

## РОЛЕВОЙ ПОДХОД К ОЦЕНКЕ КАЧЕСТВА ЭКОНОМИЧЕСКИХ ЭКСПЕРТНЫХ СИСТЕМ

*Г.П. Кожевникова, А.В. Голикова*

Всероссийский заочный финансово-экономический институт,  
123995, Москва, ГСП-5, Олеко Дундича, 23,  
тел.: 144 8559; факс 144 7819, stat@vzfei.ru

Рассмотрены проблемы оценки качества экспертных систем (ЭС) с базами знаний экономического характера. Сформулированы концептуальные основы построения метрического аппарата для измерения качества ЭС. Приведена технологическая схема многокритериальной оценки ЭС с учетом комплекса требований к системе со стороны ее разработчиков, менеджеров, пользователей.

In this paper, some problems concerning the quality measurement in economic expert systems (EES) are discussed. They include conceptual basis for description of metric systems and technology scheme used to estimate EES multi-criteria quality, determined by developers, managers and users needs.

Экономические экспертные системы (ЭЭС) находят все большее применение в таких направлениях финансово-экономической деятельности как инвестирование и налогообложение, финансовое планирование, оценка риска кредитования, анализ финансового состояния корпораций, фирм и предприятий, страхование и социальное обеспечение, стратегическое управление производством, планирование и прогнозирование и т.д. В базах знаний (БЗ) таких ЭЭС воплощены знания и опыт экспертов - высококвалифицированных специалистов в тех областях экономики, для решения задач которых предназначена та или иная ЭЭС.

Практика применения ЭЭС с экономическими БЗ показала, что в большинстве случаев ЭЭС превосходит профессиональных экономистов в выборе альтернативных вариантов и точности принимаемых решений. Во многом это обусловлено качеством знаний, заложенных в ЭЭС, а также качеством применяемых в ЭЭС механизмов логического вывода на основе БЗ решений, рекомендаций, советов. Задача обеспечения качества этих двух интеллектуальных компонент является одной из главных в области ЭЭС.

Управление качеством ЭЭС и экономичностью их проектов (проекты некоторых банковских систем оцениваются в миллионы долларов) опирается в своей основе на количественный (метрический) анализ качественных характеристик и свойств системы. Однако исследования в этой области находятся лишь на начальной стадии "первичного накопления капитала" - различных моделей качества, метрик, мер, оценочных формул показателей, методов оценки, результатов их апробации на отдельных ЭЭС [1-4].

Характеризуя в целом разработанные здесь методы и средства оценивания [5], можно сделать вывод, что они неполны, плохо систематизированы, не всегда легко применимы.

Обусловлено это, прежде всего, объективной трудностью измерения качества знаний ЭЭС, например, таких свойств, как уникальность, полнота, релевантность знаний, полнота умений, мощьность и гибкость стратегии рассуждений и др. На сегодняшний день не сформирован даже набор критериальных свойств для оценки качества знаний. Более того, почти все свойства, которые считаются релевантными понятию качества БЗ, характеризуются нечетким определением, что порождает субъективизм в выборе мер и методов их измерения.

Другая проблема связана с тем, что функционирующие ЭЭС, ориентированы на ту или другую узкую предметную область, поэтому их проектные решения, включая БЗ, носят уникальный характер. Уникальность и разнообразие проектных решений еще более затрудняют создание общего аппарата их оценки.

Накопленный в программной инженерии значительный опыт создания высококачественных программных систем очень мало используется при разработке систем с БЗ. Более того, сам подход к оценке их качества базируется на субъективном видении качества системы экспертами, специалистами в узкой прикладной области, что не содействует объективной оценке уровня качества с позиции стандартов, принятых в технологии программирования.

На наш взгляд, оценку качества ЭЭС целесообразно проводить, исходя из функций людей, имеющих дело с ЭЭС на различных стадиях ее жизненного цикла (ЖЦ) - в процессе ее разработки и функционирования. Это руководители проектов ЭЭС, разработчики, включая инженеров-когнитологов, эксперты, конечные пользователи ЭЭС, администрация фирм, финансирующие заказчики и т.д. Специалисты любой из этих групп, в соответствии с их ожиданиями от ЭЭС, выдвигают те или иные требования к качеству ЭЭС, которые могут быть отражены в определенных наборах критериальных свойств системы. Включение таких "ролевых" критериев в общую модель качества ЭЭС позволяет учесть компетентность и систему предпочтений отдельных групп специалистов в итоговой комплексной оценке качества ЭЭС.

Основные шаги такого «ролевого» подхода в оценивании конкретной ЭЭС:

- формирование групп оценщиков согласно ролям специалистов в ЖЦ ЭЭС;
- разработка «ролевых» критериев оценки качества систем;
- формирование «ролевых» моделей качества и выбор соответствующего аппарата измерения показателей качества;
- разработка тестовых примеров для получения «ролевых» оценок качества;
- расчет «ролевых» оценок показателей;
- расчет комплексного (итогового) показателя качества ЭЭС;
- предоставление рекомендаций относительно улучшения качества проекта ЭЭС.

Основываясь на ролевых критериях, качество ЭЭС можно рассматривать в различных аспектах [5]. Основными из них являются следующие:

- **исходная полезность системы**, оцениваемая с точки зрения целесообразности и уместности ее применения в данной проблемной сфере, диапазона возможностей системы в плане сложности решаемых ею задач и пригодности для их решения, а также с учетом степени модернизации, вносимой системой в процесс принятия решений «вручную»;

- **уровень интеллектуальности системы**, оцениваемый с точки зрения информативности и логической мощности ее БЗ, правильности и эффективности стратегий логических выводов, достоверности приобретаемых и выводимых знаний, качества вырабатываемых решений и рекомендаций и ряда других аналогичных свойств;

- **производительность системы**, включая оценки быстродействия вырабатываемых решений, эффективность управления выводом и ведения БЗ, ресурсные оценки поисковых процедур и других программных процедур системы;

- **комфортность работы с системой**, оцениваемая с точки зрения удобства ее эксплуатации и сопровождения, дружелюбности пользовательского интерфейса, наличия и качества различных средств помощи (подсказок, пояснений и др.), легкости использования тестовых примеров и т.д.;

- **экономичность системы**, включая оценки рентабельности затрат на ее создание и функционирование, повышение уровня производительности труда специалистов, стратегическую привлекательность в аспекте конкурентоспособности фирмы и др.;

- **социальные аспекты** в плане оценки влияния ЭЭС на среду ее применения: стимулирование пользователей к работе с системой, сокращение их рабочей нагрузки, полезность и действенность экспертов при проектировании и функционировании системы, поддержка разработки и эксплуатации системы высшим руководством и администрацией фирмы и др.

Заключительную оценку качества функционирования системы дает пользователь, и его оценка, по существу, предопределяет судьбу системы с точки зрения ее практического использования. Поэтому в системе предпочтений видения качества ЭЭС разными группами специалистов требования пользователей должны иметь в оценках систем наибольшие весовые коэффициенты.

Для пользовательской модели качества наиболее важны следующие свойства:

- готовность системы отвечать на разнообразные вопросы, касающиеся предметной области;
- правильность и полнота ответов;
- эффективность с точки зрения скорости выработки решений (рекомендаций, советов);
- комфортность взаимодействия с системой.

Пользовательская модель качества обеспечивается определенным набором критериальных свойств, важных для разработчиков – когнитологов и экспертов. В первую очередь к ним относятся:

- информативность и логическая мощность БЗ;
- правильность и эффективность стратегий логических выводов;
- достоверность приобретаемых и выводимых знаний;
- качество вырабатываемых решений и рекомендаций;
- дружелюбность пользовательского интерфейса.

Модель последнего свойства можно считать устоявшейся. Она включает следующие главные критерии:

- способность системы объяснять пользователю ход логических рассуждений в процессе выработки решений и рекомендаций, давать ответы на вопросы, «как» получено решение и «почему именно так»;
- возможность для пользователя возврата к предыдущим шагам логического вывода и попытки решения задачи с другими параметрами (т.е. прогнозирования по типу «что-если»);
- способность системы учитывать в своих ответах и пояснениях уровень подготовки пользователя;
- наличие и качество различных средств помощи – подсказок, пояснений, документации и т.д.;
- легкость в использовании тестовых примеров;
- привычный для пользователя профессиональный лексикон в вопросах, ответах и сообщениях системы;
- способность системы вырабатывать решения и рекомендации (а также обучать пользователя) на основе способов, общепринятых в сфере деятельности пользователя и являющихся привычными для него (что повышает доверие пользователя к полученным решениям и рекомендациям);
- наличие средств построения различного рода меню, экранных форм, графиков и т.д.

Критериальные модели остальных четырех свойств уровня когнитологов и экспертов пока находятся в стадии развития и весьма далеки от своего завершения. В них включают такие свойства: релевантность знаний

(их подтвержденность и достаточность), полнота умений (необходимость и достаточность процедур), сложность проблемы (сложность базы знаний и метода решения), проверка знания (их тестирование и целостность), уровень интеллектуальности (обучаемость, гибкость стратегий рассуждения и интерфейса), мощность стратегий рассуждений, наличие метазнаний и др.

Количественный анализ качества ЭЭС регламентируется аппаратом метрической системы (МС), формируемой для оценивания данной конкретной ЭЭС [5].

МС строится как совокупность 6-ти компонент:  $\langle S, R, S^*, K, F, H, F^* \rangle$ :

$S = \{s_i\}$  - совокупность свойств ЭЭС, выступающих в роли критериев качества ЭЭС;

$R = \{r_{ik}\}$  - совокупность отношений подчиненности между свойствами  $s_i$  и  $s_k$ , порождающими на S иерархическую диаграмму подчиненности критериев  $s_i \in S$ ;

$S^* = \{s_i\}$  - совокупность факторов качества ЭЭС, которые образуют модель аттестации качества ЭЭС ( $S^* \in S$ );

$K = \{K_i\}$  - совокупность метрик  $K_i = \{k_{i1}, k_{i2}, \dots, k_{in}\}$  измерения свойств  $s_i \in S$ , где  $k_{i1}, k_{i2}, \dots, k_{in}$  – индикаторы измерения  $s_i$  (параметры, доступные в системе непосредственному измерению);

$F = \{F_i\}$  – совокупность мер свойств  $s_i$ , которые являются оценочными функциями вида  $F_i: K_i \rightarrow M_i$  где  $M_i$  – шкала измерения  $s_i$  (абсолютная, порядковая или номинальная);

$H = \{H_i\}$  – совокупность формульных представлений мер  $F_i (k_{i1}, k_{i2}, \dots, k_{in}) \in F$ , которые задают количественные значения показателей  $q_i$  свойств  $s_i (q_i \in M_i)$  в зависимости от значений индикаторов  $k_{i1}, k_{i2}, \dots, k_{in}$  метрик  $K_i$  измерения  $s_i$ ;

$F^*$  - комплексная мера качества ЭЭС для аттестующей модели  $S^* \in S$ , которая определяет агрегирующую операцию над показателями  $q_j$  свойств  $s_j \in S^*$  для получения интегрального (комплексного) показателя качества  $Q^*$  ЭЭС применительно к аттестующей модели  $S^*$ .

Таким образом, количественные оценки показателей  $Q^*$  характеризуют качество ЭЭС в терминах конкретной фиксированной метрической системы  $\langle S, R, S^*, K, F, H, F^* \rangle$ . При этом для любого из свойств  $s_i$ , выступающего в роли критерия качества ЭЭС, фиксируется метрика  $K_i = \{k_{i1}, k_{i2}, \dots, k_{in}\}$ , которая формируется из индикаторов иерархической критериальной модели свойства  $s_i$ . Выбор шкалы измерения  $M_i$  и формульного представления  $H_i$  меры  $F_i: K_i \rightarrow M_i$  приводит к показателю  $q_i = F_i (k_{i1}, k_{i2}, \dots, k_{in})$ , определяющему качество ЭЭС по критерию  $s_i$ . Комплексная мера  $F^*$  интегрирует показатели качества  $q_i$  в обобщенный показатель  $Q^*$ , оценивающий качество ЭЭС применительно к фиксированной аттестующей модели  $S = \{s_i\}$ .

При «ролевом» подходе к многоаспектной оценке качеств ЭЭС с использованием метода экспертных оценок интегральный показатель  $Q$  качества ЭЭС может быть рассчитан по следующей методике.

Этап 1. **Формирование групп экспертов** – специалистов, принимающих участие в оценке ЭЭС на разных стадиях ее ЖЦ – в соответствии с теми ролями, которые они играют в ЖЦ системы. Ознакомление ролевых групп экспертов (РГЭ) с задачей, понятийным базисом и метрической системой оценивания.

Этап 2. **Определение** для каждой  $j$ -ой РГЭ и каждого этапа ЖЦ ЭЭС *значимых критериев*  $s_i \in S$ . Формирование матрицы  $\{\tau_{ij}\}$  распределения по РГЭ значимых критериев:

$$\tau_{ij} = \begin{cases} 1, & \text{если } i\text{-ый критерий является значимым для } j\text{-ой РГЭ,} \\ 0, & \text{в противном случае.} \end{cases}$$

Этап 3. **Ранжирование значимых критериев**  $s_i$  по степени их важности для каждой РГЭ. Обобщение результатов ранжирования  $s_i$  во всех группах РГЭ и расчет для каждого критерия  $s_i$  его весового коэффициента  $z_i$  в интегральной оценке ЭЭС.

Этап 4. **Экспертная оценка** специалистами каждой РГЭ показателей  $q_i$  свойств  $s_i$ , значимых для этой РГЭ, путем присвоения показателям числовых значений из диапазона  $[0,1]$ . Определение для  $j$ -ой РГЭ усредненного группового значения  $\bar{q}_{ij} \in [0 - 1]$  по количеству экспертов в  $j$ -ой РГЭ:

$$\bar{q}_{ij} = \frac{\sum_1^j q_{ij}}{m_j} \quad (1)$$

Этап 5. **Проверка согласованности групповых оценок по критериям**  $s_i$ , оцениваемым различными РГЭ. Для этого рассчитываются следующие статистические характеристики:

- среднее (межгрупповое) значение  $\bar{Q}_i$  показателей  $\bar{q}_{ij}$ , взятое по всем РГЭ;
- дисперсия  $D_i$  групповых оценок  $\bar{q}_{ij}$  относительно межгрупповой оценки  $\bar{Q}_i$ ;
- среднее квадратическое отклонение  $\sigma_i$  групповых оценок  $\bar{q}_{ij}$  от средней межгрупповой оценки  $\bar{Q}_i$ .

Согласованность межгрупповых оценок показателей  $\bar{Q}_i$  определяется величиной коэффициента вариации  $CV_i$ :

$$CV_i = \frac{\sigma_i}{Q_i} * 100\%. \quad (2)$$

Етап 6. **Проверка согласованности групповых оценок по всем  $k_i$  экспертам** показателей  $q_i$  всех РГЭ. Согласованность определяется величиной коэффициента конкордации  $CW_i$  ( по Кенделлу ).

$$CW_i = \frac{12 S}{n_i^2 (k_i^3 - k_i)}. \quad (3)$$

При этом имеется возможность введения весовых коэффициентов  $p_i$  для каждой РГЭ с целью расчета степени доминирования суждений специалистов данной РГЭ по оценке  $q_i$  на данном этапе ЖЦ ЭЭС. Если ряд показателей  $q_i$  имеет связанные ранги, то  $CW_i$  рассчитывается по формуле

$$CW_i = \frac{S}{\left( \frac{1}{12} n_i^2 (k_i^3 - k_i) - n_i \sum_{j=1}^{n_j} T_j \right) * \left( \frac{1}{n_i} \sum_{j=1}^{n_j} P_j \right)}, \quad (4)$$

где  $T_j = \frac{1}{12} \sum_{t_j} t_j (t_j^2 - 1)$ ,  $t_j$  - количество совпавших оценок (рангов) в  $j$ -ой РГЭ.

Величина  $CW_i \leq 0,2 \div 0,4$  свидетельствует о слабой согласованности межгрупповых оценок  $q_i$ , величина  $CW_i > 0,6 \div 0,8$  - о сильной их согласованности.

Если имеется достаточная согласованность результатов оценивания, то выполняются этапы 7 – 9, в противном случае – этап 10.

Етап 7. **Расчет обобщенного среднего значения показателей  $q_i$**  по всем критериям  $s_i$ , характеризующим данную компоненту ЭЭС. Расчетная формула:

$$\overline{Q_{ir}} = \frac{1}{k_j} \sum_{i=1}^{k_j} q_{ij}; \quad (r=1, \dots, N_r^i), \quad (5)$$

где  $N_r^i$  - количество критериев  $s_i$ , характеризующих  $r$ -ую компоненту ЭЭС.

Етап 8. **Комплексная оценка отдельных компонент ЭЭС.** Комплексное значение показателя  $Q_r$  качества  $r$ -ой компоненты ЭЭС оценивается с учетом весовых коэффициентов  $z_i$  критериев  $s_i$ :

$$Q_r = \frac{1}{N_r^{(i)}} \sum_{i=1}^{N_r^{(i)}} z_i \overline{Q_{ir}}. \quad (6)$$

Етап 9. **Интегральная оценка качества ЭЭС.** По результатам оценки каждой отдельной компоненты ЭЭС рассчитывается количественное значение интегрального показателя  $Q$  качества ЭЭС:

$$Q = \frac{1}{N_0} \sum_{r=1}^{N_0} Q_r, \quad (7)$$

где  $N_0$  – количество оцененных компонент ЭЭС.

Етап 10. **Повторная оценка несогласованных показателей  $q_i$ .** Анализируются причины и факторы, которые влияют на несогласованность суждений экспертов об оценках тех показателей  $q_i$ , которые имеют большой коэффициент вариации  $CV_i$ , после чего проводится их повторная оценка.

Если в результате межгрупповых консультаций и повторных туров опроса экспертов согласованность оценок  $q_i$  будет достигнута, то выполняются этапы 7 – 9. В противоположном случае обобщенное значение  $Q_r$  критерия  $s_i$  не вычисляется и по результатам экспертного опроса формулируют (корректируют) спецификации требований к следующему прототипу ЭЭС.

Рассчитанный по формулам (1) – (7) методике интегральный показатель  $Q$  принимает значения в диапазоне [0-1], что позволяет сравнить качество рассмотренного варианта ЭЭС с ее «идеальной версией», которая характеризуется значением  $Q = 1$ , или с аналогом, для которого известно или рассчитано аналогичным образом значение  $Q$ .

Предложенная методика апробирована применительно к оценкам 19-и реально эксплуатируемых ЭЭС в сфере страхования [6].

Экспертные оценки были получены от трех групп специалистов – 67 пользователей, 51 разработчика и 44 менеджера проектов ЭЭС. Множество  $S=\{s_i\}$  оцениваемых свойств качества ЭЭС включало 39 свойств, характеризующих четыре категории критериев качества:

**решаемые ЭЭС задачи** – по трудности, полноте функций ЭЭС полезности, новизне и др.;

**используемые в ЭЭС технологии** – по интерактивности, оперативности, точности, дружелюбности пользователю, наличию пояснений к выведенным ЭЭС умозаключениям и др.;

**человеческий фактор** – по уровню помощи в эксплуатации ЭЭС, полезности для разработки ЭЭС консультаций экспертов – пользователей, стимулированию использования ЭЭС в страховых компаниях и др.;

**организационные стороны ЭЭС** – по конкурентоспособности, повышению производительности труда пользователей ЭЭС, адаптируемости ЭЭС, инновациям и др.

Для оценки свойств использовалась оценочная шкала [0-1], веса важности свойств присваивались всеми тремя группами оценщиков применительно ко всем анализируемым 39 свойствам.

Время, затраченное экспертами на оценку свойств, в среднем составляло 30 мин. Некоторые свойства не были оценены (порядка 15% всех свойств, в основном не оценены пользователями). Отсутствие оценок ряда свойств, по-видимому, обусловлено недостаточной четкостью и громоздкостью формулировок этих свойств.

Интегральные показатели качества оцениваемых ЭЭС расположились в диапазоне [0-1], в основном в его центральной зоне – от 0,43 до 0,78.

Апробация предложенной методики показала ее практическую полезность для решения проблем оценки качества ЭЭС, а также направления дальнейшего развития.

В рассмотренной методике оценки интегрального качества ЭЭС на этапе 4 может быть применен **логико-лингвистический подход** к оценкам показателей  $q_i$  с использованием лингвистических (словесных) значений  $L_i$  показателей  $q_i$  и функций принадлежности, которые ставят в соответствие каждому лингвистическому описанию  $L_i$  показателя  $q_i$  соответствующее число в диапазоне [0-1].

Использование нечетких словесных понятий, которыми оперируют эксперты, позволяет ввести в процесс оценки наряду с количественными оценками качественные описания и учитывать тем самым неопределенность задачи оценки. Цель лингвистической оценки – достижение полноты анализа всех факторов, которые имеют отношение к оценке ЭЭС, но не поддаются точному количественному выражению.

1. *Слисенко А.О.* К проблеме анализа логико-алгоритмической структуры систем представления и обработки знаний // Вопросы кибернетики. Теоретические аспекты и инструментальные средства прикладных интеллектуальных систем. – М.: НС по проблеме «Кибернетика». – 1989. – С. 85-97.
2. *Chen Z., Suen C.* Measuring the Complexity of Rule-Based Expert Systems // Expert Systems with Applications. – 1994. – v. 7. – № 4. – P. 467-481.
3. *Guida G., Mauri G.* Evaluating Performance and Quality of Knowledge-Based Systems: Foundation and methodology // IEEE Trans. Knowledge and Data Eng. – 1993. – V. 5. – №2. – P. 204-244.
4. *Kaisler S.* Expert System Metrics // Proc. of the IEEE Intern. Conf. on Systems, Man and Cybernetics. – 1986. – v. 1. – P. 114-120.
5. *Кожевникова Г.П.* К построению метрического аппарата оценки качества систем с базами знаний // Праці І-ї Міжнар. наук.-практ. конф. УкрПРОГ'98. - Киев: ИК АН Украины. – 1998. – С. 351-357.
6. *Sharma R.S., Conrath D.W., Dilts D.M.*, A socio-technical model for deploying expert systems. Part I: the general theory // IEEE Transactions on Engineering Management, February 1991. – 38 (1), P. 14-23; 1991. – V. 38. – № 1. – P. 14-23.