Обобщенные интервальные оценки в моделях предметных областей систем поддержки экспертных решений*

М. Ю. Стернин[1], Н. В. Чугунов[2], Г. И. Шепелев[3]

Рассмотрены особенности применения в системах поддержки экспертных решений математических моделей некоторых предметных областей. В качестве представительного примера такой предметной области выбрана проблематика оценки ожидаемой перспективности месторождений углеводородов, находящихся на ранних стадиях изученности. Для расширения возможностей представления и использования экспертных знаний в моделях при решении различных задач подобных областей развивается метод обобщенных интервальных оценок (ОИО). Предложены процедуры агрегирования оценок, сделанных разными экспертами. Изучены возможности применения метода ОИО к зависимым величинам, в том числе при анализе динамики изменения их значений.

Введение

С усложнением проблемных ситуаций, увеличением числа и значимости междисциплинарных задач значительно возросла роль специалистов-экспертов, располагающих профессиональными знаниями, опытом и навыками в соответствующих предметных областях. Потребность в эффективном регулярном использовании этих интеллектуальных ресурсов посредством их «технологизации» материализовалась в двух классах компьютерных систем, основанных на знаниях, — экспертных системах и системах поддержки экспертных решений (СПЭР).

Указанные компьютерные системы имеют различную целевую направленность. Экспертные системы предназначены для помощи в решении

* Работа частично поддержана грантом Президента Российской Федерации для поддержки ведущих научных школ НШ 1964.2003.1, Российским фондом фундаментальных исследований (проекты 04-01-00290, 05-01-00666), Российской академией наук (программы фундаментальных исследований РАН «Математическое моделирование и интеллектуальные системы» и ОИТВС РАН «Фундаментальные основы информационных технологий и систем»).

[1] 117312, Москва, проспект 60-летия Октября, 9, ИСА РАН, mister@isa.ru.
[2] 117312, Москва, проспект 60-летия Октября, 9, ИСА РАН, nvc @isa.ru.
[3] 117312, Москва, проспект 60-летия Октября, 9, ИСА РАН, gis@isa.ru.

повторяющихся задач предметных областей, не располагающих развитыми формализованными («объективными» [Ларичев, 1987]) моделями. Они ориентированы на выявление и перенос «процедурных» знаний первоклассных специалистов в компьютеры, многократное использование этих знаний менее опытными пользователями в отсутствие экспертов [Ларичев и др., 1987а]. Эти системы, таким образом, имитируют деятельность высоко квалифицированного эксперта при решении, прежде всего, слабо структурированных проблем, когда интуиция человека и его опыт приобретают особую ценность. Они во многом утрачивают свою полезность, если такой эксперт может быть непосредственно привлечен к анализу проблемы.

СПЭР способствуют повышению эффективности труда самого эксперта в процессе анализа проблем. Взаимодействие квалифицированного эксперта и СПЭР позволяет всесторонне исследовать такие задачи, решение которых ранее было либо неудовлетворительно приближенным, либо требовало неприемлемо больших затрат времени и труда. Характерной особенностью СПЭР является наличие в их составе базы моделей предметной области, на решение задач которой система ориентирована. В базе моделей для описания одних и тех же процессов предметной области могут храниться «объективные» модели различной степени агрегированности — от приближенных, «быстрых» экспресс-моделей до детальных, теоретически более обоснованных, но требующих значительных объемов исходной информации. Различной может быть и толерантность моделей к необходимой достоверности исходной информации. Вместе с тем, СПЭР может содержать также «субъективные» модели эксперта, отражающие его опыт решения задач предметной области. Их можно хранить как в базе моделей, так и в базе знаний, связанных с базой данных, где, наряду с фактографической информацией, полезно разместить наиболее удачные предыдущие решения сходных задач («аналоги») и хранить промежуточные результаты решения текущей задачи или различные варианты решения. В ряде случаев субъективные ранее не формализованные суждения эксперта могут быть формализованы с помощью специальных методов и после этого «равноправно» использоваться вместе с объективными моделями при решении задач. «Сосуществование» в СПЭР объективной и субъективной информации, объективных и субъективных моделей и необходимость их совместного и согласованного использования требуют наличия развитого интерфейса, обеспечивающего возможность продуктивного диалога эксперта с системой. СПЭР, таким образом, могут использоваться при решении задач «частично» структурированных предметных областей, т. е. областей, располагающих отдельными объективными моделями, которые описывают решаемые задачи не исчерпывающим образом, а также «хорошо» структурированных областей, адекватное использование объективных моделей которых может требовать, по ряду причин, диалога эксперта с системой. Чаще всего эти причины обусловлены недостаточной точностью исходной информации, используемой в моделях. Пример подобной ситуации доставляют задачи оценки перспективности некоторых объектов природопользования на ранних стадиях их изученности, исследованные авторами совместно с В. Пороскуном ранее [Пороскун и др., 1999; 2002].

Модели СПЭР с экспертно задаваемыми числовыми исходными данными

Во многих областях человеческой деятельности — науке, технике, бизнесе — широко распространены проблемные ситуации, которые могут быть описаны исходными данными (параметрами), измеримыми в количественных шкалах, т. е. данными, представимыми числовыми оценками. Зачастую, однако, описание таких ситуаций «точечными» величинами (оценками, задаваемыми для каждого параметра одним числом) оказывается неадекватным достижению желаемой цели — расчету значений результирующих показателей исследуемой задачи в виде, способствующем принятию обоснованных решений. Недостаточно точное априорное знание исходных параметров требует тогда, чтобы для целей анализа они были представлены как некоторые связные области на шкалах, характеризующих возможные допустимые в анализируемой ситуации значения каждого параметра.

Исходные данные D, таким образом, представляются интервалами («интервальными числами») $[D_l, D_r]$, задаваемыми их левыми D_l и правыми D_r границами. Результирующие показатели, рассчитанные на объективных моделях предметной области, которые связывают исходные данные с показателями, также оказываются теперь интервальными числами. Подобного рода ситуации, когда исходные параметры представлены как интервалы, типичны для естественных, инженерно-технических наук и техники, где измерениям принципиально присуща некоторая погрешность, которую требуется учитывать в дальнейших расчетах, в том числе в расчетах искомых значений непосредственно неизмеримых величин. В экономических исследованиях, бизнесе в виде интервальных чисел могут быть представлены прогнозируемые параметры, при этом погрешности значения параметров, отвечающие прошлым и текущему моментам времени, часто известны вполне точно.

Один из аспектов отмеченного выше усиления роли экспертов — практическая возможность получения значений многих исходных данных требующих решения задач преимущественно в виде экспертных оценок, формируемых на базе экспертных знаний, суждений по аналогии и т. п.

Широкий класс экспертных оценок во многих прикладных задачах доставляют именно интервальные экспертные оценки. Актуальность введения в практику экспертиз методов сбора и обработки интервальных экспертных оценок подчеркивалась, например, в работе [Орлов, 1996]. Целесообразным также оказалось интервальное задание экспертом значений исходных параметров в задачах подсчета запасов углеводородов и прогнозирования перспективности нефтесодержащих объектов на ранних стадиях их изученности [Пороскун и др., 1999; 2002].

Однако для того, чтобы подобные неформализованные экспертные знания могли быть использованы в СПЭР, следует получить их математическое представление. Это обстоятельство стимулирует развитие различных методов анализа данных, преобразующих исходные параметры в формализованные структуры, методов выявления и представления экспертных знаний. Один из таких методов, разработанный авторами метод обобщенных интервальных оценок, представлен в работах [Стернин и др., 2003; Shepelyov et al., 2003] и развит в этой статье.

Для числовых исходных данных их интервальное задание отвечает ситуации с наибольшей неопределенностью. Диапазоны допустимых изменений значений результирующих показателей моделей оказываются при этом, как правило, чрезвычайно широкими. Этот недостаток частично устраняется, а возможности количественного, математического анализа исходных данных становятся гораздо богаче, если эксперт в дополнение к интервальной оценке параметра выдвинет гипотезы о шансах на реализацию тех или иных значений в заданном интервале $[D_l, D_r]$. Достаточно распространенным на практике математическим аппаратом анализа неопределенности числовых данных являются теоретико-вероятностные методы.

Хотя возможности анализа свойств изучаемой задачи, связанных с описанием присущей ей неопределенности, не ограничиваются теоретико-вероятностными методами, в ряде случаев применение именно их наиболее оправдано. Основная причина этого состоит в том, что, как показало обобщение практического опыта работы с экспертами, использующими СПЭР, экспертный анализ наиболее продуктивен, если диалог с системой ведется на привычном для пользователя языке, с использованием знакомой ему терминологии [Петровский, 1996]. Не отрицая привлекательных черт других методов анализа неопределенности, следует признать, что анализ, базирующийся на теоретико-вероятностных понятиях, более распространен в среде специалистов-практиков различных предметных областей, в частности, среди специалистов-нефтяников. Именно эти понятия образуют часть их профессионального языка. Привычность теоретико-вероятностного инструментария для основного круга экспертов, занимающихся геолого-экономической оценкой, — геологов, технологов, экономистов, понимание ими математической и содержательной структуры используемого аппарата, а, в силу этого, влияния возможных изменений в описании исходных данных на результирующие показатели, оказывается существенным для продуктивности диалога «эксперт — СПЭР».

В ряде случаев преимущественное применение при моделировании того или иного математического аппарата может быть связано и с некоторыми формальными обстоятельствами. Такой формальный, но существенный аргумент в пользу использования в качестве меры достоверности различных суждений на интервалах изменения исходных параметров теоретико-вероятностных конструкций имеется, например, в задачах оценки запасов углеводородов. Дело в том, что международная классификация запасов основывается именно на теоретико-вероятностных представлениях [Золотухин, 2001].

В задачах оценки запасов исследуемые параметры были исходными данными некоторых объективных моделей предметной области, значительная часть исходной информации получалась либо непосредственно от эксперта, либо выбиралась им из списка, содержащегося в базе данных СПЭР. В зависимости от точности имеющейся информации исходные параметры моделей, их результирующие показатели могли трактоваться экспертом, в порядке увеличения неопределенности, как точечные величины (обычные числа); случайные величины или интервальные числа.

Модели предметной области при этом имели одну и ту же математическую структуру как в точечной, так и в стохастической, и в интервальной интерпретациях. Аналогичность математической структуры в данном случае означает, что одноименные модели различных интерпретаций обладают одинаковыми исходными параметрами и связывающими их операциями, а также теми же результирующими показателями. При переходе от одной интерпретации к другой меняются лишь точки зрения на фигурирующие в моделях величины, которые рассматриваются либо как обычные числа, либо как случайные величины или интервальные числа. Для получения результирующих показателей при этом используются обычные арифметические операции, теоретико-вероятностный метод статистических испытаний (метод Монте-Карло) и методы интервального анализа соответственно.

Основными методами учета неопределенности при решении задач в рассматриваемой СПЭР были методы теоретико-вероятностного моделирования. Параметры моделей трактовались экспертами, как правило, как независимые случайные величины. Расчетам по вероятностным моделям предшествовали расчеты по интервальным и детерминистским версиям моделей. Такие вспомогательные расчеты помогали эксперту оценить границы диапазонов, в которых по имеющейся информации могли находиться результирующие показатели. Эти границы соответствуют худшим и лучшим (допустимым по имеющейся информации) вариантам прогноза результатов. Анализ результатов расчетов по детерминистским геолого-экономическим моделям с исходными данными, отвечающими наиболее вероятным, по мнению эксперта, значениям, позволяли эксперту получить представление о правдоподобности предположений о числовых характеристиках исходных параметров вероятностных моделей. Существенную роль в этом процессе играл также анализ чувствительности результирующих показателей по отношению к возможным изменениям исходных параметров.

Для каждого из параметров задачи, рассматриваемых как случайные величины, эксперт выбирал тип распределения вероятности, отражающий его представления о шансах на возникновение тех или иных диапазонов значений применительно к анализируемому объекту. Задавались минимальное, наиболее вероятное и максимально возможное, согласно имеющейся информации, значения оцениваемого параметра на объекте, специфицируя (априорное) распределение вероятности ранее выбранного типа. Среди типов распределений наиболее широко использовались экспертами равномерное, нормальное, треугольное и β-распределение.

Поскольку анализируемые нефтегазовые объекты во многих отношениях уникальны, было признано, что при привлечении теоретико-вероятностного моделирования для оценки запасов, подсчета других результирующих показателей традиционное применение средних по распределению величин («математическое ожидание»), адекватных массовым повторяющимся явлениям, не всегда оправдано.

Более адекватным подходом в указанных условиях представляется использование дихотомических оценок, разделяющих весь интервал возможных значений результирующего показателя на две части, — одну, содержащую все исходы, классифицируемые экспертом как благоприятные, и другую, включающую все неблагоприятные случаи. Такой оценкой в теоретико-вероятностной картине является пара $(L,\ P(R \geq L))$, где L — некоторое граничное значение результирующего показателя R, анализируемое в данный момент экспертом, а $P(R \geq L)$ — вероятность того, что в «игре с природой» реализуются лишь благоприятные возможности, т. е. такие, значения результирующего показателя R для которых превосходят L.

(Предполагается, что увеличение значений результирующего показателя увеличивают привлекательность возможных исходов для эксперта). Например, извлекаемые запасы углеводородов, окажутся больше, чем заданное экспертом их значение, или реализуются исходы, стоимостные оценки перспективности анализируемого объекта для которых превосходят желаемую граничную величину, и т. д.

Применительно к задачам оценки запасов эта оценка названа нами вероятностью гарантированного результата, имея в виду, что она показывает, какова вероятность допустимых, по имеющейся информации, исходов, реализация любого из которых приведет к результату R, значение которого, при прочих равных условиях, будет не хуже величины L, играющей роль эталона сравнения. Эта вероятность служит также (при наблюдаемых с неопределенностью исходных данных) мерой надежности суждений о величинах возможных запасов и перспективности освоения анализируемого нефтегазового объекта.

Введенная выше мера надежности реализации различных значений результирующего показателя напрямую связана с такой широко используемой мерой риска, как VAR (Value at Risk), а знание распределений вероятностей результирующих показателей моделей позволяет рассчитать также и значения такой часто используемой в последнее время меры риска как Shortfall [Artzner et al., 1997].

Полиинтервальные оценки исходных данных

В процессе анализа и обобщения результатов исследований, проведенных при решении задач оценки перспективности месторождений, находящихся на ранних стадиях изученности, стало ясно, что не все ситуации экспертного задания исходных параметров, измеримых в количественных шкалах, могут быть описаны в рамках «моноинтервального» подхода, подхода, при котором эксперт ограничивается заданием для каждого исходного параметра единственного интервала изменчивости данных. Оказалось, что знания эксперта о проблеме могут иметь другую природу, и он посчитает задание значений какого-либо параметра в виде единственного интервального числа неадекватным его истинной точке зрения. Фактически его знаниям об исходном параметре может отвечать задание целой совокупности интервалов. Для спецификации такой совокупности эксперт может задать оценки нескольких характерных интервалов, входящих в совокупность. Это могут быть, например, пессимистическая, оптимистическая и наиболее вероятная оценки для интервала, описывающего, по мнению эксперта, изменчивость анализируемого исходного параметра, или какие-либо другие оценки. Тогда, исходя из предъявленных экспертом оценок, можно попытаться восстановить всю совокупность интервалов. Ясно, что и при фиксированном числе оценок последующий анализ зависит от того, какие именно интервалы их совокупности выбраны в качестве характерных. Рассмотрим далее более подробно важный для приложений случай наличия экспертных оценок для двух специально выбранных характерных интервалов.

Эксперт может, например, задать оценку для наименее широкого («мини») интервала из входящих в совокупность, $I_m = [D_l^{(m)},\ D_r^{(m)}]$ длины $M = D_r^{(m)} - D_l^{(m)}$ и для наиболее широкого («базового») интервала совокупности $I_b = [D_l^{(b)},\ D_r^{(b)}]$ длины $B = D_r^{(b)} - D_l^{(b)}$. Все «промежуточные» (между I_m и I_b) интервалы I_l, такие что $I_m \subset \cdots \subset I_l \subset \cdots \subset I_b$, также входят в указанную совокупность. Длина интервала I_m в определенном смысле отражает уверенность эксперта в его знаниях о параметре D: чем лучше эксперт представляет себе вариабельность параметра, тем уже этот интервал. Вместе с тем, сужение интервала I_m увеличивает риск принятия недостаточно обоснованного решения, если знания эксперта окажутся ошибочными. (См. в этой связи работу [Kahneman et al., 1982]).

Расположим все возможные отдельные интервалы заданной таким образом экспертом совокупности на плоскости ($Y = h,\ X = D$) друг над другом, от большего (по длине) I_b к меньшему I_m, так чтобы левые границы промежуточных интервалов лежали на левой $[D_l^{(b)},\ D_l^{(m)}]$, а правые на правой $[D_r^{(m)},\ D_r^{(b)}]$ боковой стороне соответствующей криволинейной трапеции. На оси h расположены, таким образом, «метки» отдельных интервалов совокупности, а ось D показывает возможную изменчивость параметра на каждом интервале.

Нормируем высоту h этой трапеции на единицу, при этом базовый интервал совокупности отвечает $h = 0$, а мини-интервал $h = 1$. Полученную конструкцию назовем «полиинтервальной» оценкой (ПИО) возможных, по мнению эксперта, значений исходного параметра. При таком задании ПИО интервалы системы оказываются вложенными друг в друга. В общем случае это не обязательно.

Итак, ПИО определяются не только заданием базового интервала и мини-интервала, но и формой левой и правой границ соответствующей криволинейной трапеции. В простейшем, но важном для практики, случае эти границы прямолинейны. Отметим, что возможность варьирования границ ПИО и задания меры уверенности в реализации границ той или иной формы предоставляет эксперту дополнительное средство отображения своих знаний об анализируемом исходном параметре. Отметим, наконец, что ПИО доставляет пример мультимножества [Петровский, 2003] специального вида: его опорное множество совпадает с базовым интервалом, а функция кратности точек этого интервала с соответствующими ординатами границ трапеции.

Обобщенные интервальные оценки исходных данных

Построим теперь на базе ПИО такую теоретико-вероятностную модель представления экспертных знаний, которая сочетала бы в себе описание двух видов неопределенности, присущих, вообще говоря, суждениям эксперта. С одной стороны, он не знает с достоверностью, какое подмножество интервалов, включающее базовый, из всей совокупности интервалов, образующих вышеуказанную трапецию, наилучшим образом характеризует неопределенность состояния его знаний о возможном размахе изменения значений рассматриваемого параметра. Вместе с тем все множество интервалов, составляющих названную трапецию, характеризует эту неопределенность, по мнению эксперта, исчерпывающим образом. Тогда меру уверенности эксперта в адекватности описания параметра на базе имеющихся знаний каким-либо подмножеством интервалов, включающим базовый, которое в графическом представлении ПИО оценки имеет высоту h, $0 \le h \le 1$, можно связать с некоторой функцией от этой высоты. С другой стороны, как и в моноинтервальном случае, знаниям эксперта возможно присуща неопределенность, связанная с изменчивостью значений исходного параметра на отдельных интервалах их совокупности.

Таким образом, в нашей теоретико-вероятностной модели для каждого исходного параметра мы имеем дело с системой двух случайных величин $\alpha = h$ и D, заданных на его ПИО. Эта система обладает совместной функцией распределения вероятностей с плотностью $\Psi(\alpha, D)$. Случайная величина α имеет плотность распределения $f_1(\alpha)$, обычным образом связанную с вероятностью $P(\alpha < \alpha_o)$. Эту вероятность будем считать в рамках модели вероятностью того, что при заданных боковых сторонах соответствующей криволинейной трапеции совокупность интервалов, отграниченных интервалами с $h = 0$ и $h = \alpha_o$, адекватно описывает возможную изменчивость длины интервалов, характеризующих вариабельность какого-либо исходного параметра. Отметим, что разные диапазоны изменения α могут быть снабжены разными функциями распределения вероятностей $f_1(\alpha)$. Случайная величина D, лежащая в интервале, отвечающем $h = \alpha$, обладает (условной) плотностью распределения $f_2(D/\alpha)$. Для разных α тип распределения f_2 может быть разным.

ПИО исходного параметра, обладающую известными f_1 и f_2, назовем его обобщенной интервальной оценкой (ОИО).

Следует отметить, что известная неадекватность, для некоторых задач, моноинтервальной оценки с заданием единственной функции распределения вероятностей осознавалась и другими исследователями, предложившими свои модели представления данных.

Так в монографии [Кузнецов, 1991] в рамках теоретико-вероятностной концепции при (моно)интервальном задании данных дополнительно предполагалось, что средние соответствующих распределений имеют интервальную оценку. Это эквивалентно тому, что вероятности наступления анализируемых событий также оказываются заданными интервально. В работах [Аркин и др., 1985; Смоляк, 2001] при (моно) интервальном задании данных их стохастические свойства описывались, в условиях недостатка информации, не однозначно установленной функцией распределения, а целым классом возможных таких функций. При этом степень возможности возникновения каждого из распределений класса характеризуется экспертно задаваемой функцией принадлежности (в нечетко-множественном понимании). Думается, что разнообразие возможных интервально-вероятностных моделей представления данных должно найти отражение в базах моделей СПЭР. Это позволит пользователю привлекать к исследованию те модели, которые в наибольшей степени согласуются с привычным для него языком анализа.

Наша следующая задача состоит теперь в том, чтобы, исходя из ОИО, получить функцию распределения $P(D < D_s)$, заданную на базовом интервале. После того, как такие соотношения для $P(D < D_s)$ установлены, обычным образом можно получить вероятностные оценки гарантированного результата и некоторые другие часто интересующие пользователя показатели, такие как, например, измерители риска.

Функции распределения вероятностей для параметров, описываемых обобщенными интервальными оценками

Пусть $D_{l,r}(\alpha)$ соответственно левые (правые) границы интервалов I_α, отвечающих на криволинейной трапеции ОИО значениям $0 < \alpha < 1$. Тогда интересующая нас вероятность $P(D < D_s)$ имеет вид:

$$P(D < D_S) = \int_0^1 f_1(\alpha) \int_{D_l^{(b)}}^{D_S} f_2\left(D/_{\alpha}\right) dD d\alpha = \int_0^1 f_1(\alpha) P\left(D < D_S/_{\alpha}\right) d\alpha \, . \qquad (1)$$

При этом

$$P\left(D < D_S \middle/ \alpha\right) = \begin{cases} 0, & D_S \le D_l(\alpha), \\ <1, & D_l(\alpha) < D_S < D_r(\alpha), \\ 1, & D_S \ge D_r(\alpha). \end{cases} \tag{2}$$

Ясно, кроме того, что $P(D < D_s) = 0$ при $D_s \le D_l^{(b)}$, $P(D < D_s) = 1$ при $D_s \ge D_r^{(b)}$.

Разделив базовый интервал на три части, $(D_l^{(b)}, D_l^{(m)})$, $(D_l^{(m)}, D_r^{(m)})$ и $(D_r^{(m)}, D_r^{(b)})$, соотношение (1) можно переписать в виде (3), где $\alpha_l(D_s)$ и $\alpha_r(D_s)$ ординаты точек пересечения соответствующих перпендикуляров D_s с боковыми сторонами криволинейной трапеции полиинтервальной оценки на плоскости ($Y = \alpha$, $X = D$).

$$P(D < D_S) =$$
$$= \begin{cases} \int\limits_0^{\alpha_l(D_S)} f_1(\alpha) \int\limits_{D_l(\alpha)}^{D_S} f_2\left(D/\alpha\right) d\alpha dD & \text{при } D_l^{(b)} \le D_s < D_l^{(m)}, \\ \int\limits_0^1 f_1(\alpha) \int\limits_{D_l(\alpha)}^{D_S} f_2\left(D/\alpha\right) d\alpha dD & \text{при } D_l^{(m)} \le D_s \le D_r^{(m)}, \\ \int\limits_0^{\alpha_r(D_S)} f_1(\alpha) \int\limits_{D_l(\alpha)}^{D_S} f_2\left(D/\alpha\right) d\alpha dD + \int\limits_{\alpha_r(D_S)}^1 f_1(\alpha)\, d\alpha & \text{при } D_r^{(m)} < D_s \le D_r^{(b)}. \end{cases} \tag{3}$$

Будем в дальнейшем в этой работе для простоты предполагать, что эксперт задает для случайных величин D, определенных на разных интервалах, один и тот же тип распределения вероятностей. При этом положение наиболее вероятного значения (моды) параметра на разных интервалах будем нормировать на их длину. Это означает, что если, например, эксперт выбрал в качестве закона распределения для D треугольное распределение, то положение моды $D^{(m)}$ для всех интервалов задается им коэффициентом формы $0 < K < 1$.

При этом $D^{(m)}(\alpha) = (1 - K)D_l(\alpha) + KD_r(\alpha)$, где $D_{l,r}(\alpha)$ соответственно левая и правая границы интервала, отвечающего $h = \alpha$. При желании от этого ограничения, как уже отмечалось, можно отказаться. Практическая задача, в которой эксперт выбрал треугольное распределение на базовом интервале, переходящее в равномерное на мини-интервале, рассмотрена, например, в статье одного из авторов (НВЧ) в этом сборнике.

Для частного случая прямолинейных боковых сторон трапеции ОИО имеем

$$\alpha_l(D_S) = \frac{D_S - D_l^{(b)}}{D_l^{(m)} - D_l^{(b)}},\ \alpha_{\text{r}}(D_S) = \frac{D_r^{(b)} - D_S}{D_r^{(b)} - D_r^{(m)}},$$

$$D_{l,\,r}(\alpha) = \alpha D_{l,r}^{(m)} + (1-\alpha) D_{l,r}^{(b)}.$$

В качестве примера применения соотношений (3) для прямолинейных боковых границ ОИО рассмотрим простейший случай равномерных распределений для случайных величин α:

$$f_1(\alpha) = 1 \quad \text{и} \quad D\colon f_2(D/\alpha) = 1/(D_r(\alpha) - D_l(\alpha)).$$

Можно вычислить теперь интегралы в (2) и получить:

При $D_l^{(b)} \le D_s < D_l^{(m)}$:

$$P(D < D_S) =$$
$$= \frac{1}{B-M}\left((D_l^{(m)} - D_l^{(b)})\alpha_l + \frac{B(D_l^{(m)} - D_s) + M(D_s - D_l^{(b)})}{B-M} ln \ \frac{B - \alpha_l (B-M)}{B} \right),$$

при $D_l^{(m)} \le D_s \le D_r^{(m)}$:

$$P(D < D_S) = \frac{1}{B-M}\left(D_l^{(m)} - D_l^{(b)} + \frac{M(D_s - D_l^{(b)}) - B(D_s - D_l^{(m)})}{B-M} ln \frac{M}{B} \right),$$

при $D_r^{(m)} < D_s \le D_r^{(b)}$:

$$P(D < D_S) =$$
$$= \frac{1}{B-M}\left((D_l^{(m)} - D_l^{(b)})\alpha_r + \frac{M(D_s - D_l^{(b)}) - B(D_s - D_l^{(m)})}{B-M} ln \frac{B - \alpha_r (B-M)}{B} \right) + (1-\alpha_r).$$

Для плотности функции распределения тогда получаем:
при $D_l^{(b)} \le D < D_l^{(m)}$:

$$f(D) = \frac{1}{B-M} \ln \frac{B(D_l^{(m)} - D_l^{(b)})}{B(D_l^{(m)} - D) + M(D - D_l^{(b)})},$$

при $D_l^{(m)} \le D \le D_r^{(m)}$:

$$f(D) = \frac{1}{B-M} \ln \frac{B}{M},$$

при $D_r^{(m)} < D \le D_r^{(b)}$:

$$f(D) = \frac{1}{B-M} \ln \frac{B(D_r^{(b)} - D_r^{(m)})}{B(D - D_r^{(m)}) + M(D_r^{(b)} - D)}.$$

Можно показать, что при $M \to B$ это распределение переходит в обычное равномерное распределение: $f(D) = 1/M$. В этом смысле полученные соотношения являются обобщением равномерной функции распределения вероятностей, заданной на моноинтервале, на случай обобщенных интервальных оценок.

Интересно сопоставить поведение функций распределения вероятностей для равномерных распределений, заданных на I_b и I_m и на соответствующей ОИО.

Это сравнение показано на рис. 1 на плоскости ($Y = P(D < D_s)$, $X = D$) для $D_l^{(b)} = 10$, $D_r^{(b)} = 50$, $D_l^{(m)} = 30$, $D_r^{(m)} = 35$. Все три графика имеют одну общую точку пересечения с координатами:

$$D_i = (BD_l^{(m)} - MD_l^{(b)})/(B - M), \quad P_i = (D_l^{(m)} - D_l^{(b)})/(B - M).$$

Можно видеть, что поведение функции распределения для ОИО более сбалансировано, чем для соответствующих кривых моноинтервальных оценок. Именно, ОИО увеличивает вероятность реализации гарантированных исходов для $D < D_i$ и уменьшает эту вероятность для $D > D_i$. Это означает, что ОИО устраняют недооценку малых значений исходных параметров, с одной стороны, и переоценку больших значений этих параметров, с другой.

Аналитические соотношения для $P(D < D_s)$ могут быть получены также для случаев, когда эксперт выбрал для распределения D на интервалах их системы треугольный закон, а для α-распределения по-прежнему принято равномерное распределение. Аналитические формулы получаются также для различных комбинаций (по D и α) треугольного и равномерного распределений.

В случае отказа от задания распределения вероятностей для D на отдельных интервалах и ограничения равномерным распределением для их совокупности получающуюся обобщенную интервальную оценку можно трактовать с позиций теории нечеткости. На факт возможности рассматривать совокупность α-уровней нечеткого множества как обобщенное интервальное число указано ранее в работе [Бессарабов, 1983]. Случай введения равномерного вероятностного распределения на α-уровнях нечеткого числа с треугольной функцией принадлежности рассмотрен в работе [Недосекин, 2000]. Полученные там соотношения следуют из приведенных выше при $M \to 0$, $D_l^{(m)} = D_r^{(m)} = D^{(m)}$.

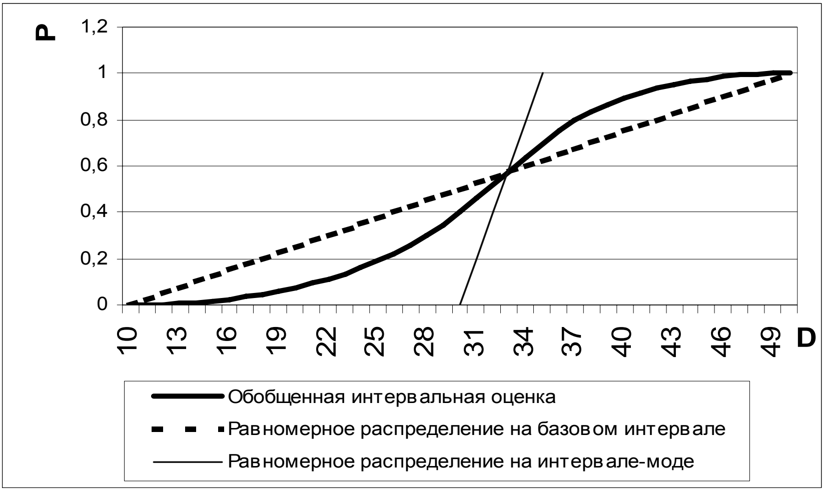

Рис. 1. Функции распределения вероятностей для моноинтервальных и обобщенной интервальной оценок

Более удобными, чем (3), для анализа задач с распределениями исходных параметров произвольного вида являются соотношения для плотностей распределений. Они имеют вид (4):

$$f(D) = \begin{cases} \int\limits_0^{\alpha_l(D)} f_1(\alpha) f_2(D|\alpha) d\alpha & \text{при } D \in [D_l^b; D_l^m), \\ \int\limits_0^1 f_1(\alpha) f_2(D|\alpha) d\alpha & \text{при } D \in [D_l^m; D_r^m), \\ \int\limits_0^{\alpha_r(D)} f_1(\alpha) f_2(D|\alpha) d\alpha & \text{при } D \in [D_r^m; D_r^b). \end{cases} \tag{4}$$

Мы показали таким образом, что совокупность распределений ОИО эквивалентна определенной функции распределения, заданной на базовом интервале. Различные подинтервалы базового интервала приобретают при этом различный «вес». Другими словами, на базовом интервале индуцируется некоторая мера, определяющая ОИО на базовом интервале как вероятностную смесь распределений [Айвазян и др., 1989]. При этом ПИО специфицирует на базовом интервале выбранную экспертом систему подинтервалов, а распределение на α меру на нем[4]. Теперь ясно, что, в случае отказа эксперта, по каким-либо причинам, от построения на основе характерных интервалов полиинтервальной оценки путем задания ее границ, ОИО может быть получена стандартным для подобной ситуации способом. Именно, пусть эксперт задал в качестве оценки параметра несколько характерных интервалов (например, пессимистический, оптимистический и наиболее вероятный), а также шансы на их реализацию, сумма которых равна единице. Следует при этом иметь в виду, что для каждого из характерных интервалов эксперт может построить их ОИО. Тогда результирующее распределение на базовом интервале, построенном как интервал с минимальной левой и максимальной правой границами характерных интервалов, совпадает с вероятностной смесью распределений, первоначально заданных на характерных интервалах и продолженных на базовый интервал, как в (2). Такой подход можно использовать в корпоративных информационных системах при агрегировании мнений нескольких экспертов, предложивших различные оценки для одних и тех же исходных данных.

В качестве первого этапа формирования агрегированной оценки может быть при этом рекомендована следующая процедура. Для каждого исходного параметра моделей получают от экспертов их ОИО и рассчитывают (частные) распределения вероятностей на соответствующих базовых интервалах. Агрегированное распределение задают на результирующем базовом интервале, построенном как интервал с минимальной левой и максимальной правой границами базовых интервалов всех экспертов. Значение агрегированного распределения для каждой точки результирующего базового интервала получают как нормированную на число экспертов сумму частных распределений вероятностей, отвечающих рассматриваемой точке результирующего базового интервала. Отметим, что эта процедура не зависит от того, ПИО какого вида пользуется каждый эксперт.

[4] Думается, что аппарат ОИО более удобен для работы с экспертом, чем попытка выявления его знаний в форме вероятностной смеси.

Ясно, что при этом предполагается равная компетентность всех экспертов. Можно, однако, проводить вышеописанное агрегирование частных оценок с учетом «веса» каждого эксперта. Веса устанавливаются либо «суперэкспертом», либо формируются автоматически, если существуют соответствующие решающие правила.

Интерпретация ОИО как вероятностной смеси открывает дополнительные возможности привлечения методов анализа смесей [Айвазян и др., 1989], таких как методы их идентификации и расщепления, для дальнейшего анализа полученной от эксперта информации. Интерес представляет, например, решение задачи согласования «целостного» знания эксперта об анализируемом показателе с его суждениями о нем, основанными на оценках отдельных параметров.

Причиной рассогласования целостной и параметрической оценок могут быть как ошибочные мнения эксперта, так и неосознанное использование экспертом привходящих соображений, не нашедших отражения в используемой модели представления функции исходных параметров, согласно которой рассчитывается анализируемый показатель.

Например, при вероятностном расчете оценки запасов углеводородов объемным методом предполагается независимость параметров, входящих в расчетную модель, однако это предположение может быть нарушено экспертом при попытке получить целостную оценку запасов. Допустим, что опыт эксперта позволяет ему сделать вывод о том, что его целостная оценка гораздо ближе к истине, чем результаты, полученные на расчетной модели. Обнаруженное противоречие может привести к необходимости пересмотра расчетной модели (например, за счет введения ПИО или учета коррелированности параметров). В ситуации, когда эксперт склонен доверять скорее результатам расчетов, ему, вероятно, необходимо пересмотреть свои взгляды на рассматриваемую проблему в целом.

Таким образом, предлагаемый метод позволяет согласовать суждения эксперта и уточнить их в ходе такого согласования, а также уточнить модели, используемые для расчета функций, требующих оценки.

Важный пример, демонстрирующий актуальность наличия в СПЭР метода согласования частных и общих экспертных оценок, доставляет задача прогнозирования коэффициента извлечения углеводородов. С одной стороны, он может быть задан экспертом как один из коэффициентов в объемном методе оценки запасов (целостное восприятие), а с другой, получен в результате решения комплекса геолого-экономических моделей по величине накопленной добычи за весь срок реализации проекта по освоению объекта.

Другой задачей, которая может быть решена в рамках представления ОИО как смеси, является задача оценивания конечного числа компонентов в модели смесей распределений для $f(D)$, полученной по (4). Ее решение дает эксперту-геологу дополнительную информацию о возможной структуре анализируемого природного объекта.

Метод ОИО позволяет учитывать неопределенность в знаниях экспертов о зависимости параметров модели. Указанная неопределенность может быть представлена как «трубка» значений одного параметра при заданных значениях, принимаемых вторым параметром (рис. 2). При этом аппарат ОИО позволяет эксперту естественным образом отразить свои предпочтения для различных участков этой «трубки».

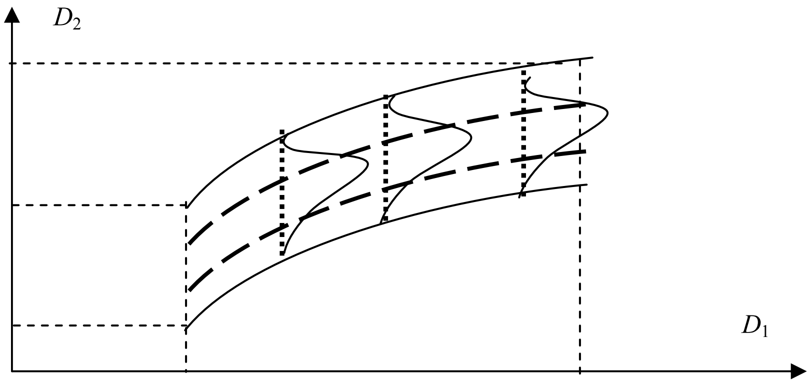

Рис. 2. Учет зависимости параметров модели с помощью ОИО

Для каждого значения параметра D_1 можно построить ОИО, отражающую мнение эксперта о возможных значениях параметра D_2. При этом базовые интервалы задаются сплошными линиями трубки, а мини-интервалы ее пунктирными границами. Плотности распределений для всех промежуточных сечений трубки можно вычислять интерполяцией по значениям характерных сечений.

Частным случаем решения задачи по учету зависимости параметров является учет динамики — зависимости значения анализируемого параметра от времени. Эта задача довольно часто встречается при прогнозных расчетах. Работа с получаемыми таким образом динамическими ОИО аналогична процедурам, рассмотренным выше. Можно предположить, что со временем неопределенность возрастает, приводя к расширению трубки.

Если, по мнению эксперта, ОИО для разных сечений меняются со временем по одному закону, для ряда комбинаций типов распределений по α и по D плотности для базовых интервалов последовательных сечений можно вычислить по явным формулам, либо, если явные формулы не могут быть получены, как и выше, интерполяцией.

Заключение

Метод обобщенных интервальных оценок является новым методом представления экспертных знаний в задачах, исходные данные которых могут быть описаны количественными оценками. Метод предназначен, прежде всего, для учета неопределенности в (числовой) исходной информации моделей предметных областей в базе моделей СПЭР. Он может быть использован при агрегировании интервальных оценок, заданных разными экспертами, и в задачах с зависимыми исходными данными.

Помимо задач оценки ожидаемой перспективности природных объектов, потенциально содержащих углеводороды, аналогичный подход может быть применен и для анализа других задач. В классе задач оценки перспективности природных объектов отметим задачу определения рациональной стратегии пополнения фондов запасов (ресурсов) различной степени изученности, обеспечивающей выполнение заданной динамики добычи; задачу теоретико-вероятностного суммирования запасов различных блоков одного объекта или различных объектов; задачу рационального выбора стратегии освоения группы объектов, ориентированных на один и тот же магистральный трубопровод.

Литература

[Айвазян и др., 1989] *Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989.

[Аркин и др., 1985] *Аркин В. И., Смоляк С. А.* О структуре критериев оптимальности в условиях неопределенности // Вероятностные проблемы управления и математическая экономика. М.: ЦЭМИ РАН, 1985.

[Бессарабов, 1983] *Бессарабов Н. В.* Множества уровня как обобщенные интервальные числа // Управление при наличии расплывчатых категорий: Тезисы VI научного семинара. Пермь: НИИУМС, 1983.

[Золотухин, 2001] *Золотухин А. Б.* Начальные и извлекаемые запасы нефти и газа // Процесс принятия управленческих решений на основе экономического анализа работ по поискам и разведке нефти и газа. М.: ВНИИОЭНГ, 2001.

[Кузнецов, 1991] *Кузнецов В. П.* Интервальные статистические модели. М.: Радио и связь, 1991.

[Ларичев, 1987] *Ларичев О. И.* Объективные модели и субъективные решения. М.: Наука, 1987.

[Ларичев и др., 1987а] *Ларичев О. И., Мошкович Е. М.* Качественные методы принятия решений. М.: Наука, 1987.

[Недосекин, 2000] *Недосекин А. О.* Применение теории нечетких множеств к задачам управления финансами // Аудит и финансовый анализ. 2000. № 2.

[Орлов, 1996] *Орлов А. И.* Экспертные оценки // Заводская лаборатория. 1996. Т. 62. № 1.

[Петровский, 1996] *Петровский А. Б.* Компьютерная поддержка принятия решений: современное состояние и перспективы развития // Системные исследования. Методологические проблемы: Ежегодник. М.: УРСС, 1996.

[Петровский, 2003] *Петровский А. Б.* Пространства множеств и мультимножеств. М.: УРСС, 2003.

[Пороскун и др., 1999] *Пороскун В. И., Стернин М. Ю., Шепелев Г. И.* Вероятностная оценка запасов на начальных стадиях изучения залежей нефти и газа // Геология нефти и газа. 1999. № 5–6.

[Пороскун и др., 2002] *Пороскун В. И., Стернин М. Ю., Шепелев Г. И.* Экспертные модели прогнозных оценок подготовки и освоения запасов нефти // Искусственный интеллект. 2002. № 2.

[Смоляк, 2001] *Смоляк С. А.* Оценка эффективности проектов в условиях нечеткой вероятностной неопределенности // Экономика и математические методы. 2001. Т. 37. № 1.

[Стернин и др., 2003] *Стернин М. Ю., Шепелев Г. И.* Метод представления знаний в интеллектуальных системах поддержки экспертных решений // Новости искусственного интеллекта. 2003. № 4.

[P. Artzner, 1997] *Artzner P., Delbaen F., Eber J.-M., Heath D.* Definition of Coherent Measures of Risk / Symposium on Risk Management at the European Finance Association 24th Annual Meeting. Viena, Austria, 1997.

[Kahneman et al., 1982] *Kahneman D., Slovic P., Tversky A* (eds). Judgment Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press, 1982.

[Shepelyov et al., 2003] *Shepelyov G., Sternin M.* Method of generalized interval estimations for intelligent DSS // Proceedings o international conference «DSS in the uncertainty of the Internet age». Katowice: University of economics, 2003.