		*	*	
M_{32}				*
M_{33}			*	*
M_{34}	*			
$P(A_3 H_i)$	1/4	1/4	1/2	1/2

Полученные вероятностные оценки используются для получения итогового распределения вероятностей на множестве гипотез об интегральном качестве ПО, как показано в таблице 4.

Таблица 4
Байесовская интегральная оценка качества программного продукта

Table 4
Integral Bayesian software quality evaluation

Вероятности		Гипотезы об уровне качества			
		Низкий, H_1	Средний, H_2	Хороший, H_3	Отличный, H_4
Априорные	$P(H_i)$	1/4	1/4	1/4	1/4
Оценки соответствия гипотезам	$P(A_1 H_i)$	1/3	2/3	1/3	1/3
	$P(A_2 H_i)$	1	1/2	0	0
	$P(A_3 H_i)$	1/4	1/4	1/2	1/2
Апостериорные	$P(H_i A_1, A_2, A_3)$	1/2	1/2	0	0

В данном примере по результатам измерений, выполненных для выбранных атрибутов и метрик, можно сделать вывод о том, что качество программного продукта не выше среднего уровня. При этом у лица, принимающего решение, нет оснований предпочесть гипотезу о среднем уровне качества гипотезе о том, что уровень качества низкий.

Особенности байесовского оценивания качества

Рассмотренный подход обладает рядом полезных особенностей, позволяющих, в частности, обнаруживать и учитывать несогласованность частных оценок по различным атрибутам [10, 11]. Остановимся на них подробнее.

1. Если в ходе оценивания всем метрикам M_{ji} по каждому атрибуту A_j присвоены отметки о принадлежности *только одному* уровню ранжирования L_i , то апостериорное распределение вероятностей на множестве гипотез о качестве становится *вырожденным*: $P(H_i | A_1, \dots, A_n) = 1$, $P(H_h | A_1, \dots, A_n) = 0$, $h = 1, \dots, k$, $h \neq i$. При этом единственную гипотезу H_i можно рассматривать как достоверную, а остальные гипотезы – как невозможные.

2. Если по *каждому* атрибуту A_j все используемые метрики M_{jl} , $l = 1, \dots, s_j$, в ходе оценивания получают *разные* уровни ранжирования, то распределения условных вероятностей $P(A_j | H_i)$, $i = 1, \dots, k$, становятся равномерными. В этом случае совпадение априорного распределения $P(H_i)$ и апостериорного распределения $P(H_i | A_1, \dots, A_n)$, $i = 1, \dots, k$, на множестве гипотез показывает, что измерения метрик *противоречивы* и не дают данных, влияющих на итоговое распределение.

3. Другим характерным случаем противоречий является ситуация, когда для каждого уровня ранжирования L_i найдется *хотя бы один* атрибут A_j , такой, что $P(A_j | H_i) = 0$. В итоге полная вероятность в знаменателе формулы (2) примет значение 0, что не позволяет вынести *обоснованное свидетельствами* суждение об интегральном качестве ПО.

Отметим, что в случае использования метода взвешенных средних при наличии противоречивых оценок качества в силу вступает *эффект компенсации* [18]: низкий уровень качества по одному атрибуту компенсируется (интенсивность такой компенсации и регулируется весами) высоким уровнем качества по другим. Данный эффект может в существенной мере исказить представление лица, производящего оценивание, об общем уровне качества ПО.

Заключение

Признанному авторитету в области программной инженерии Р. Глассу принадлежит ироничное высказывание о том, что нет ни приемлемого определения качества программного продукта, ни согласия по поводу того, на ком лежит ответственность за это качество. И даже если найдено определение, которое всех устроит, то придется придумать, как измерить достигнутый уровень качества для любого программного продукта [19]. Разделяя эту иронию и осознавая сложность задачи, авторы данной статьи, тем не менее, все же хотели бы внести свой вклад в дело ее решения и, будучи солидарными с мнением академика В.И. Ар-

нольда о том, что сложные модели редко бывают полезными [20], предлагают простой подход с использованием формулы Байеса для формирования интегральной оценки качества ПО на основе показателей (атрибутов) и метрик, представленных в ISO/IEC 9126 (ГОСТ Р ИСО/МЭК 9126-93) и в заменившем его стандарте ISO/IEC 25010:2011.

Под байесовской оценкой качества ПО понимается апостериорное распределение вероятностей на множестве гипотез о качестве, пересмотренное и уточненное в процессе оценивания по различным показателям (атрибутам) и позволяющее лицу, производящему оценивание, обоснованно отдать предпочтение той или иной гипотезе о качестве ПО.

Апостериорные вероятности рассматриваются как обобщенные оценки соответствия результатов неточных измерений выбранных метрик и/или субъективных экспертных оценок гипотезам о достижении качеством ПО одного из predetermined уровней ранжирования.

К преимуществам предложенного подхода можно отнести отсутствие ограничений на количество, конкретный перечень, содержание используемых атрибутов и точность измерения выбранных метрик; вероятностный характер интегральной оценки качества программного продукта, объективно снижающий требования к точности; а также упрощение процедуры итогового оценивания и естественное выявление случаев несогласованности, противоречивости в оценках качества продукта по отдельным показателям.

Литература

1. Кожомбердиева Г.И. Оценка качества программного обеспечения. СПб: Изд-во ПГУПС, 2010. 44 с.
2. Ахен Д.М., Клауз А., Тернер Р. СММИ®: Комплексный подход к совершенствованию процессов. Практическое введение в модель; [пер. с англ.]. М.: МФК, 2005. 330 с.
3. Кожомбердиева Г.И., Бураков Д.П., Гарина М.И. Использование формулы Байеса при оценивании выполнения практик модели СММИ® // Программные продукты и системы. 2017. № 1. С. 67–74.
4. Жарко Е.Ф. Сравнение моделей качества программного обеспечения: аналитический подход // XII Всерос. совещ. по пробле-

мам управления (ВСПУ-2014): сб. тр. 2014. С. 4585–4594. URL: <http://vspu2014.ipu.ru/proceedings/prcdngs/4585.pdf> (дата обращения: 02.08.2018).

5. Черников Б.В., Поклонов Б.Е. Оценка качества программного обеспечения: практикум. М.: ФОРУМ; ИНФРА-М, 2012. 400 с.
6. Intel® VTune™ Amplifier 2017. URL: <https://software.intel.com/sites/default/files/managed/d7/ba/intel-vtune-amplifier-2017-product-brief.pdf> (дата обращения: 02.08.2018).
7. Дроботун Е.Б., Козлов Д.В. Оценка степени влияния антивирусных программных средств на качество функционирования информационно-вычислительных систем // Программные продукты и системы. 2016. № 4. С. 129–134.
8. Картавенко М. Тест антивирусов на быстродействие. 2012. URL: http://www.anti-malware.ru/antivirus_test_performance_2012 (дата обращения: 02.08.2018).
9. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и ее инженерные приложения. М.: Наука, 1988. 480 с.
10. Кожомбердиева Г.И., Бураков Д.П. Получение интегральной оценки качества программного обеспечения на основе формулы Байеса // Транспортные интеллектуальные системы (TIS-2017): сб. матер. I Междунар. науч.-практич. конф. 2017. СПб: Изд-во ПГУПС, 2017. С. 209–220.
11. Бураков Д.П., Кожомбердиева Г.И. Особенности применения байесовского подхода при оценивании качества программных продуктов // XX Междунар. конф. по мягким вычислениям и измерениям (SCM'2017): сб. докл. 2017. СПб: Изд-во СПбГЭТУ «ЛЭТИ», 2017. Т. 1. С. 35–38.
12. ISO/IEC 9126:1991. Software engineering – Product quality. URL: <https://www.iso.org/standard/16722.html> (дата обращения: 02.08.2018).
13. ISO/IEC 25010:2011. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models. URL: <https://www.iso.org/standard/35733.html> (дата обращения: 02.08.2018).
14. Райфа Г. Анализ решений (введение в проблему выбора в условиях неопределенности); [пер. с англ.]. М.: Наука, 1977. 408 с.
15. Моррис У.Т. Наука об управлении: байесовский подход; [пер. с англ.]. М.: Мир, 1971. 304 с.
16. Уткин Л.В. Анализ риска и принятие решений при неполной информации. СПб: Наука, 2007. 404 с.
17. Ветров А.Н., Прокопчина С.В., Нестеров А.О. Управление инвестиционными рисками строительных организаций на основе байесовских информационных технологий // Программные продукты и системы. 2014. № 1. С. 212–216.
18. Микони С.В. Теория принятия управленческих решений. СПб: Лань, 2015. 448 с.
19. Гласс Р. Факты и заблуждения профессионального программирования; [пер. с англ.]. СПб: Символ-Плюс, 2007. 240 с.
20. Арнольд В.И. О преподавании математики // Успехи математических наук. 1998. Т. 53. Вып. 1. С. 229–234.

Software & Systems
DOI: 10.15827/0236-235X.125.034-041

Received 06.08.18
2019, vol. 32, no. 1, pp. 034–041

Using the Bayes' theorem within software quality evaluation according to ISO/IEC 9126 standard

D.P. Burakov¹, Ph.D. (Engineering), Associate Professor of Mathematics and Modeling Department, burakovdmitry8@gmail.com

G.I. Kozhombardieva¹, Ph.D. (Engineering), Associate Professor of Information and Computing Systems Department, kgi-liizht@yandex.ru

¹ Emperor Alexander I St. Petersburg State Transport University, St. Petersburg, 190031, Russian Federation

Abstract. The paper discusses a way to use the approach based on well-known Bayes' rule to evaluate software quality according to quality model and evaluation process described in the ISO/IEC 9126 standard. In addition, it briefly describes software quality

models and evaluation process that are proposed by the abovementioned standard, as well as by the improved ISO/IEC 25010:2011 standard. The authors define the field of using the proposed approach during the evaluation process.

The software quality evaluation is presented as a probability distribution on a set of hypotheses that software quality has reached one of the predefined quality levels proposed by the model. The Bayes' formula is used to build a posteriori probability distribution based on revised and refined during quality evaluation a priori probability distribution that is defined before evaluation. The source data for calculating probabilities is the results of measurement of heterogeneous quality metrics for arbitrary set of quality attributes that are specified in the software quality model.

The proposed approach allows using both directly measured metrics and the metrics estimated by experts. In fact, the approach gives reasonable software quality evaluation even if there are incomplete, inaccurate and inconsistent quality metrics.

Keywords: ISO/IEC 9126; software quality; quality metrics; software evaluation methods; expert estimation; software validation; Bayes' rule; Bayesian approach.

References

1. Kozhombierdieva G.I. *Software Quality Evaluation*. St. Petersburg State Transport Univ. Publ., 2010, 44 p.
2. Ahern D.M., Clouse A., Turner R. *CMMI Distilled: A Practical Introduction to Integrated Process Improvement*. Addison-Wesley Prof. Publ., 2004, 305 p. (Russ. ed.: Moscow, MFK Publ., 2005, 330 p.).
3. Kozhombierdieva G.I., Burakov D.P., Garina M.I. Using Bayes' theorem to estimate CMMI® practices implementation. *Software & Systems*. 2017, no. 1, pp. 67–74 (in Russ.).
4. Zharko E.F. Comparison of software quality models: analytical approach. *Proc. 12th All-Russian Meeting on Control Problems VSPU-2014*. 2014, pp. 4585–4594. Available at: <http://vspu2014.ipu.ru/proceedings/prcdngs/4585.pdf> (accessed August 2, 2018).
5. Chernikov B.V., Poklonov B.E. *Software Quality Evaluation. Practice*. Moscow, Forum Publ., 2012, 400 p.
6. *Intel®VTune™ Amplifier 2017*. Available at: <https://software.intel.com/sites/default/files/managed/d7/ba/intel-vtune-amplifier-2017-product-brief.pdf> (accessed August 2, 2018).
7. Drobotun E.B., Kozlov D.V. Assessment of Influence of Anti-Virus Software on Quality of Information-Computing System Functioning. *Software & Systems*. 2016, no. 4, pp. 129–134.
8. Kartavenko M. *Anti-Virus Performance Testing (March, 2002)*. Available at: http://www.anti-malware.ru/antivirus_test_performance_2012 (accessed August 2, 2018).
9. Venttsel E.S., Ovcharov L.A. *The Theory of Probability and its Engineering Applications*. Moscow, Nauka Publ., 1988, 480 p.
10. Kozhombierdieva G.I., Burakov D.P. Integrated evaluation of software quality based on Bayes' theorem. *Proc. 1st Int. Sci. and Pract. Conf. "TIS-2017"*. St. Petersburg State Transport Univ. Publ., 2017, pp. 209–220 (in Russ.).
11. Burakov D.P., Kozhombierdieva G.I. Features of using Bayesian approach for software products quality evaluation. *Proc. 20th Int. Conf. on Soft Computing and Measurements ("SCM'2017")*. St. Petersburg Electrotechnical Univ. "LETU" Publ., 2017, vol. 1, pp. 35–38 (in Russ.).
12. *ISO/IEC 9126:1991. Software engineering – Product quality*. Available at: <https://www.iso.org/standard/16722.html> (accessed August 2, 2018).
13. *ISO/IEC 25010:2011. Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality Models*. Available at: <https://www.iso.org/standard/35733.html> (accessed August 2, 2018).
14. Raiffa H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Addison-Wesley Publ., Reading, MA, 1968 (Russ. ed.: Moscow, Nauka Publ., 1977, 408 p.).
15. Morris W. *Management Science: A Bayesian Introduction*. Prentice-Hall Publ., Englewood Cliffs, NY, 1968 (Russ. ed.: Moscow, Mir Publ., 1971, 304 p.).
16. Utkin L.V. *Risk Analysis and Decision Making with Incomplete Information*. St. Petersburg, Nauka Publ., 2007, 404 p.
17. Vetrov A.N., Prokopchina S.V., Nesterov A.O. Investment risks management for construction organizations based on Bayesian information technologies. *Software & Systems*. 2014, no. 1, pp. 212–216 (in Russ.).
18. Mikoni S.V. *Theory of Management Decision Making*. St. Petersburg, Lan Publ., 2015, 448 p.
19. Glass R. *Facts and Fallacies of Software Engineering*. Addison-Wesley Publ., 2002 (Russ. ed.: St. Petersburg, Simvol-Plus Publ., 2007, 240 p.).
20. Arnold V.I. On teaching mathematics. *Russian Mathematical Surveys*. 1998, vol. 53, iss. 1, pp. 229–236 (in Russ.).