

ПРИМЕНЕНИЕ ЭКСПЕРТНЫХ МЕТОДОВ ПРИ ОЦЕНИВАНИИ КАЧЕСТВА ПРИКЛАДНЫХ ПРОГРАММНЫХ СРЕДСТВ И СИСТЕМ АВТОМАТИЗАЦИИ ИНФОРМАЦИОННО- МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА И УПРАВЛЕНИЯ ОБРАЗОВАТЕЛЬНЫМ УЧРЕЖДЕНИЕМ

В.П. Граб

Применение математико-статистических методов значительно расширяет возможности использования информации, полученной от специалистов. Практика последних лет показала, что даже простые статистические методы в сочетании с этой информацией при выборе перспективных решений часто приводят к более успешным результатам, чем «точные» расчеты с ориентацией на средние показатели и экстраполяцию существующих тенденций.

Использование информации, полученной от специалистов, особенно плодотворно, если для ее сбора, обобщения и анализа применяются специальные процедуры, логические приемы и математические методы, получившие название «*Методы экспертных оценок*».

Для того чтобы получаемая экспертная информация была качественной, необходимо выполнение следующих условий:

- наличие экспертной комиссии, состоящей из специалистов, профессионально знакомых с объектом экспертизы, и имеющих опыт работы эксперта;
- наличие аналитической группы, профессионально владеющей технологией организации и проведения экспертиз, методами получения и анализа экспертной информации:
 - получение достоверной экспертной информации;
 - корректная обработка и анализ экспертной информации.

Отсутствие любого из перечисленных условий ставит под сомнение эффективность и корректность проводимой экспертизы. Экспертное оценивание представляет собой процесс измерения, который можно определить

как процедуру сравнения объектов по выбранным показателям (признакам). В качестве показателей сравнения могут использоваться пространственно-временные, физические, психические и другие. Для целей экспертного оценивания возникают следующие базисные задачи:

- построение обобщенной оценки понятий и объектов на основе индивидуальных оценок экспертов;
- построение обобщенной оценки на основе парного сравнения объектов каждым из экспертов;
- оценка компетентности экспертов;
- оценка достоверности результатов.

Применение экспертных методов эффективно при оценке актуальности и легитимности, при оценке значений содержательно-педагогических характеристик прикладных программных средств и систем автоматизации информационно-методического обеспечения образовательного процесса и управления образовательным учреждением (далее продукция образовательного назначения (ПОН)). Это обусловлено тем, что при квалиметрическом подходе к интегрированной оценке качества инновационной педагогической продукции, наиболее достоверные результаты получаются при применении совокупности комплексных, статистических, экспертных методов наряду с мониторингом и измерениями продукции.

Экспериментальная оценка, безусловно, в связи новизной проблемы и в силу отсутствия накопленного опыта проведения оценки качества продукции образовательного назначения приобретает первостепенное значение.

Нормативные документы, используемые при экспертизе продукции, определяют основные требования к проведению экспертного оценивания ПОН, методы и способы проверки их соответствия требованиям.

Экспертным методом получают оценку исследуемых характеристик путем фиксации факта их наличия в заданной форме (в эксплуатационных документах, на экране монитора, на бумажных носителях), оценивают реализацию функций ПОН при их выполнении.

При особо ответственных измерениях экспертным методом могут учитываться весовые коэффициенты квалификации экспертов.

Традиционные способы получения групповой оценки с помощью средних величин оказываются применимы только тогда, когда коллектив экспертов однороден в смысле характера ответов.

И все же практика показывает, что экспертные методы дают более надежные результаты, чем традиционные методы групповых решений (метод комиссий).

Одним из наиболее перспективных методов формирования групповой оценки экспертов является метод Дельфы. В основу метода Дельфы положены следующие предпосылки:

1) поставленные вопросы должны допускать возможность выражения ответа в виде числа;

2) эксперты должны располагать достаточной информацией для того, чтобы дать оценку;

3) ответ на каждый из вопросов (оценка) должен быть обоснован экспертом.

Метод Дельфы чаще всего используется для определения показателей групповой самооценки и численности группы [1].

Количество экспертов тоже играет важную роль [1]. С ростом числа экспертов в группе точность измерения повышается.

Исходная численность экспертной группы составляет обычно не менее 5 человек. В отдельных случаях она достигает 15–20 экспертов (массовый опрос проводится, как правило, только при социологических исследованиях). Если в подготовительный период число экспертов не определено, то достижение требуемой точности за счет расширения экспертной группы достигается уже в процессе измерения экспертным методом.

По тому, в какой форме эксперты выражают свое мнение, т.е. по способу проведения экспертизы, различают [2, 3, 4]:

- непосредственное измерение;
- ранжирование;

- сопоставление.

При непосредственных измерениях экспертным методом значения физических величин или показателей качества определяются сразу в установленных единицах (в единицах СИ, в баллах, нормо-часах, рублях, единицах условного топлива и т.д.). Непосредственное измерение весовых коэффициентов, сумма которых должна равняться единице, производится по шкале интервалов. Значения этих коэффициентов рассчитываются по формуле:

$$g_j = \frac{\sum_{i=1}^n G_{i,j}}{\sum_{i=1}^n \sum_{j=1}^m G_{i,j}}$$

где n – количество экспертов, m – число «взвешиваемых» показателей, G_{ij} – коэффициент весомости j -го показателя в баллах, данный i -м экспертом.

Метод непосредственной оценки состоит в том, что диапазон изменения какой-либо качественной переменной разбивается на несколько интервалов, каждому из которых присваивается определенная оценка (балл).

Непосредственное измерение экспертным методом является наиболее сложным и предъявляет к экспертам наиболее высокие требования. При решении многих практических задач часто оказывается, что факторы, определяющие конечные результаты, не поддаются непосредственному измерению. Расположение этих факторов в порядке возрастания (или убывания) какого-либо присущего им свойства называется ранжированием. Ранжирование позволяет выбрать из исследуемой совокупности факторов наиболее существенный.

В последнее время метод ранжирования широко применяется при измерении уровня качества продукции, оценке деятельности научных подразделений и организаций и т.п.

Процедура проведения экспертных оценок требует, чтобы существовала какая-либо мера, позволяющая хотя бы субъективно сравнивать эти оценки.

Трудности использования ранжирования, непосредственной оценки и метода последовательных сравнений при выявлении предпочтений для

большого числа факторов можно в определенной степени уменьшить, если предложить экспертам произвести сравнение этих факторов попарно, с тем чтобы установить наиболее важный в каждой паре.

Анализ прогнозов, выполненных отдельными специалистами и оказавшихся неверными, показал, что одна из наиболее распространенных ошибок таких прогнозов заключалась в том, что принимались во внимание факторы, которые впоследствии оказались малозначащими, и, наоборот, упускались наиболее существенные факторы.

При создании экспертной комиссии и подборе экспертов большое внимание уделяется согласованности их мнений [1, 2, 3, 4], которая характеризуется смещенной или несмещенной оценкой дисперсии отсчета. С этой целью на этапе формирования экспертной группы проводятся контрольные измерения с математической обработкой их результатов. Нередко при этом используется не один, а сразу несколько объектов измерений, которые в зависимости от их ценности или качества нужно расставить по шкале порядка, т.е. определить их ранг, ибо измерение по шкале порядка называется ранжированием. За меру согласованности мнений экспертов в этом случае принимается так называемый коэффициент конкордации [1].

$$W = 12 S / (nI * (mi - m)),$$

где S – сумма квадратов отклонений суммы рангов каждого объекта экспертизы от среднего арифметического рангов, n – число экспертов, m – число объектов экспертизы.

В зависимости от степени согласованности мнений экспертов коэффициент конкордации может принимать значения от 0 (при отсутствии согласованности) до 1 (при полном единодушии).

Если степень согласованности мнений экспертов оказывается неудовлетворительной, принимают специальные меры для ее повышения. Сводятся они, в основном, к проведению тренировок с обсуждением результатов и разбором ошибок. Если возможности для предварительной подготовки экспертов нет, измерение экспертным методом проводится по методу Дельфы. Характерными чертами этого метода являются:

- анонимность (эксперты не встречаются друг с другом, чтобы избежать влияния авторитета и красноречия кого-либо из них);
- многоэтапность (после каждого тура опроса все эксперты знакомятся с мнениями друг друга и при необходимости представляют письменные обоснования своих точек зрения). Соглашаясь или не соглашаясь с мнениями своих коллег, они могут пересматривать свою точку зрения;
- контроль (после каждого тура проверяется согласованность мнений экспертов до тех пор, пока разброс отдельных мнений не снизится до заранее выбранного значения).

Если ранжирование проводилось с целью определения весовых коэффициентов g_i , то они рассчитываются по приведенной формуле, в которой $G_{i,j}$ – ранг j -го показателя, установленный i -м экспертом.

Сопоставление бывает последовательным и попарным.

Последовательное сопоставление каждого объекта экспертизы с совокупностью всех тех, которые ниже рангом, позволяет откорректировать ранжированный ряд, уточнить позиции входящих в него объектов с учетом их важности. Оно имеет смысл тогда, когда несколько объектов экспертизы можно рассматривать как один составной объект той же природы. Порядок последовательного сопоставления следующий:

1. Объекты экспертизы располагаются в порядке их предпочтения (ранжирование).

2. Наиболее важному объекту приписывается наибольший балл или весовой коэффициент, всем остальным в порядке уменьшения их относительной значимости.

3. Сопоставляется первый объект с совокупностью всех остальных. Если по мнению эксперта он предпочтительнее, чем совокупность всех остальных вместе взятых, то результат его измерения в баллах или весовой коэффициент корректируется в сторону увеличения с таким расчетом, чтобы он стал больше (иногда определяют и на сколько больше) суммы баллов или весовых коэффициентов всех остальных объектов экспертизы, которые ниже рангом. В противном случае результат измерения или весовой коэффициент первого

объекта корректируется в сторону уменьшения так, чтобы он оказался меньше суммы баллов или весовых коэффициентов остальных объектов.

4. Сопоставляется второй объект с совокупностью всех остальных, стоящих ниже рангом. По установленному выше правилу корректируется результат его измерения или значение весового коэффициента (при этом нужно следить, чтобы не нарушилось предпочтение первого объекта перед совокупностью всех остальных, если оно установлено на предыдущем этапе). Такая процедура сопоставлений и корректировок продолжается вплоть до предпоследнего объекта.

5. Полученные результаты измерений или весовые коэффициенты нормируют, т.е. делят на общую сумму баллов или весовых коэффициентов. После этого они принимают значения в пределах от 0 до 1, их сумма становится равной 1.

Попарное сопоставление, самое простое и наиболее оправданное с психологической точки зрения.

Балл j -го объекта или весомость j -го показателя рассчитывается по формуле для последовательного сопоставления. В данном случае:

$$G_{i,j} = F_{i,j} / C$$

где $F_{i,j}$ – частота предпочтения i -м экспертом j -го объекта экспертизы; C – общее число суждений одного эксперта, связанное с числом объектов экспертизы m (числом измеряемых показателей или коэффициентов весомости) соотношением

$$C = \frac{m \cdot (m-1)}{2}$$

Если значение $\sum G_j = 1$, можно рассматривать G_j уже как нормированные и, в частности, использовать как весовые коэффициенты.

Опыт попарного сопоставления показывает, что в силу особенностей человеческой психики эксперты иногда бессознательно отдают предпочтение не тому объекту в очередной рассматриваемой паре, который важнее, а тому, который стоит в перечне первым. Чтобы избежать этого, используют свободную часть таблицы и проводят попарное сопоставление дважды

(например, сначала первого объекта со вторым, третьим, четвертым и т.д., затем второго с первым, третьим, четвертым – и так до последнего, а потом в обратном порядке: последнего с предпоследним и т.д. до первого, предпоследнего с последним, предыдущим и вновь до первого. Таким образом, каждая пара объектов сопоставляется дважды, причем в разном порядке и по истечении некоторого времени). При таком сопоставлении, называемом полным или двойным, удастся избежать случайных ошибок, кроме того, выявить экспертов, небрежно относящихся к своим обязанностям или не имеющих определенной точки зрения. Иначе говоря, двойное попарное сопоставление обладает более высокой надежностью, чем однократное. Порядок расчетов при нем остается прежним, за исключением того, что:

$$C = m * (m - 1).$$

Уточнить результаты измерений или значения весовых коэффициентов, полученные попарным сопоставлением, можно методом последовательного приближения. Первоначальные результаты рассматриваются в этом случае как первое приближение. Во втором приближении они используются как весовые коэффициенты G_j [1] суждений экспертов. Полученные с учетом этих весовых коэффициентов новые результаты в третьем приближении рассматриваются опять как весовые коэффициенты G_j тех же мнений экспертов и т.д. Согласно теореме Перрона-Фробениуса, при определенных условиях, которые на практике всегда выполняются, этот процесс сходится, т.е. нормированные результаты измерений g_j или весовые коэффициенты стремятся к некоторым постоянным значениям, строго отражающим соотношения между объектами экспертизы при установленных экспертами исходных данных.

Каждый из приведенных способов экспертной оценки имеет наряду с преимуществами определенные недостатки. Анализ показал возможность использования конкретных методов при проведении экспертизы по выбранным группам показателей.

Таблица 1

Метод	Способы расчета	Применение	Рекомендации
Метод группового опроса	Рассчитывается среднеарифметическое значение определяемых показателей	Для определения различных видов показателей	Применяется тогда, когда коллектив экспертов однороден в смысле характера ответов
Метод Дельфы	Формирование группового мнения путем проведения много турового опроса экспертов	Для определения различных видов показателей, предпочтительно – для определения прогнозируемых величин	Применять для определения нормируемых значений единичных и обобщенных показателей качества изделий
Метод непосредственных измерений	<p>Определяется усредненная оценка (вес) для каждого показателя</p> $q_i = \frac{\sum_{j=1}^m G_{ij}}{\sum_{j=1}^m G_{ij}}$ <p>(Ф.1.1)</p>	<p>Для расчета значений коэффициентов весомости.</p> <p>При $W \gg 0.5$ принимаются значения q_i, при $W \ll 0.5$ производится оценка другим</p>	Применять при высокой компетентности экспертов, для расчета значений коэффициентов весомости единичных показателей качества

С проверкой согласованности мнений	<p>Определяется согласованность мнений</p> $W = 12 S / n^2 (m^3 - m)$	способом	
Ранжирование	<p>Составляется ранжированный ряд, определяется ранг показателя. ($\Phi 1$), где G_{ij} – ранг показателя</p>	<p>Для расчета значений коэффициентов весомости</p>	<p>Применяется совместно с методом непосредственных измерений</p>
Сопоставление последовательное	<p>Сопоставляется каждый объект с совокупностью всех тех, что ниже рангом</p> $\sigma_i = F_{ij} / C,$ <p>где</p> $C = m * (m - 1) / 2$	<p>Для расчета значений коэффициентов весомости</p> <p>Если $\sum \sigma_i = 1$, σ_i является уже нормированным и используется</p>	<p>В связи с большой трудоемкостью данного метода целесообразно применять его при $n < 6$, или разбить показатели по группам</p>
Сопоставление попарное	<p>Сопоставляются пары объектов в прямой последовательности</p> $\sigma_i = F_{ij} / C,$ <p>где</p> $C = m * (m - 1) / 2$	как весовой коэффициент	<p>Применяется в том числе при невозможности использования метода последовательного сопоставления. Возможно предпочтение не того объекта</p>

Сопоставление двойное	Сопоставляются пары объектов в прямой и обратной последовательности $\sigma_i = F_{ij} / C$, где $C = M(m - 1)$ (ф. 1.4)		Удается избежать случайных ошибок возможных при попарном сопоставлении
-----------------------	---	--	--

Анализ применяемых экспертных методов (таблица 1) показал:

- наиболее простым в применении на практике является метод групповых решений. Метод групповых решений (метод комиссий) при определении показателей по группам изделий и коэффициентов весомости по подгруппам показателей является наиболее доступным, не требует проведения дополнительных мероприятий. Групповая оценка с помощью средних величин применима только тогда, когда коллектив экспертов однороден в смысле характера ответов;

- метод Дельфы также удобен для применения при экспертном оценивании средств и систем образовательного назначения в сфере информатизации.

В основу метода Дельфы положены следующие предпосылки:

- поставленные вопросы должны допускать возможность выражения ответа в виде числа;

- эксперты должны располагать достаточной информацией для того, чтобы дать оценку;

- ответ на каждый из вопросов (оценка) должен быть обоснован экспертом.

Применение метода Дельфы предполагает учет квалификации экспертов, наличие специальных знаний и подготовки специалистов.

Таким образом, при формировании экспертных групп для проведения экспертной оценки качества ПОН руководствуются следующими положениями:

- исходная численность экспертной группы составляет обычно не менее трех человек;
- эксперты должны обладать навыками работы с данным видом продукции;
- эксперты должны пройти специальную подготовку и иметь «Аттестаты экспертов» Системы добровольной сертификации аппаратно-программных и информационных комплексов образовательного назначения;
- необходимость проведения оценки достоверности полученных результатов и оценки согласованности мнений экспертов определяет руководитель экспертной группы.

При экспертном оценивании продукции определяются группы показателей и требования:

- к составу технических средств и документации ПОН;
- к характеристикам идентификации ПОН;
- к характеристикам безопасности технических средств ПОН;
- к техническим, функциональным характеристикам и характеристикам совместимости ПОН;
- к системным характеристикам ПОН;
- к характеристикам технического обслуживания ПОН;
- к характеристикам базового ПОН;
- к тестам и правилам их применения;
- к методам оценки характеристик качества ПОН.

Методические подходы и принципы, применяемые при экспертном оценивании показателей ПОН, для некоторых из групп представлены ниже.

Идентификация и документация ПОН

Экспертный метод включает в себя проведение проверки наличия необходимой документации и проверки содержания по каждому пункту требований.

Характеристиками идентификации являются:

- эксплуатационные документы;
- сведения о разработчике и версии;
- сведения о регистрации;
- описание контрольных вариантов для проверки корректности функционирования.

Характеристики ПОН

Экспертным методом осуществляется проверка содержания по каждому пункту требований, проверка соответствия представленных в документации характеристик установленным нормам.

Характеристики ПОН подразделяются:

- на технико-технологические;
- на эргономические;
- на содержательно-педагогические.

Экспертное оценивание показателей ПОН и значений этих показателей осуществляются по разработанным методикам. Общий подход к проведению экспертиз по группам некоторых наиболее значимых показателей при оценке характеристик продукции представлен в таблице 2.

Таблица 2

Состав характеристик и методы их оценки

Оцениваемая характеристика	Метод оценки
Характеристика функционального назначения	По каждому нормативному, правовому документу задаются определенные наборы входных данных, результаты оцениваются по содержимому экрана и выходных форм на бумажных носителях
Функции обработки данных	По каждому пункту требований задается несколько различных наборов данных, результаты оцениваются по содержимому экрана.

Характеристика информационной совместимости	1. Проверка наличия описания форматов файлов обмена, описания допустимых характеристик; 2. Проверка наличия словарей при вводе данных
Характеристика целостности и сохранности программ и данных	Непосредственная проверка контроля неизменяемости программ и данных при архивировании и восстановлении архивирования данных по команде администратора и автоматически через заданные промежутки времени, фактического разграничения прав доступа всех категорий пользователей к режимам и данным.
Максимальное время: - ожидания отклика на помощь пользователю; - поиска по запросу	Определение времени: – ожидания отклика на запрос пользователя; – поиска по случайному запросу
Содержание	Оценить соответствие издания регламентируемым Министерством образования РФ учебным планам, нормативным требованиям и иным документам
Дидактические принципы	Оценить соответствие основным дидактическим принципам научности, доступности и т.д.
Уровень сложности учебного материала	Убедиться в наличии нескольких уровней сложности, соответствующих уровням усвоения учебного материала и т.д.

Особые требования при экспертном оценивании ПОН предъявляются к **эксплуатационным документам**, которые должны включать:

- формуляр;
- руководство пользователя;
- руководство оператора (руководство администратора, руководство по установке и настройке и др.) – в зависимости от типа и сложности ПОН.

Формуляр должен содержать:

- перечень предметных областей, в которых может использоваться ПОН (универсальное; для интегрированной предметной области; для конкретной предметной области; для конкретного учебного предмета; для конкретного учебного предмета и класса; для конкретной учебной программы и т. д.);

- возможный возраст обучаемого (начальная школа; основная школа; средняя школа; профессиональное образование; высшая школа);

- тип (назначение) ПОН (информационно-справочное; обучающее программное средство; программированное учебное пособие; демонстрационное программное средство; практикум (деятельностная среда); тренажер (среда) для создания тестов учителем и проведения тестирования; учебно-игровое программное средство; досуговое программное средство; информационная система для управления образованием; инструментальное программное средство и т.д.).

Экспертное заключение составляется по каждой из групп показателей и в целом о качестве ПОН, оформляется не менее чем тремя специалистами той области знания, на которую ориентирована продукция, утверждается директором Института информатизации образования Российской академии образования.

Литература

1. Бешелев С.Д., Гурвич Ф.Г. Математико-статистические методы экспертных оценок. М.: Статистика, 1980.

2. Недикова Т.Н., Фризнер Д.А. Модели приближенных оценок качества сложных систем // Надежность и качество-2002: Труды международного симпозиума. Пенза, 2002.

3. Ермолаев Ю.П., Саиткулов В.Г., Харитонов М.В., Гафуров Ю.Р. Многокритериальная оптимизация технических решений электронной аппаратуры с ограничениями на показатели качества и с применением экспертных оценок // Электронное приборостроение. Казань: КГТУ.

4. Граб В.П. Экспертная оценка показателей качества // Надежность и качество: Труды Международного симпозиума. Пенза: ПГТУ, 2005.