

ОБНАРУЖЕНИЕ ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ (Вычислительные системы)

1999 год

Выпуск 166

УДК 681.142

МЕТРОЛОГИЧЕСКИЕ СВОЙСТВА ЭКСПЕРТА¹

Н.Г. Загоруйко

В в е д е н и е

Человек часто используется в качестве средства измерений, особенно при изучении слабо формализованных характеристик объектов или явлений. Типичным примером может служить использование экспертных оценок в исследованиях гуманитарного характера.

Современная теория измерений фиксирует группу допустимых преобразований протоколов, полученных с помощью приборов с тсм или иным типом измерительной шкалы. Это обеспечивает сохранение эмпирического содержания протоколов при переходе с одного языка на другой эквивалентный ему язык и позволяет применять адекватные математические методы анализа экспериментальных данных. К сожалению, теория измерений традиционно ориентирована, в основном, на применение в физико-технической сфере и по этой причине слабо влияет на методологический уровень проведения эмпирических исследований в социологии, психологии и других гуманитарных областях.

Способности человека измерять отдельные характеристики объектов, сравнивать объекты по комплексу свойств, оценивать ситуации и на этой основе принимать решения исследуются с разными целями. Знание психофизиологических механизмов, используемых человеком в процессе выработки экспертных оценок,

¹Работа выполнена при поддержке гранта РФФИ № 97-06-80312 и Гранта МинВУЗа РФ № ЗН-228-98.

позволяет создавать аппаратно-программные измерительные системы, моделирующие эти механизмы. В этом русле лежат работы по моделированию отдельных органов чувств человека — зрения, слуха и т.п., а также более сложных человекоподобных технических систем (распознавания образов, принятия решений и пр.).

Во многих случаях мы не знаем, как устроена человеческая измерительная система и как она функционирует. Это касается измерения сложных явлений, таких, например, как упорядочение объектов по привлекательности или сравнения абстрактных понятий по сложности. Вряд ли можно рассчитывать на скорое появление моделей человеческих механизмов таких измерений. В этом случае нам необходимо установить закономерную связь между сигналами на ее входе (X) и выходе (Y), чтобы по выходной реакции Y судить о входном воздействии X , что позволит использовать систему в качестве измерительного средства. При этом выходную реакцию человека можно фиксировать двумя путями. Один из них состоит в наблюдении за показаниями приборов, измеряющих физиологические характеристики человека до и после входного воздействия на него. Такой способ использования человека в качестве физиологического измерительного средства предложен в [1]. Второй, более распространенный способ заключается в том, что человек, воспринимая внешнее воздействие, осмысливает его и формулирует оценки своих ощущений. Именно в таком режиме интеллектуального прибора используется человек-эксперт, участвующий в измерении неформализованных характеристик.

Но, как и всякий измерительный прибор, эксперт требует тестирования: нужно знать такие его метрологические свойства, как пропускная и разрешающая способности, объективность, компетентность, стабильность показаний во времени, помехоустойчивость, наличие случайных и систематических погрешностей, типы используемых измерительных шкал и т.д. Эти свойства экспертов пока изучены слабо. Следует отметить, что большую трудность представляет собой разработка методов тестирования экспертов, особенно если это касается тестирования его трудно формализуемых характеристик, в чем можно убедиться на при-

мере методики тестирования такой характеристики, как "конформизм" [2]. Данная работа содержит результаты исследований некоторых других метрологических характеристик экспертов.

§1. Погрешности и стабильности оценок во времени

Погрешностью в метрологии называется разность между результатом измерения и истинным значением измеряемой величины. При измерении любых характеристик, а тем более, гуманитарных, мы не можем знать их истинных значений. Наличие случайных погрешностей можно обнаружить при многократном измерении одной и той же предположительно не меняющейся во времени характеристики.

В нашем эксперименте группе из 14 экспертов ($n = 14$), хорошо знающих друг друга, предлагалось оценить некоторую неформализованную характеристику, например, деловую квалификацию или чувство юмора каждого из членов этой группы. Диапазон возможных значений характеристики задавался в пределах от 0 до 100. Каждый эксперт получал ($n - 1$) листов бумаги с фамилиями оцениваемых людей и, не советуясь с другими участниками эксперимента, выставлял им свои оценки. Нет ничего удивительного в том, что оценки, выставленные разными экспертами одному и тому же человеку, заметно отличались друг от друга (отклонения от среднего значения колебались в пределах от 10 до 40%).

Через 20 дней эксперимент был повторен. Эксперты не знали общих результатов первого эксперимента и снова выставляли свои оценки независимо друг от друга. Сравнение прежних и новых результатов показало, что средние оценки и их дисперсии изменились мало, но индивидуальные оценки, выставленные одному и тому же человеку одним и тем же экспертом в первом и втором эксперименте, колебались в пределах от 7 до 20%.

Следующая серия из двух аналогичных экспериментов отличалась от первой тем, что экспертам были созданы условия, позволявшие им видеть все свои оценки одновременно. Для этого было изготовлено простое приспособление в виде n параллельно расположенных стержней с кольцом, движущимся вдоль каждого стержня. Под стержнями помещалась бумага с равномерно нане-

сенными на нее метками в диапазоне от 0 до 100. Каждый стержень был помечен фамилией оцениваемого человека. Эксперт выставлял оценки каждому человеку, устанавливая кольцо над соответствующим числом. После того, как эксперт выставил все оценки, он получал общую картину взаимного расположения колец на всех стержнях. Это давало ему возможность видеть множественные отношения между оценками и вносить поправки в первоначальный вариант. После нескольких итераций эксперт приходил к заключению, что теперь оценки хорошо соответствуют его представлению о значении оцениваемой характеристики у данной группы людей, и этот результат фиксировался в протоколе.

Повторный эксперимент проводился через месяц после первого. Было обнаружено, что эксперты воспроизводили свои прежние оценки с ошибкой, не превышавшей двух единиц. Этот эксперимент позволил сделать следующий важный вывод: случайная погрешность эксперта при умозрительном оценивании неформализованных характеристик может достигать +20%. Но если дать ему возможность подключить к процессу принятия решения зрительный анализатор, то восприятие множественных отношений снижает эту погрешность на порядок. Это означает, что таким путем повышается разрешающая способность эксперта, как измерительного прибора. При этом стабильность во времени оценок хорошо известных эксперту характеристик достаточно высока.

Простую модель устройства ("визуализатора отношений"), позволяющего визуализировать множественные отношения между оцениваемыми объектами, можно сделать на экране дисплея. Отобразим на нем протокол испытания и позволим эксперту двигать вдоль горизонтальных линий некоторый символ, устанавливая его в любом месте шкалы, как это показано на рис.1.

	0	1	2	3	4	...	100
Иванов*
Петров*
Сидоров*
.....
Федоров*

Рис. 1. Модель визуализатора отношений

§2. Типы используемых измерительных шкал

В опытах с экспертами было обнаружено, что они могут работать со шкалами любого типа, но чем выше информативность шкалы, тем труднее эксперту с ней работать. В одном из опытов, в котором участвовало 28 экспертов, требовалось оценить в шкале порядка и в шкале отношений некоторое неформализованное свойство 10 объектов. При оценивании в шкале порядка эксперты приписывали объектам порядковые места от 1 до 10. Оценка в шкале отношений могла находиться в пределах от +100 до -100. Эксперты без особых затруднений и охотно упорядочивали объекты по измеряемому свойству, в то время как его оценивание в шкале отношений (в процентах) делалось ими с трудом и неохотно.

Индивидуальные экспертные оценки, выставленные объекту, усреднялись по каждому типу шкал отдельно. В таблице показаны средние значения измеряемой характеристики в шкале отношений x_1 и в шкале порядка x_2 . Оказалось, что зависимость между этими средними значениями очень близка к линейной (см. рис. 2).

Результаты этого эксперимента позволяют сделать предположение о возможности упростить задачу группового экспертного оценивания неформализованных характеристик. Процедура при этом должна иметь два этапа. На первом этапе экспертам предлагается оценить характеристику в шкале порядка. По результатам усреднения порядковых мест выбирается два объекта: с самым малым средним порядковым местом и с самым большим. На втором этапе экспертам предлагается оценить ту же характеристику двух выбранных объектов, но в более сильной шкале отношений. После этого оценки в сильной шкале для остальных объектов можно получить по графику, аналогичному приведенному на рис. 2.

Т а б л и ц а

x_1	66.5	62	58	14	6	5	-1	-3	-42	-58
x_2	3.08	2.85	3.02	5.54	6.02	5.86	5.97	6.30	7.93	9.03

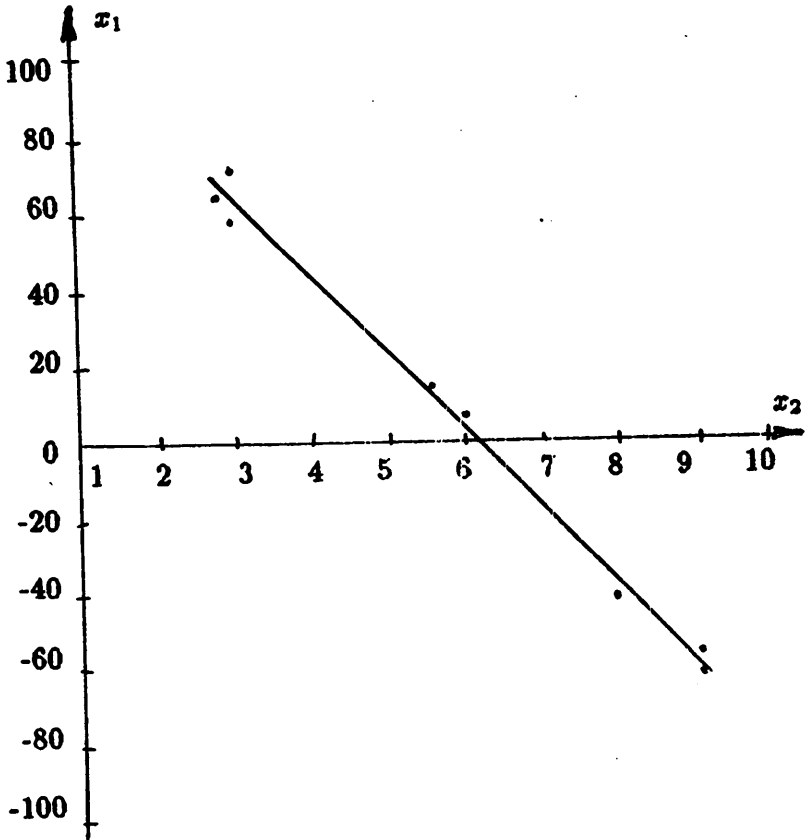


Рис. 2.

§3. Случайные погрешности и компетентность экспертов

Если есть основания считать, что ошибки экспертных измерений носят случайный характер, то их влияние можно уменьшить путем усреднения индивидуальных оценок. Если ошибки распределяются по нормальному закону и отклонение индивидуальной оценки x_i i -го эксперта от среднего значения X' равно u_i , то среднее квадратичное отклонение S среднего значения X' от истинного X с ростом числа экспертов n будет умень-

$$S = \sqrt{\frac{\sum_{i=1}^n u_i^2}{n(n-1)}}.$$

В описанных выше экспериментах средние значения стабилизировались уже при усреднении 10–15 индивидуальных оценок. Средние значения оценок 28 экспертов отличались от средних значений 15 экспертов на величину, не превышавшую +5%.

Величина погрешности индивидуальных оценок существенно зависит от компетентности эксперта. В работе [3] описана методика автоматической оценки компетентности эксперта. Там же исследовался вопрос о том, как влияет на средние оценки учет компетентности экспертов. Выяснилось, что при небольшом числе экспертов их индивидуальная компетентность влияет на средний результат существенно. При увеличении числа экспертов n это влияние ослабевает. Так при $n > 50$ максимальная разница средних значений, полученных с учетом и без учета компетентности экспертов, не превышала +10%.

§4. Метрика пространства восприятия

Эксперту предъявляется группа из m объектов, каждый из которых описан набором из k характеристик, и предлагается разделить их на три класса по комплексной характеристике качества: S_1 ("хорошие"), S_2 ("средние") и S_3 ("плохие"). Пусть для начала все k свойств измерены в сильных ("количественных") шкалах. В этом k -мерном метрическом пространстве эксперт сравнивает между собой характеристики объектов и создает свое представление о качестве объектов, т.е. синтезирует новую (целую) характеристику, измеряемую в шкале порядка. При разделении объектов на три класса по качеству он строит грани-

цы (решающие функции), которые делят k -мерное пространство описывающих свойств на 3 непересекающиеся области.

Если бы нам удалось узнать уравнения этих решающих функций, то мы могли бы построить модель эксперта, оценивающего качество объектов данного вида. В дальнейшем для оценки качества некоторого нового объекта не было бы необходимости снова пригласить эксперта. Зная описывающие характеристики объекта, можно по уравнениям решающих функций автоматически определить, к какому классу по качеству нужно отнести этот объект.

Эксперт обычно не может объяснить механизм принятия своих решений. Построить модель этого механизма можно с помощью методов распознавания образов. Предположим, что эксперт использует самый простой вид разделяющих поверхностей — k -мерные гиперплоскости. Зафиксируем в протоколе результаты серии экспериментов, которые состоят в том, что эксперту предъявляется вектор k описывающих характеристик объекта, а он называет номер его класса качества. Этот протокол играет роль обучающей выборки, использование которой позволяет построить линейные решающие функции, описываемые уравнениями вида $Y = \sum_{i=1}^k a_i * x_i + a_0$.

Чем больше коэффициент a_i , тем сильнее i -я характеристика влияет на оценку качества объекта. Это дает возможность понять, как оценивает эксперт относительную значимость разных характеристик при различении объектов двух любых классов.

Реальные эксперименты показали, что в пространстве даже небольшой размерности (при $k > 2$) человек обычно не может оперировать гиперплоскостями, наклоненными к координатным осям [4]. В этом случае он рассматривает характеристики в качестве независимых и переходит к использованию логических решающих правил [5] типа:

Если $(x_1 > 5) \wedge (x_2 = 3 - 6) \wedge \dots \wedge (x_k = 10)$, то $(Y = S_1)$.

Здесь границы между классами представляют собой набор плоскостей, перпендикулярных координатным осям.

Такие решающие правила позволяют работать и в разнотипном пространстве. Необходимость в этом возникает тогда, когда

эксперту приходится сравнивать между собой характеристики различной физической природы, например, вес, цвет, твердость, скорость, внешний вид и т.п. Решающие правила при этом будут иметь такой вид:

Если $(x_1 < 2) \wedge (x_2 = \text{синий}) \wedge (x_3 = \text{тверже стали}) \wedge \dots \wedge (x_k = 0)$, то $(Y = S_3)$.

Логические решающие правила удовлетворяют требованиям, предъявляемым к математическому аппарату, используемому в области Data Mining [6]: ими легко пользоваться как машине, так и человеку. Модель эксперта в виде используемых им логических решающих правил позволяет понять особенности пространства его восприятия.

З а к л ю ч е н и е

1. Если эксперт оценивает большое число объектов чисто умозрительно, то различающая способность эксперта, т.е. способность надежно различать две близкие ситуации, и стабильность его показаний во времени имеют низкие значения. Если же создать условия, при которых эксперт в процессе измерений может воспользоваться своими органами чувств (например, зрением), то эти характеристики становятся вполне приемлемыми для многих гуманитарных измерений.

2. Эксперту трудно работать в режиме измерительного прибора с сильными шкалами. Он охотнее и увереннее оценивает свойства объектов в порядковой и номинальной шкалах. Имеется возможность получения оценок в сильных шкалах по результатам измерений в шкале порядка, что существенно облегчает проведение групповых экспертных оценок.

3. Для получения результата с погрешностью, приемлемой для гуманитарных целей, достаточно привлечь к оцениванию сравнительно небольшое число экспертов (30–50).

4. При небольшом числе экспертов учет компетентности экспертов играет существенную роль. С ростом числа экспертов важность учета индивидуальной компетентности быстро падает.

5. Методами распознавания образов можно построить модель пространства восприятия эксперта, которая позволяет не только имитировать процессы синтеза целевых характеристик, но и

понять механизм соизмерения разнотипных и неравнозначных характеристик.

Литература:

1. КОСАРЕВ Ю.Г. Человек как универсальный измерительный прибор (в порядке постановки проблемы) // Искусственный интеллект и экспертные системы. — Новосибирск, 1997. — Вып. 160: Вычислительные системы. — С. 36–42.
2. КАРГАЛЬЦЕВА С.Л., САМОХВАЛОВ К.Ф. Как измерить склонность к конформизму? // Искусственный интеллект и экспертные системы. — Новосибирск, 1997. — Вып. 160: Вычислительные системы. — С. 18–35.
3. ЗАГОРУЙКО Н.Г. Доверие к информации и ее источнику в экспертной системе // Экспертные системы и распознавание образов. Новосибирск, 1988. — Вып. 126: Вычислительные системы. — С. 3–23.
4. ЗАГОРУЙКО Н.Г. Какими решающими функциями пользуется человек? // — Новосибирск, 1987. — Вычислительные системы. Вып.28. — С. 69–78.
5. ЛБОВ Г.С. Методы обработки разнотипных экспериментальных данных. — Новосибирск: Наука, 1981.
6. ЗАГОРУЙКО Н.Г. Прикладные методы анализа данных и знаний. — Новосибирск, 1999. (Изд. ИМ СО РАН).

Поступила в редакцию
23 ноября 1999 года