

УДК 519.237.7

## Об интерпретации латентных переменных в задачах многомерного шкалирования

**Михно Галина Алексеевна** – кандидат технических наук, доцент кафедры Вычислительной математики Тверского государственного университета (ТвГУ, г. Тверь)

*Аннотация:* Предлагаются процедуры реализации основных этапов процесса решения задачи интерпретации латентных переменных. В основу процедур положены экспертизы формирования множества факторов-интерпретаторов и ранжирования анализируемых объектов, а также критерий ранговой корреляции латентных переменных и факторов-интерпретаторов.

*Ключевые слова:* Многомерное шкалирование, латентная переменная, интерпретация, нестрогая ранжировка, коэффициент ранговой корреляции.

### Введение

Методы многомерного шкалирования (МШ) составляют одно из направлений прикладной статистики [1] и находят широкое применение в технических и социальных науках, в экономике, психологии [2, 3]. Одной из основных задач МШ является смысловая (содержательная) интерпретация латентных (т.е. скрытых, непосредственно не наблюдаемых) переменных, объясняющих попарные сходства, связи или отношения множества анализируемых объектов (см. [2]). Известные методы решения указанной задачи условно можно разделить на три группы по степени влияния в них субъективного мнения на конечный результат: 1 – чисто субъективные методы; 2 – методы завершающей субъективной интерпретации; 3 – методы, основанные на формальных критериях.

К первой группе относятся полностью субъективные методы отождествления каждой латентной переменной с ее интерпретатором. Здесь субъект (исследователь, эксперт,

испытываемый) определяет множество факторов, которые, по его мнению, обуславливают исходные попарные меры близостей объектов. Число таких факторов (интерпретаторов) совпадает с числом латентных переменных. Далее на основе анализа геометрической конфигурации множества анализируемых объектов в пространстве латентных переменных и сущности каждого из определенных гипотетических интерпретаторов субъект либо сразу, либо после предварительного “ручного” преобразования конфигурации осуществляет отождествление латентной переменной с ее интерпретатором. При этом преобразование конфигурации заключается в таком повороте координатных осей пространства, которое подтверждает предполагаемую интерпретацию.

Ко второй группе относятся методы, в которых предварительное интерпретируемое преобразование геометрической конфигурации множества объектов в пространстве латентных переменных производится “объективно”, т.е. с использованием формальных критериев. Полагается, что в результате “объективного” поворота может быть получена простая конфигурация объектов, обеспечивающая возможность субъективного определения интерпретаторов и их субъективного отождествления с соответствующими латентными переменными.

Третью группу составляют методы, в которых субъективная составляющая относится к этапу формирования множества гипотетических интерпретаторов и оценивания их численных значений для каждого анализируемого объекта. Причем, число таких интерпретаторов превосходит или равно числу латентных переменных. Задача собственно отождествления интерпретаторов и латентных переменных решается здесь с использованием формальных критериев и решающих статистик.

Первые две группы методов практически применимы лишь для случая размерности пространства латентных переменных, не более, чем два. Кроме того, существенный недостаток первой и второй групп методов состоит в необходимости субъективного эвристического отождествления (по сути угадывания) смыслового фактора, отражаемого каждой из латентных переменных. Последнее не всегда возможно и достоверность интерпретации в рамках данных методов не проверяется. Основным недостатком третьей группы методов является необходимость численного оценивания значений гипотетических интерпретаторов для анализируемых объектов. Известно [4], что задача предоставления требуемой оценки в рассматриваемых методах относится к классу задач большой сложности для субъектов, что может привести к недопустимым (в смысле результатов интерпретации) погрешностям. Согласно тем же исследованиям (см. [4]) наиболее простой задачей для субъекта является предоставление вербальных (качественных) суждений относительно объектов, например, лучше, хуже, равноценно и т.п. для пар объектов.

Изложенные обстоятельства определяют актуальность задачи совершенствования методов интерпретации латентных переменных с целью повышения достоверности получаемой от субъектов информации и снижения влияния субъективных суждений на результаты интерпретации.

В статье дается обоснование и конкретизируются процедуры реализации составных этапов решения задачи интерпретации латентных переменных, обеспечивающие (процедуры) достижение сформулированных целей.

## 1. Постановка задачи

Пусть в результате использования произвольной (метрической или не метрической) модели МШ решены задача поиска латентных переменных и оценки координат каждого анализируемого объекта. Обозначим через  $m$  число объектов, через  $K$  число найденных латентных переменных, т.е. размерность пространства латентных переменных, через  $L=\{1, \dots, K\}$  множество номеров латентных переменных, через  $X = [x_{ij}]$ ,  $i=1, \dots, m; j=1, \dots, K$  матрицу оценок координат объектов в пространстве латентных переменных. Здесь  $x_{ij}$

$x_{ij}$  – оценка  $j$ -й координаты  $i$ -го объекта.

Перечисленная информация определяет исходные данные для решения задачи интерпретации латентных переменных, которая заключается в необходимости найти (правильное, соответствующее истине) смысловое содержание каждой из выявленных переменных. Постановка и решение данной задачи включает постановку и решение следующих подзадач:

1) формирование полного множества гипотетических интерпретаторов для множества выявленных латентных переменных;

2) конкретизация вида и вербального способа выявления информации, характеризующей рассматриваемые объекты посредством предполагаемых

интерпретаторов;

3) конкретизация формального критерия отождествления каждой латентной переменной с ее интерпретатором.

## 2. Метод решения задачи интерпретации латентных переменных

### 2.1. Формирование множества гипотетических интерпретаторов.

В основу решения подзадачи формирования исходного гипотетического множества смысловых факторов (интерпретаторов) для латентных переменных из множества  $L$  положим процедуру экспертного перечисления [5]. Сущность соответствующего данной процедуре подхода состоит в следующем. Пусть к решению рассматриваемой подзадачи привлекается группа из  $N$  экспертов. Каждый из экспертов перечисляет факторы, имеющие смысловое содержание и влияющие, по его мнению, на попарную близость анализируемых объектов. Число называемых каждым экспертом факторов не обязательно равно числу латентных переменных. В общем случае оно может быть как больше, так и меньше числа латентных переменных. Обозначим через  $Z_t, t=1, \dots, N$  множество факторов, указанных  $t$ -м экспертом. Введем суммарное экспертное множество  $Z = \bigcup_{t=1}^N Z_t$

смысловых факторов. Число факторов, вошедших в  $Z$ , обозначим через  $n$ . Далее строится матрица  $R = \|r_{tj}\|$

$r_{tj} = \begin{cases} 1, & \text{если фактор } j \text{ принадлежит } Z_t, \\ 0 & \text{в противном случае.} \end{cases}$ ,  $t=1, \dots, N; j=1, \dots, n$ , где

$$r_{tj} = \begin{cases} 1, & \text{если фактор } j \text{ принадлежит } Z_t, \\ 0 & \text{в противном случае.} \end{cases}$$

Вычисляются оценки вероятностей принадлежности факторов из множества  $Z$  формируемому исходному гипотетическому множеству смысловых факторов по формуле

$P_j = 1/N \sum_{t=1}^N r_{tj}$ ,  $j=1, \dots, n$ . Искомое гипотетическое множество  $Y = \{y_1, \dots, y_T\}$  смысловых факторов-интерпретаторов формируется путем включения в него факторов из  $Z$ , для которых  $P_j \geq P$ .

Здесь  $P$  – задаваемая исследователем пороговая вероятность включения факторов в искомое множество интерпретаторов. При решении рассматриваемой подзадачи требуется выполнение условия  $T \geq K$ . Последнее (при необходимости) может быть обеспечено настройкой параметров приведенного подхода, например изменением величины  $P$  или требованием к экспертам на включение в множество  $Z$

$t$   
числа факторов не меньшего, чем  $K$ .

## 2.2. Характеристики объектов посредством интерпретаторов.

С учетом указанной во введении цели ориентации на вербальную информацию будем характеризовать анализируемые объекты посредством их (нестрогой) упорядочения по степени проявления смыслового фактора. В получаемой ранжировке рассматриваемых объектов по конкретному интерпретатору первое место присваивается объекту с наиболее присущим по сравнению с другими объектами проявлением данного фактора и далее объекты располагаются в порядке уменьшения проявления фактора. Если для нескольких объектов проявление фактора одинаково, то им присваивается одинаковый ранг (место), равное среднему арифметическому мест, приходящихся на эти объекты. В последнем случае ранги (места) объектов могут быть дробными. Например, если двум объектам присуще одинаковое проявление фактора и перед ними уже располагаются четыре объекта, то каждый из этих двух объектов получит ранг (место), равный 5.5, т.к. на них приходятся места пятое и шестое.

Наиболее простым (в смысле требуемой от экспертов информации) способом получения рассматриваемых нестрогих ранжировок объектов по каждому из интерпретаторов является экспертиза, основанная на методе парных сравнений объектов [6]. Сущность экспертизы состоит в следующем. Каждый из экспертов производит  $C_m^2$  сравнений пар объектов (каждый объект сравнивается с остальными). В результате формируются матрицы  $A^{(t)}$

$t$   
 $a_{ij}^{(t)}$ ,  $i=1, \dots, m; j=1, \dots, m; t=1, \dots, N$ . Здесь  $a$

$t$   
 $a_{ij} \in \{0, 1\}$ ;  $a$

$t$   
 $a_{ij}$

$a_{ij} = 1, i \neq j$ , если по мнению  $t$ -го эксперта  $i$ -му объекту рассматриваемый фактор присущ в большей степени, чем  $j$ -му объекту;  $a_{ij} = 0, i = j$ ;  $a_{ij} = a_{ji}$

$t$

$i, j$

$+ a_{ij}$

$t$

$j, i$

$= 1$  при  $i \neq j$ ;  $a_{ij} = 0, i = j$

$t$

$i, j$

$= 0, i = 1, \dots, m$ . Далее вычисляются суммарная матрица  $A = \sum_{t=1}^T a_{ij}^{(t)}$

$i, j$

$\square = \sum_{t=1}^T a_{ij}^{(t)}$

$t$

$a_{ij}^{(t)}$

$t$

$i, j$

и величины  $a_{ij}^{(t)}$

$s$

$= \sum_{m=1}^m$

$m$

$i=1$

$a_{ij}^{(s)}$

$i, s$

,  $s = 1, \dots, m$ . Результирующая ранжировка объектов по конкретному интерпретатору определяется в соответствии с величинами  $a_{ij}^{(s)}$

$s$

, а именно в порядке возрастания данных величин. При равенстве оценок  $a_{ij}^{(s)}$

$s$

для нескольких объектов они получают средний арифметический ранг (место), как указывалось выше. Величины  $a_{ij}^{(s)}$

$s$

отражают суммарное число случаев, когда эксперты считали, что объекту  $s$  фактор был присущ в меньшей степени, чем другим объектам. Нетрудно видеть, что получаемая ранжировка отражает учет степени проявления фактора-интерпретатора в объектах.

В результате реализации рассмотренной экспертизы получим  $T$  ранжировок (по числу факторов-интерпретаторов) множества анализируемых объектов. Введем для них обозначение  $\hat{y}_s = (y_{s1}, \dots, y_{sm})$ , где  $y_{sj}$  – ранг  $j$ -го объекта в упорядочении объектов по  $s$ -му фактору-интерпретатору. Пусть далее  $\hat{Y} = \{\hat{y}_1, \dots, \hat{y}_T\}$  – множество полученных ранжировок, а  $I = \{1, \dots, T\}$  – множество номеров данных ранжировок.

### 2.3. Критерий интерпретации латентных переменных.

Правомочно положить, что значения координат объектов в пространстве латентных переменных, полученные в результате применения модели МШ (см. п.1), отражают степень влияния (важности) латентных переменных на анализируемые объекты в смысле изучаемого с помощью МШ явления. По аналогии с процедурой из п.2.2 каждый вектор-столбец матрицы X координат может быть использован для определения ранжировок объектов по каждой латентной переменной. В определяемом упорядочении по конкретной латентной переменной на первое место ставится объект с наибольшим значением координаты и далее объекты располагаются по убыванию оценок данной латентной переменной. Для объектов с одинаковым значением координат ранг определяется как среднее арифметическое соответствующих мест (см. п.2.2). Обозначим получаемые таким образом ранжировки через  $\hat{x}_j = (\hat{x}_{j1}, \dots, \hat{x}_{jm})$ ,  $j=1, \dots, K$ , где  $\hat{x}_{jq}$  – ранг q-го объекта в ранжировке объектов по j-й латентной переменной.

Теперь в качестве решающей статистики в задаче отождествления конкретной латентной переменной с конкретным фактором-интерпретатором, т.е. в задаче интерпретации латентных переменных, будем использовать коэффициент ранговой корреляции Кендалла [7], который для ранжировок  $\hat{x}_j = (\hat{x}_{j1}, \dots, \hat{x}_{jm})$  и  $\hat{y}_s = (y_{s1}, \dots, y_{sm})$ ,  $j \in L, s \in I$  вводится следующим образом. Для ранжировки  $\hat{x}_j$

положим

$$c_{pq}^{(j)} = \begin{cases} 0, & \text{если } \hat{x}_{jp} = \hat{x}_{jq}, \\ 1, & \text{если } \hat{x}_{jp} < \hat{x}_{jq}, \\ -1, & \text{если } \hat{x}_{jp} > \hat{x}_{jq}. \end{cases}$$

Аналогично определим величины  $d^{(s)pq}$  для ранжировки  $\hat{y}_s$ . Тогда коэффициент ранговой корреляции Кендалла рассматриваемых ранжировок задается величиной

$$\tau_{js} = \frac{\sum_{p=1}^m \sum_{q=1}^m c_{pq}^{(j)} d^{(s)pq}}{[m(m-1)/2]}.$$

Величины  $\tau_{js}, j \in L, s \in I$  определяют оценку связи между соответствующими ранжировками и, следовательно, могут использоваться для

интерпретации латентных переменных. Полагаем, что смысловой фактор  $s_1$  является интерпретатором латентной переменной  $j$   $\tau_{j1} = \max_{s_1} \tau_{js_1}$ .

Далее считаем, что смысловой фактор  $s_2$

2

$\tau_{j2}$  является интерпретатором латентной переменной  $j$

2

$\tau_{j2}$ , если  $\tau_{j2} > \tau_{j1}$

$\tau_{j2} > \tau_{j1}$

$\tau_{j2} > \tau_{j1}$

$\tau_{j2} > \tau_{j1}$

$\tau_{j2}$

$\tau_{j2}$

. Последовательно применяя приведенную процедуру получим в результате множество пар  $\{(j, s_1), (j, s_2), \dots, (j, s_k)\}$ , каждая из которых определяет латентную переменную и интерпретирующий ее смысловой фактор.

1

,  $s_1$

1

), ( $j, s_1$ )

2

,  $s_2$

2

), ..., ( $j, s_k$ )

к

,  $s_k$

к

)), каждая из которых определяет латентную переменную и интерпретирующий ее смысловой фактор.

## Заключение

Предложенный подход на этапах участия субъектов (экспертов или испытуемых) в процессе решения задачи интерпретации латентных переменных ориентирован на получение качественной информации, которую субъекты способны предоставить, как правило, без особых затруднений. Этап же непосредственной интерпретации переменных реализуется объективно с использованием формального критерия связи латентной переменной и гипотетических факторов-интерпретаторов. Отмеченные обстоятельства обуславливают возможность широкого применения рассмотренного подхода в практических приложениях.

## Список литературы

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Прикладная статистика. Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. – 471 с.

2. Дейвисон М. Многомерное шкалирование: Методы наглядного представления данных / Пер. с англ.- М.: Финансы и статистика, 1988. – 254 с.

3. Трусова А.Ю. Многомерное шкалирование структуры общества. //Вестник СамГУ, 2008. № 7 (66).

4. Ларичев О.И. Новое направление в теории принятия решений: вербальный анализ решений. //Информационные технологии и вычислительные системы (ИТиВС), №1, 1995. - сс. 24 – 34.

5. Макаров И.М., Виноградская Т.М., Рубчинский А.А., Соколов В.Б. Теория выбора и принятия решений. – М.: Наука, 1982. – 328 с.

6. Бешелев С.Д., Гурвич Ф.Г. Математико-статистические методы экспертных оценок. – М.: Статистика, 1980. – 264 с.

7. Кендалл М. Ранговые корреляции. – М.: Статистика 1975. – 302 с.

{social}