

УДК 519.873

Ю.В. СОСНОВСКИЙ

Таврический национальный университет им. В.И. Вернадского, Украина

МУЛЬТИСИТУАТИВНАЯ ОЦЕНКА ПОТРЕБИТЕЛЬСКИХ КАЧЕСТВ ПРОГРАММНЫХ СРЕДСТВ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ В СПЕЦИАЛЬНОЙ ЗАДАЧЕ ОЦЕНИВАНИЯ

Выполнен квалиметрический анализ программных комплексов статистической обработки данных с привлечением специалистов гуманитарного профиля, что дало возможность учесть мнения пользователей конкретной специфической области науки и выделить наиболее приемлемое программное обеспечение. Рассматривается задача мультиситуативной оценки потребительских качеств программных пакетов для статистического анализа данных. Представлено дерево свойств, по которым выполняется квалиметрия. Оценка выполняется группой экспертов определенной предметной области.

Ключевые слова: статистический анализ данных, квалиметрия, программные средства.

Введение

Проблема эффективной статистической обработки результатов экспериментов не нова. С учетом бурного развития информационных технологий, создан ряд программных систем для обработки и анализа статистических данных, некоторые системы насчитывают более десяти выпусков. Большинство подобных программных систем позиционируется создателями как универсальный инструмент для решения практически любой задачи анализа данных.

Как свидетельствуют публикации в периодической печати, во многих случаях пользователю предлагаются десятки, сотни программных продуктов одного назначения, существенно отличающихся по качеству и цене. И цена, и число продаж, зачастую, мало зависят от реального качества представленного на рынок программного продукта. Существующие методы оценки программных систем для статистического анализа данных разработаны и ориентированы на специалистов в области математики и статистики.

Целью работы является проведение квалиметрического анализа потребительских качеств систем статистического анализа данных применительно к обработке результатов медико-биологических исследований, а также поиск технических предпосылок низкого уровня выполнения статистического анализа исследователями гуманитарной направленности, математико-статистическая подготовка которых значительно отличается от соответствующего специалиста-математика.

Принимая во внимание значительное число критических публикаций об уровне качества статистических расчетов в рассматриваемой области, вы-

брана нестандартная методика оценки качества рассматриваемых систем.

Особенностью методики является привлечение медицинских специалистов к исследованию и их использование в качестве экспертов для проведения квалиметрического анализа.

Специализированные программные комплексы не могут быть измерены количественно сами по себе, сравнить их возможно с привлечением аппарата квалиметрического анализа [1]. Для сравнения привлекается пользователь, который может оценить различные характеристики пакета с точки зрения опыта использования программы. Таким образом, получаемая оценка является субъективной и относительной, ввиду подсознательного сравнения с другим статистическим пакетом или, может быть, гипотетическим эталоном.

С целью сокращения записи программные продукты, которые могут быть использованы для статистической обработки результатов медико-биологических исследований, будем относить к классу «программный объект» или, сокращенно ПО.

Все рассматриваемые программные продукты являются широко известными, использовались в виде демо-версий и в рамках оценочной задачи обладали сопоставимой функциональностью. Ввиду особенности рассматриваемой ситуации квалиметрии, а также некоторой доли субъективности в исследовании, зависящей от мнения экспертов, конкретные названия и торговые марки программных продуктов автором не называются и заменены идентификаторами ПО1-ПО6.

Задачи квалиметрической оценки могут быть сформулированы на основе [2], применительно к рассматриваемой ситуации, следующим образом:

1. Выбор лучшего по критерию качества варианта программного объекта;

2. Сравнение качества программного объекта с уровнем качества другого, имеющего близкую функциональность;

3. Поиск количественного выражения наиболее общей характеристики программного объекта – его интегрального качества в текущей ситуации применения;

4. Выявление свойств программного объекта, по которым он доминирует над аналогичными продуктами.

С целью разработки методики оценки качества используется алгоритм [1], который модифицируется для квалиметрического анализа программных объектов. Основными его узлами, согласно [2], являются:

1. Идентификация ситуации, в которой производится оценка;

2. Построение «дерева свойств» ПО;

3. Выделение показателей свойств ПО;

4. Выбор значений коэффициентов весомости свойств ПО;

5. Нахождение значений показателя качества (или интегрального качества Q) ПО.

В данной работе специальная ситуация оценки заключается в рассмотрении программных интегрированных комплексов статистического анализа, которые используются специалистами гуманитарного профиля для решения задач своей прикладной области.

Результаты квалиметрического анализа могут быть использованы для принятия решений о применении соответствующих программных продуктов, а также с целью формулирования требований к созданию экспертной системы по обработке медико-биологических данных.

Дерево свойств программного объекта

Учитывая, что первичные оценки программному объекту в текущей ситуации оценивания выставляются экспертами (в роли которых выступают участники эксперимента), такие оценки относительно и могут выполняться в шкале рангов. Подобный подход дает право представить свойства объекта, в отношении которого проводится квалиметрический анализ, в виде открытого иерархического дерева свойств, изображенного на рисунке 1. Ряд свойств в рамках проводимого исследования не может быть оценен, и при этом не является значимо важным для текущей ситуации оценки. В тоже время, при необходимости, дерево свойств может быть дополнено.

Таким образом, сформировалась объективная проблема разрыва между специфическими возмож-

ностями программного обеспечения и уровнем квалификации реальных пользователей медико-биологической предметной области. И с выходом новых версий программных продуктов данная проблема продолжает усугубляться ввиду наращивания возможностей программных средств с одной стороны, и недостаточным объемом специальных знаний у пользователей – с другой.

Первый иерархический уровень дерева свойств состоит из трех компонент: функциональности (1), требований к квалификации пользователя (2), и интегральной группы второстепенных свойств (3). Под функциональностью подразумевается комплексная потенциальная способность рассматриваемого ПО выполнять набор функций, определенных в его внешнем описании и удовлетворяющим заданным или подразумеваемым потребностям пользователей.

Данное свойство I-го уровня иерархии объединяет два свойства II-го уровня – функциональную полноту (1.1) и комплексность решения задачи (1.2). Свойство «функциональная полнота» определяет возможность выполнения необходимых пользователю функций, определенных во внешнем описании оцениваемого ПО.

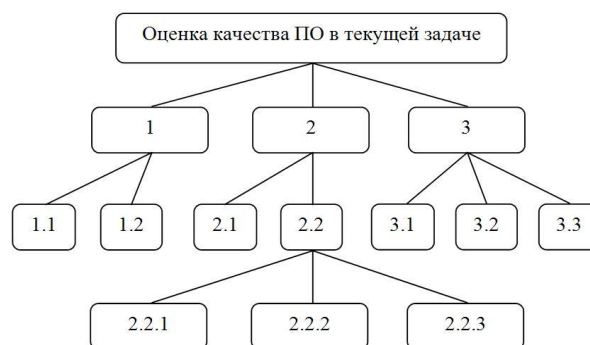


Рис. 1. Дерево свойств программного объекта

Исходя из необходимости комбинировать различные статистические процедуры, рассматриваемая целевая задача представляется как совместное решение задач обработки данных различными требуемыми методами, их выбор, применение и представление данных, что отражено в свойстве «комплексность решения задачи»

В текущей ситуации оценивания значительную роль играет свойство «требования к квалификации пользователя» I-го уровня, которое комплексует целый набор свойств: «трудоемкость освоения ПО» (2.1) и «качество помощи» (2.2) II-го уровня. Свойство «трудоемкость освоения ПО» важно ввиду того, что исследуется процесс взаимодействия специалиста гуманитарного профиля и программного продукта технической направленности. Данное свойство тесно связано с «качеством помощи», которое

комплексирует набор свойств III-го уровня: «качество и дружелюбность интерфейса» (2.2.1), «качество и интеллектуализация подсказок в области рассматриваемых задач» (2.2.2) и «качество встроенной помощи» (2.2.3).

Важной может являться такая характеристика, как стоимость. Интуитивно понятно, что она должна положительно коррелировать с качественными показателями продукта. Однако в ряде квалиметрических исследований стоимость не рассматривается в виде меры качества продукта, поэтому стоимость не включается в дерево свойств [3].

Интегрированная группа свойств (3) приводится для обеспечения целостности рассмотрения качества ПО и комплексирует ряд свойств II-го уровня, рассмотрение которых целесообразно в рамках исследования: «возможность перенастройки на другой класс задач» (3.1), «возможность работы в сети» (3.2), «операционное быстроедействие» (3.3).

В ходе оценки единичных показателей качества используется трехуровневая шкала. Экспертным методом определяется показатели качества из трех вариантов: Н – низкий показатель, С – средний показатель, В – высокий показатель качества рассматриваемого свойства. Исходная предпосылка в определении комплексного показателя качества (Q) состоит в том, что при высоком уровне всех единичных показателей значение комплексного показателя качества должно равняться 1. Аналогично, при всех средних единичных показателях $Q=0,5$ и при всех низких единичных показателях $Q=0$.

Распределение коэффициентов единичных показателей качества выбирается, в известном смысле, произвольно с учетом ситуации оценивания, описанной выше. Основываясь на предварительном опросе экспертов, составлена табл. 1, в которой приведены описания и коэффициенты единичных показателей, иерархия которых показана на рис. 1.

Таблица 1

Коэффициенты единичных показателей качества ПО

Описание показателя	Коэффициент показателя		
	r_i/S_1	r_i/S_2	r_i/S_3
Функциональная полнота	0,1	0,01	0,01
Комплексность решения задачи	0,1	0,75	0,2
Трудоемкость освоения ПО	0,2	0,2	0,75
Дружелюбность интерфейса	0,15	0,01	0,01
Качество подсказок в области рассматриваемых задач	0,2	0,01	0,01
Качество встроенной помощи	0,15	0,01	0,01
Интегрированная группа свойств	0,1	0,01	0,01

Значение комплексного показателя качества определяется по известной формуле:

$$Q = 1 - 0,5 \sum_{j=1}^{m_c} p_{c_j} - \sum_{j=1}^{m_n} p_{n_j}, \quad (1)$$

где p_{c_j} , p_{n_j} – нормированные коэффициенты показателей качества среднего и низшего уровней; m_c , m_n – количество показателей качества среднего и низшего уровней.

Таким образом, комплексный показатель качества (Q), рассчитываемый по формуле (1), является нормированным, принимает значения $Q \in (0,1)$. Значение $Q = 0$ в пределе свидетельствует о низшем уровне качества, $Q = 1$ о высшем уровне качества исследуемого объекта в текущей ситуации оценивания.

Число экспертов определено исходя из предположек, описанных в [3], где указывается минимальное значение – 7. В текущей ситуации оценивания это число увеличено, в исследовании принимали участие 15 экспертов. Их работа разбита на два этапа, различающиеся также уровнем сложности. Пер-

вый этап состоит в использовании описательной статистики для получения статистических характеристик вариационных рядов, построении графических зависимостей, расчете значения коэффициента корреляции между значениями двух вариационных рядов и построении линии регрессии.

Второй этап более сложен и включал в себя оценку типа статистического распределения данных с использованием непараметрических критериев (Колмогорова-Смирнова, к примеру), построение гистограммы и оценку достоверности различия медиан двух вариационных рядов.

Кроме этого, экспертам предоставлялось время на освоение статистических пакетов в пределах 3-х часов. По завершению отведенного времени контролировалось выполнение этапов анализа и представления данных, а также эксперты выставляли оценку единичным свойствам каждого программного объекта по трехуровневой шкале оценки. Полученные результаты заносились в соответствующие таблицы и по формуле (1) проводился расчет комплексного показателя качества (Q) оцениваемых программных объектов.

Результаты мультиситуативной квалиметрии ПО

Как показал пробный эксперимент, результаты сильно подвержены влиянию распределения коэффициентов единичных показателей качества ПО. Основываясь на конечной задаче исследования, в рассмотрение вводятся три управляемых ситуации $S_1 - S_3$, различающиеся векторами коэффициентов единичных показателей качества.

Результаты обработки мнения экспертов – усредненные по ситуациям значения интегрального показателя для рассматриваемых программных объектов, коэффициенты конкордации (W) и их уровни

значимости (p), относительное число экспертов, полностью выполнивших первый (N_1) и второй (N_2) этапы задания представлены в таблице 2.

В ситуации оценивания S_1 среднее значение интегрального показателя качества $Q_{cp} = 0,69$ при среднем отклонении $Q_{cp,откл.} = 0,07$, что свидетельствует об усреднении значения Q при равномерном распределении коэффициентов единичных показателей качества. Разница между максимальным и минимальным значением Q : $\Delta Q = 0,25$. Также, учитывая высокий класс рассматриваемых программных объектов, фактически отсутствуют значения среднего показателя качества менее 0,5.

Таблица 2

Значения \bar{Q}

	Интегральный показатель качества ПО Q					
	ПО1	ПО2	ПО3	ПО4	ПО5	ПО6
S_1	0,57	0,67	0,74	0,77	0,8	0,74
S_2	0,46	0,63	0,81	0,73	0,84	0,78
S_3	0,79	0,77	0,69	0,64	0,67	0,59
W	0,59	0,36	0,34	0,23	0,34	0,36
p	<0,01	<0,01	<0,01	0,08	<0,01	<0,01
N_1	86,7	93,3	80,0	93,3	93,3	80,0
N_2	53,3	53,3	40,0	46,7	60,0	26,7
\bar{Q}	0,61	0,69	0,75	0,71	0,77	0,70

С некоторой долей субъективности, а также учитывая последующие результаты, порог оценки принимается равным 0,75. В этом случае порог по критерию интегрального показателя качества преодолели ПО4 и ПО5.

Ситуация S_2 характеризуется смещением акцентов оценки в сторону комплексности решения задачи посредством рассматриваемого ПО. $Q_{cp} = 0,56$ и $Q_{cp,откл.} = 0,25$. Значение $\Delta Q = 0,75$ максимально в ходе анализа для всех рассматриваемых ситуаций. Учитывая выбранное значение порога 0,75, его преодолели ПО3, ПО5 и ПО6.

Ситуация S_3 характеризуется смещением акцентов оценки в сторону оценки трудоемкости освоения ПО. $Q_{cp} = 0,67$ и $Q_{cp,откл.} = 0,07$. Заслуживает рассмотрения тот факт, что заявленный порог преодолели ПО1 и ПО2, и это иные объекты в отличие от S_1 и S_2 .

Значения коэффициента конкордации (W) для каждого оцениваемого ПО существенно превосходят достигаемый уровень значимости и находятся в пределах от 0,23 до 0,59, таким образом мнения экспертов можно считать согласованными.

Анализ результатов оценки был бы неполным без учета относительного числа экспертов, успешно завершивших выполнение задания, особенно II эта-

па. По критерию максимального числа экспертов, успешно выполнивших задание II уровня и преодоления порогового $Q' = 0,75$ выделен доминирующий над остальными программный комплекс (ПО5). Он характеризуется максимальным относительным числом экспертов, успешно выполнивших задание II этапа (60,0%) и максимальным значением усредненного интегрального коэффициента качества $Q_{cp} = 0,77$, в равной степени учитывающего результаты трех ситуаций оценки $S_1 - S_3$.

Выводы

Выполнен квалиметрический анализ программных комплексов статистической обработки данных с привлечением специалистов гуманитарного профиля, что дало возможность учесть мнения пользователей конкретной специфической области науки и выделить наиболее приемлемое ПО.

Программные комплексы с относительно простым интерфейсом (ПО1, ПО2) получили высшие баллы в ситуации оценки простоты освоения программного объекта ($Q > 0,75$), а более мощные и универсальные пакеты (например, ПО4 и ПО5) получили низший балл ($Q < 0,7$). В тоже время мнение экспертов достаточно точно отражает функциональ-

ность пакета, в ситуации оценки комплексности решения задачи три наиболее мощных пакета преодолели порог $Q=0,75$. Анализ интегрального показателя качества, усредненного по ситуации оценки, выполнен в совокупности относительным числом экспертов, успешно выполнивших задания второго, более сложного этапа.

Выделен доминирующий по рассматриваемым показателям качества программный продукт (ПО5), получивший максимальный интегральный показатель качества $Q=0,77$.

Учитывая сложности на этапе формализации мнения экспертов и перевода его в шкалу рангов, что не в полной мере отражает особенности человеческого мышления, целесообразным может быть применение аппарата нечеткой логики для более

точного перевода мнения эксперта в формат оценки программного продукта.

Литература

1. Квалиметрическая экспертиза. Руководство по организации экспертизы и выполнению квалиметрических расчетов. Книга первая. Организация экспертизы [Текст] / под ред. В.М. Маругина, Г.Г. Азгальдова. – СПб., М.: Русский Регистр, 2002.

2. Техническое творчество: теория, методология, практика. Энциклопедический словарь-справочник [Текст]/ под ред. А. И. Половинкина, В. В. Попова. – М.: НПО «Информ-система», 1995. – 408 с.

3. Азгальдов Г.Г. Теория и практика оценки качества товаров. Основы квалиметрии [Текст]/ Г.Г. Азгальдов. – М.: Экономика, 1982. – 256 с.

Поступила в редакцию 14.03.2012

Рецензент: д-р техн. наук, проф. В.И. Хаханов, Харьковский национальный университет радиоэлектроники, Харьков, Украина

МУЛЬТСИТУАТИВНА ОЦІНКА СПОЖИВЧИХ ЯКОСТЕЙ ПРОГРАМНИХ ЗАСОБІВ СТАТИСТИЧНОГО АНАЛІЗУ ДАНИХ У СПЕЦІАЛЬНІЙ ЗАДАЧІ ОЦІНЮВАННЯ

Ю.В. Сосновський

Виконаний кваліметричний аналіз програмних комплексів статистичної обробки даних із залученням фахівців гуманітарного профілю, що дало можливість врахувати думки користувачів конкретної специфічної галузі науки і виділити найбільш прийнятне програмне забезпечення. Розглядається задача мультиситуативної оцінки споживчих якостей програмних пакетів для статистичного аналізу даних. Представлено дерево властивостей, за якими виконується кваліметрія. Оцінка виконується групою експертів певної предметної галузі.

Ключові слова: статистичний аналіз даних, кваліметрія, програмні засоби.

MULTI-SITUATIONAL ASSESSMENT OF STATISTICAL SOFTWARE IN THE SPECIFIC PROBLEM OF ESTIMATION

Y.V. Sosnovskij

The qualities analysis of programmatic complexes of statistical treatment of the specialists of humanitarian type given with bringing in is executed, that enabled to take into account opinions of users of concrete specific area of science and select the most acceptable software. The problem of multi-situational assessment of statistical software is considered. The properties of qualities tree is described. The assessment of quality is performed by the group of experts.

Key words: statistical analysis of data, assessment of quality, software.

Сосновський Юрій Вячеславович – канд. техн. наук, доцент, доцент кафедри комп'ютерної інженерії і моделювання Таврицького національного університету ім. В.И. Вернадського, Сімферополь, Україна.