НЕКОТОРЫЕ ПОСТАНОВКИ ЗАДАЧ РАНЖИРОВАНИЯ И ИХ РЕШЕНИЕ НА ОСНОВЕ ПРИНЦИПА САМОСОГЛАСОВАННОСТИ[1]

В.З. Беленький, В.Г. Гребенников

Рассматривается задача ранжирования (упорядочения) N объектов, каждый из которых характеризуется значениями M различных признаков (как позитивных, так и негативных), так что исходной информацией является матрица $P(N \times M)$. Главная особенность работы в том, что ранжируются не только сами объекты, но определяются также и оценки (значимости) характеризующих их признаков. Решение задачи дается двухконтурным итеративным алгоритмом, сходимость которого интерпретируется как достижение самосогласованности полученных результатов. Приведены несколько примеров экспериментальных расчетов; рассмотрен пример реальной задачи ранжирования субъектов РФ.

Ключевые слова: двухуровневая задача ранжирования, веса и расстановка объектов, значимости характеристических признаков, принцип самосогласованности.

0. ВВЕДЕНИЕ

Методы парных сравнений занимают сегодня важное место в факторном анализе систем различной природы, в том числе социально-экономических, политических и др. Они находят применение в широком спектре задач выбора, выявления предпочтений, диагностики изменений в структуре анализируемой совокупности объектов или структуре состояний изучаемого объекта во времени и т.п. Проблеме упорядочения в группе объектов посвящена обширная литература мы, упомянем здесь наиболее близкие нам работы Т. Саати (Саати, Кернс, 1991) и Б.Г. Миркина (Миркин, 1974).

В работе рассматриваются два вида постановок задачи ранжирования – *одноуровневая* и *двухуровневая*.

0.1. Одноуровневая постановка

В этой постановке задача состоит в упорядочении группы объектов (в количестве N). Решение этой задачи мы даем на основе результатов их парных сравнений, которые задаются квадратной матрицей $A(N \times N)$ (первичная информация). Конечным результатом решения задачи является «ранжирование» объектов (т.е. упорядочение их по рангам); мы пользуемся также синонимом этого термина – «расстановка» (по местам).

В основе нашего подхода к построению расстановки лежит присвоение (исходя из результатов парных сравнений) каждому объекту его *веса*, так что расстановка объектов строится по убыванию их весов. Таким образом, задача построения расстановки сводится к расчету весов объектов. В статье предложен метод расчета весов, основанный на *принципе самосогласованности*[2]; этому посвящен раздел 1, материал которого – это несколько сокращенный вариант нашей предварительной публикации (Беленький, Гребенников, 2012), в которой одноуровневая постановка задачи ранжирования рассмотрена подробно.

[1] Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда (проект № 11-02-00-261).

[2] Выдвинутом В.Г. Гребенниковым в докладе «К построению расстановок на основе парных сравнений» на семинаре ЦЭМИ РАН «Неизвестная экономика» в июне 2012 г.

0.2. Двухуровневая постановка

В этом случае каждый объект описывается вектором значений *характеристических признаков* (в количестве M), так что первичная информация задается матрицей признаков $P(N \times M)$, на основе которой строится квадратная матрица $A = A(P)$ парных сравнений, и затем решается задача первого уровня, описанная в п. 0.1. Второй уровень состоит в том, что по результатам решения задач первого уровня (точнее, по полученному вектору $x \in R_+^N$ весов объектов) самим характеристическим признакам присваиваются оценки их значимости, так что решением задачи второго уровня является вектор *оценок признаков* $w \in R_+^M$.

С учетом оценок, матрица признаков P корректируется $P \to P(w)$, и на ее основе строится новая матрица парных сравнений $A = A(P(w))$. Возникает, таким образом, двузвенный итеративный контур

$$w_0 := (1, 1, ..., 1),$$

$$A_k := A(P(w_k)) \to x_k \to w_{k+1}, \; k = 0, 1, \dots . \quad (1)$$

Принцип самосогласованности в широком смысле состоит в том, что решением двухуровневой задачи в целом считается такая пара векторов (x, w), что в процессе (1) последовательные векторы w_k, w_{k+1} совпадают (соответственно, совпадают и векторы x_k, x_{k+1}). Постановка и метод решения этой задачи описан в разделе 2.

1. ПЕРВЫЙ УРОВЕНЬ – РАССТАНОВКА ОБЪЕКТОВ ПО РЕЗУЛЬТАТАМ ИХ ПАРНЫХ СРАВНЕНИЙ

1.1. Исходная информация

Исходной информацией является количество объектов N и матрица $A(N \times N)$ результатов их парных сравнений, предполагаемая неотрицательной. Элементы a_{ij}, a_{ji} отражают результат сравнения i-го и j-го объекта. Сравнение объектов может проводиться в количественной либо в качественной (*булевой*) форме.

Количественную форму сравнения удобнее всего описать на языке спортивных соревнований, считая, что элемент a_{ij} матрицы результатов показывает количество «голов» (в широком смысле), забитых участником (объектом) i во встрече с участником j. Например, в футболе это число голов (в узком смысле), в баскетболе – число попаданий в корзину соперника, в фехтовании – число нанесенных уколов и т.п. Вообще говоря, результирующие числа могут быть и нецелыми, выражающими, например, среднюю силу ударов, наносимых боксерами друг другу. Количество забитых голов можно связать с числом признаков одного объекта, по которым он превосходит (если признак позитивный) или уступает (если признак негативный) в сравнении с другим объектом.

При качественном сравнении принимается $a_{ii} = 0$

$$\begin{cases} (a_{ij} = 1, \; a_{ji} = 0) & \text{если } i \text{ лучше, чем } j, \\ (a_{ij} = a_{ji} = 1) & \text{если } i \text{ и } j \text{ эквивалентны.} \end{cases} \quad (2)$$

Отношение превосходства может быть задано непосредственно (например, экспертом) или синтезировано на основе ряда характеристических признаков (свойств).

Представление результата качественного сравнения в форме (2) позволяет рассматривать его как количественный; поэтому в дальнейшем мы будем в любом случае иметь дело только с количественной матрицей A, интерпретируя ее в спортивном («футбольном») смысле. Тогда элементы фиксированной строки i показывают количества голов, *забитых* игроком i всем остальным игрокам (полагаем $a_{ii} = 0$); элементы фиксированного столбца j суть количества голов *пропущенных* участником j во встречах с остальными участниками.

Подчеркнем, что важным преимуществом отношения предпочтения, построенного на парных сравнениях, является то, что оно не обязано быть транзитивным.

1.2. Веса участников

Как можно было бы характеризовать силу участников футбольного турнира с помощью весовых коэффициентов, составляющих в совокупности неотрицательный вектор $x = (x_1, \ldots, x_N)$? Наиболее простой способ – это взять отношение сумм забитых и пропущенных голов, т.е. по формуле

$$x_i = \sum_{j \neq i}^{N} a_{ij} \Big/ \sum_{j \neq i}^{N} a_{ji}, \; i = 1, \ldots, N. \tag{3}$$

Такой подход исходит из неявного предположения, что априори все участники равносильны: этим объясняется прямое суммирование в (3). Однако апостериори (по результатам турнира) это предположение, вообще говоря, не подтверждается: значения x_i оказываются различными; это означает, что формула (3) внутренне противоречива и нуждается в корректировке. Логично наделить всех участников весами (x_i) так, чтобы скорректированное с их помощью выражение в правой части (3) давало, при всех i, то же значение, что и слева – *принцип самосогласованности*.

Один из возможных подходов к корректировке правой части (3) таков[3]: гол, забитый участнику j, засчитывать с весом x_j, а гол, пропущенный от участника j, засчитывать с обратным весом $1/x_j$. При таком подходе правая часть (3) заменится выражением

$$\frac{\sum_{j \neq i} a_{ij} x_j}{\sum_{j \neq i} (a_{ji} / x_j)} =: f_i(x), \; i = 1, \ldots, N, \tag{4}$$

и условие самосогласованности выражается векторным нелинейным уравнением

$$x = f(x), f = (f_1, \ldots, f_N), x \in R_+^N. \tag{5}$$

Таким образом, искомые апостериорные веса участников турнира (назовем их *абсолютными*) образуют вектор, являющийся решением уравнения (5).

Примечание. С математической точки зрения уравнение (5) требует специального анализа относительно существования и единственности решения, а также вычислительного метода его нахождения. Здесь мы этих вопросов не касаемся, ограничившись приводимым ниже методом решения, показавшим[4] свою эффективность в экспериментальных расчетах.

Поскольку веса используются нами как инструмент ранжирования участников, в этом качестве вектор весов $x = (x_i)$, может определяться с точностью до произвольного нормирующего множителя. Удобно нормировать веса, введя вектор $y := x/v$, подобрав знаменатель v (скаляр) так, чтобы *среднее значение*

$$\overline{y} := \frac{1}{N} \cdot \sum_{i=1}^{N} y_i \tag{6}$$

равнялось единице (черта сверху будет использоваться и далее для обозначения среднего, в смысле (6), значения компонент того или иного вектора в R_+^N); определенный таким способом вектор нормированных весов (y_i) назовем *равновесным*.

1.3. Метод решения

Для равновесного вектора правая и левая части (5) отличаются некоторым положительным множителем $v > 0$, и уравнение самосогласованности имеет вид

$$y = vf(y). \tag{7}$$

Условие нормировки $\overline{y} = 1$ позволяет выразить скаляр v непосредственно через y:

---

[3] Описываемый вариант определения весов опробован нами хронологически первым (раньше других, которые мы считаем его модификациями). Именно в этом варианте был сформулирован принцип самосогласованности, который здесь выступает в наиболее логичной форме, и который был заложен затем во все последующие модификации.

[4] В регулярном случае (т.е. за исключением особых ситуаций (см. (Саати, Кернс, 1991, п. 2.3)).

$$1=\overline{y}=v\overline{f}(y)\Rightarrow v=\frac{1}{\overline{f}(y)}, \qquad (8)$$

и, таким образом, уравнение самосогласованности приобретает *нормированную форму*

$$y=F(y):=\frac{f(y)}{\overline{f}(y)}, \qquad (9)$$

где вектор f определен в (4).

Экспериментальные расчеты на компьютере показывают, что уравнение (9) в регулярном случае эффективно решается методом прямых итераций:

$$y^0:=(1,...,1),\;\; y^k:=F(y^{k-1}),\;\; k=1,2,\dots . \qquad (10)$$

1.4. Некоторые варианты модификаций

Можно предложить различные модификации описанного выше подхода, которые сохраняют основную идею самосогласованности, выраженную уравнением (5). В этом пункте мы приведем несколько вариантов, в которых метод решения прошел экспериментальную проверку.

1.4.1. Первый вариант модификации

Идея состоит в том, что гол, забитый участником j во встрече с участником i, засчитывается не с абсолютным весом x_j, а с относительным весом x_j/x_i. Функции (4) теперь таковы

$$f_i^{(1)}(x):=\frac{\sum_{j\neq i}a_{ij}(x_j/x_i)}{\sum_{j\neq i}a_{ji}(x_i/x_j)},\; i=1,\dots,N. \qquad (11)$$

Уравнение (9) и итеративный процесс (10) полностью сохраняются.

1.4.2. Второй вариант модификации

Вторая идея, дополняющая первую, состоит в том, что вместо отношений сумм, как в (4) берется их разность; тогда имеем

$$x_i=\sum_{j\neq i}a_{ij}(x_j/x_i)-\sum_{j\neq i}a_{ji}(x_i/x_j),$$

$$i=1,\dots,N. \qquad (12)$$

Эта форма уравнения прямо неприемлема, так как правая часть может быть отрицательной; но, учитывая, что в процессе суммирования значение x_i остается постоянным, можно записать (11) в виде

$$x_i^2\left(1+\sum_{j\neq i}a_{ji}/x_j\right)=\sum_{j\neq i}a_{ij}x_j \Leftrightarrow$$

$$\Leftrightarrow x_i=\left(\frac{\sum_{j\neq i}a_{ij}x_j}{1+\sum_{j\neq i}a_{ji}/x_j}\right)^{1/2} =: f_i^{(2)}(x). \qquad (13)$$

Далее записываем нормированное уравнение (9) и итеративный процесс (10).

1.4.3. Третий вариант модификации

Этот вариант можно считать «гибридом» абсолютного и относительного, но относительное надо понимать не в смысле дроби x_j/x_i, а в смысле разности x_j-x_i. Уравнение (5) запишем в виде

$$x_i=\sum_{j\neq i}a_{ij}x_j-\sum_{j\neq i}a_{ji}(x_i-x_j),\; i=1,\dots,N,$$

(забитый гол считается с абсолютным весом, а пропущенный – с относительным), или

$$x_i(1+c_i)=\sum_{j\neq i}(a_{ij}+a_{ji})x_j,\;\; c_i:=\sum_{j\neq i}a_{ji},$$

$$i=1,\dots,N.$$

Таким образом, уравнение самосогласованности таково:

$$x_i=\sum_{j\neq i}s_{ij}x_j,\;\; s_{ij}:=\frac{a_{ij}+a_{ji}}{1+c_i},\;\; i=1,\dots,N. \qquad (14)$$

Это система линейных уравнений, которая в векторной форме записывается в виде $x=Sx=:f^{(3)}(x)$, где квадратная матрица $S(N\times N)$ состоит из элементов s_{ij}. Так как система (14) однородна, то она имеет в общем случае (когда матрица S невырождена – ее определитель отличен от нуля) только тривиальное решение $x_i = 0$ при всех i. Чтобы получить нетривиальное нормированное решение y надо заменить условие равенства (5) на условие пропорциональности (7), которое, как обычно, сводится к нормированному уравнению (9) с последующим применением процесса (10).

Примечание. В данном случае мы получили стандартную задачу линейной алгебры на нахождение нормированного собственного вектора y и собственного числа v данной неотрицательной матрицы S. В силу теоремы Фробениуса-Перрона она всегда (в случае невырожденности) имеет положительное собственное число v (называемое *спектральным радиусом матрицы* S) и отвечающий ему нормированный собственный вектор $y \in R_+^N$. Решение единственно, и итеративный процесс (10) всегда сходится.

1.4.4. Четвертый вариант модификации

Этот вариант отличается от базового варианта тем, что в формуле (4) вместо отношения сумм забитых и пропущенных голов берется их разность. Уравнение самосогласованности таково

$$x_i = \sum_{j\neq i} a_{ij} x_j - \sum_{j\neq i} (a_{ji} / x_j), \; i = 1, ..., N. \tag{15}$$

Заметим, что хотя функция (15) монотонна на R_+^N, но она обращается в $-\infty$ при $x = 0$, поэтому прямое применение формулы (15) неприемлемо. Однако если мы ищем нормированное (в смысле (6)) решение, то всегда найдется константа $K > 0$ такая, что для любого нормированного вектора y будет выполняться неравенство

$$f_i^{(4)}(y) := \sum_{j\neq i} a_{ij} y_j - \sum_{j\neq i} (a_{ji} / y_j) + K \geq 0,$$

$$i = 1, ..., N. \tag{16}$$

Можно показать, что при больших K решение нормированного уравнения (9) с функцией (16) стремится к единичному вектору $e = (1, 1, ..., 1)$. Однако, оказывается, что существует такое (минимальное) значение $K_{мин}$, после которого упорядочение объектов (т.е. их места в расстановке) перестает зависеть от K, и эту установившуюся расстановку можно считать решением задачи, а в качестве весов принять вектор, отвечающий значению $K_{мин}$. Значение $K_{мин}$ может быть вычислено теоретически, но для практики достаточно найти его прямым численным экспериментом.

Отметим, что при каждом данном K решение уравнения (9) находится, как обычно, прямым итеративным алгоритмом (10).

1.5. Анализ экспериментальных расчетов

По всем описанным вариантам были проведены экспериментальные расчеты. Исходная матрица A генерировалась либо случайным образом, либо на основе какой-либо замысловатой формулы, дающей достаточно разнообразный спектр значений.

Приведенный ниже в табл. 1 пример для $N = 8$ иллюстрирует типичную картину результатов. Исходная матрица генерировалась формулой

$$a_{ij} = TRUNC\left[2\,|\sin(ij)|\cdot(2+\cos(i-2j))\right],$$

в которой функция TRUNC (в языке PASCAL) вычисляет целую часть аргумента; очевидно, элементы a_{ij} – целые числа в интервале [0, 5].

Вверху таблицы показана матрица A, а затем идут результаты расчета по основному варианту (его номер «0»), и четырем вариантам модификаций; показаны нормированные веса участников (равновесный вектор y) и соответствующее ранжирование (с номером ранга m) в порядке убывания веса. Дополним эти данные числом итераций k в процессе типа (10) с критерием остановки

$$\frac{1}{N}\cdot\sum_{j=1}^{N} |\, y_j^k - y_j^{k-1} | < \varepsilon, \text{ при } \varepsilon = 10^{-6}.$$

Вариант	0	1	2	3	4
$k =$	423	10	15	10	9

1.6. Сопоставительный анализ

Для сопоставления вариантов построим следующие две симметричные сопоставительные C–таблицы, в каждой из которых строки и столбцы отвечают различным вариантам 0 – 4 табл. 1.

В табл. 2 элемент c_{kl} показывает суммарное отклонение в ранжировании участников при сравнении данной пары вариантов, именно

$$c_{kl} := \sum_{j=1}^{N} | m_j^k - m_j^l |;$$

в дополнительном столбце в строке k показано среднее отклонение

$$\frac{1}{4} \cdot \sum_{l \neq k} c_{kl}.$$

В табл. 3 этот же элемент показывает расхождение (количество несовпадающих позиций) между вариантами (k, l).

1.7. Обсуждение результатов

1. Как видим разные варианты моделей дают разные результаты, причем, как показывают табл. 2, 3, различия существенны. Наиболее «сбалансированным» оказался вариант 2 – его отличие от других вариантов наименьшее, но это не означает, что его надо использовать во всех случаях. Думается, выбор того или иного варианта построения весов, должен диктоваться смыслом исследуемой проблемы.

2. При заданной точности, число требуемых итераций k во всех вариантах примерно одинаково, кроме варианта «0»; не ясно, почему это так.

Таблица 1

Результаты экспериментальных расчетов

i \ j	1	2	3	4	5	6	7	8
1	0	1	0	4	2	1	3	2
2	5	0	0	5	2	1	5	1
3	0	1	0	2	3	1	3	5
4	2	5	1	0	5	3	0	3
5	1	2	3	1	0	5	0	2
6	0	1	4	2	2	0	3	1
7	3	2	4	1	0	4	0	1
8	5	0	2	3	2	2	3	0
0. вес	0,922	1,484	0,934	1,137	0,874	0,742	0,862	1,025
M	5	1	4	2	7	8	6	3
1. вес	0,948	1,150	1,010	1,028	0,958	0,913	0,962	1,029
M	7	1	4	3	6	8	5	2
2. вес	0,931	1,219	1,004	1,051	0,939	0,874	0,946	1,036
M	7	1	4	2	6	8	5	3
3. вес	0,754	1,386	1,175	1,231	0,972	0,716	0,769	0,998
M	7	1	3	2	5	8	6	4
4. вес	0,954	1,130	1,010	1,030	0,963	0,919	0,964	1,030
M	7	1	4	3	6	8	5	2

Примечание. В четвертом варианте константа $K = K_{\text{мин}} = 50$.

Таблица 2

Суммарные отклонения между вариантами ранжирования

k \ l	0	1	2	3	4	Среднее
0	–	6	4	6	6	6,50
1	6	–	2	6	0	3,50
2	4	2	–	4	2	3,00
3	6	6	4	–	6	5,50
4	6	0	2	6	–	3,50

Таблица 3

Количество расхождений между вариантами

k \ l	0	1	2	3	4	Среднее
0	–	5	3	4	5	4,25
1	5	–	2	5	0	3,00
2	3	2	–	4	2	2,75
3	4	5	4	–	5	4,50
4	5	0	2	5	–	3,00

3. В общем виде предложенная нами схема описывается так. Подсчитываются взвешенные суммы числа забитых S_1 и пропущенных S_2 голов, они имеют вид

$$S_1 = \sum_j a_{ij}\varphi_1(x_i, x_j), \quad S_2 = \sum_j a_{ji}\varphi_2(x_j, x_i), \qquad (17)$$

где φ_1, φ_2 – взвешивающие функции, зависящие от весов участников. Содержательный смысл требует, чтобы функция φ_1 возрастала по второму аргументу (кому забивают) и не возрастала по первому (кто забивает), а функция φ_2 – наоборот. В условии самосогласованности (5) берется либо отношение, либо разность этих двух сумм.

Представляется, что такая схема плодотворна и допускает многие модификации взвешивающих функций. Кроме приведенных пяти вариантов нами были испробованы и некоторые другие; те варианты, для которых мы не смогли дать простой итеративный метод решения вида (10), остались вне рамок статьи.

2. ДВУХУРОВНЕВАЯ ПОСТАНОВКА ЗАДАЧИ РАНЖИРОВАНИЯ

В данном разделе дается сначала описание задачи ранжирования объектов не в «футбольных» терминах (как в предыдущем разделе), а в более содержательной постановке, когда матрица парных сравнений A не является первичной, а строится на основе матрицы характеристических признаков $P(N \times M)$, которая и является исходной информацией. Затем дается постановка двухуровневой задачи и итеративный метод ее решения.

2.1. Построение матрицы парных сравнений на основе таблицы (матрицы) признаков

Каждый объект $i \in [1, N]$ в подлежащей ранжированию группе характеризуется вектором признаков $p_i \in R_+^M$; совокупность этих векторов образует матрицу признаков $P(N \times M)$. Предполагаем, что каждый признак имеет свою единицу измерения; для позитивных признаков большее его значение повышает вес объекта, для негативных – наоборот. Будем говорить, что значение признака лучше, если оно повышает вес объекта.

Опишем предлагаемую в статье процедуру построения по заданной матрице P матрицы парных сравнений $A = A(P)$. Полагаем, что для любой пары участников (объектов) элемент a_{ij} вычисляется по формуле

$$a_{ij} := \begin{bmatrix} \text{число компонентов вектора } p_i, \\ \text{которые лучше соответствующих} \\ \text{компонент вектора } p_j, \\ \text{плюс половина числа равных} \\ \text{компонент этой пары векторов} \end{bmatrix}; \quad (18)$$

на «футбольном» языке это означает, что каждая лучшая компонента интерпретируется как «гол», забитый участником i во встрече с участником j, а при равенстве компонент каждому участнику добавляется 1/2.

Примечание. Значения a_{ij} не зависят от единиц измерения признаков, при этом сумма $a_{ij} + a_{ji}$ всегда равна M.

Формула (18) неявно предполагает (аналогично (3)), что все признаки, безотносительно к их содержанию, эквивалентны в смысле влияния на результаты сравнений. Ясно, что в реальных приложениях описываемой модели такое предположение выполняется далеко не всегда; модель требует корректировки, содержательный смысл которой – приписать каждому признаку *оценку его значимости* в контексте рассматриваемой задачи ранжирования. Приписывая признаку $k \in [1, M]$ оценку w_k, мы корректируем матрицу P, заменяя в столбце k элемент p_{ik} на $w_k p_{ik}$, и с полученной матрицей $P(w)$ применяем описанный выше подход.

Таким образом, предлагаемая процедура требует задания в исходной информации помимо матрицы P также и вектора оценок $w \in R_+^M$.

2.2. Двухуровневая задача ранжирования

Расширим постановку задачи ранжирования так, чтобы требуемый для построения матрицы A вектор оценок w не задавать *экзогенно* (в качестве дополнения к матрице P), а находить его *эндогенно*, т.е. так, чтобы на выходе мы получали не только равновесный вектор весов объектов y, но и вектор оценок признаков w; такую постановку мы и назвали двухуровневой. По аналогии с теорией двойственности можно назвать задачу нахождения вектора y (задача первого уровня) *прямой*, а задачу нахождения вектора w (задача второго уровня) *двойственной*.

Каждый вектор w определяет матрицу $P = P(w)$, и процедура, описанная в предыдущем пункте, строит матрицу парных сравнений $A = A(P(w))$ и находит равновесный вектор $y \in R_+^N$ как решение соответствующей прямой задачи (получаемое по одному из вариантов, описанных в разделе 1). Это означает, что мы имеем *отображение* $w \to y = Y(w)$. Если теперь мы построим отображение обратного перехода $y \to w = W(x)$, то можно определить двузвенный *оператор* T: $R_+^M \to R_+^M$, действующий по формуле $T(w) := W(Y(w))$. Принцип самосогласованности в широком смысле выражается условием

$$w = T(w), \quad w \in R_+^M, \tag{19}$$

Выполнение условия (19) означает, что найдена пара векторов (w, y) такая, что выполняются равенства

$$y = Y(w), \quad w = W(y); \tag{20}$$

нахождение этой согласованной пары и является двухуровневой задачей.

Отображение $Y(w)$ уже описано в предыдущем пункте. Для полного описания задачи осталось определить отображение $W(y)$, где y – произвольный нормированный вектор в R_+^N; этому посвящен следующий пункт.

2.3. Процедура второго уровня – построение оценок характеристических признаков

В п. 2.1 мы обозначили i-ю строку матрицы P как вектор $p_i \in R_+^M$ значений различных признаков на объекте i; обозначим здесь через $z_k \in R_+^N$ столбец k матрицы P, т.е. вектор значений признака k на различных объектах. Можно считать, что столбцы исходной матрицы P нормированы в смысле (6) (отметим, что нормировка не влияет на отображение $Y(w)$), и тогда можно считать, что каждый вектор z_k представляет собой вектор весов объектов «с точки зрения» k-го признака.

Пусть y – некоторый вектор весов объектов, получаемый как равновесный в результате итеративной процедуры первого уровня; сравним с ним каждый из векторов z_k, т.е. вычислим *вектор отклонений* $\delta = (\delta k)$

$$\delta := \|z_k - y\|, \; k = 1, ..., N, \tag{21}$$

где $\|\cdot\|$ обозначает некоторую норму (метрику) в пространстве R_+^N. Величина δ_k отражает отклонение в ранжировании объектов с точки зрения k-го признака от равновесного ранжирования. Логично полагать, что чем больше такое отклонение, тем менее значим соответствующий признак, и поэтому в качестве оценки признака можно взять некоторую убывающую функцию от его отклонения, т.е. принять

$$w_k := \psi(\delta_k), \; k = 1, ..., N, \tag{22}$$

где ψ – некоторая неотрицательная убывающая функция, задаваемая вариантно; таким образом, в качестве оценок характеристических признаков мы принимаем вектор $w = W(y)$ с компонентами (22).

2.4. Итеративный метод решения двухуровневой задачи

Полная постановка двухуровневой задачи требует исходного информационного паспорта со следующими позициями:

1) количество объектов N и признаков M;

2) матрица признаков P;

3) вариант решения прямой задачи (см. п. 1.4), или, в общем случае, функций φ_1, φ_2 (см. п. 1.7);

4) метрика $\|\cdot\|$, см. (21);

5) функция ψ, см. (22).

Эта информация однозначно определяет отображения $y = Y(w)$, $w = W(y)$ и тем самым двухуровневую задачу (20); задача решается итеративным процессом (1), использующим указанные отображения. Экспериментальные расчеты показывают, что итеративный процесс (1) эффективен: если он сходится, то сходится быстро (возможны нерегулярные случаи, когда возникают особые ситуации, см. сноску 2).

3. КОМБИНИРОВАННАЯ ПОСТАНОВКА ДВУХУРОВНЕВОЙ ЗАДАЧИ

В п. 2.1 была описана постановка задачи первого уровня на основе матрицы признаков P и задаваемого дополнительно вектора оценок w значимостей различных признаков, который задавался экспертом *экзогенно*. В следующем пункте 2.2 сформулирована постановка двухуровневой задачи, в которой вектор w не задается, а находится *эндогенно* в процессе решения, и тогда участие эксперта не требуется. В данном разделе описывается промежуточный вариант постановки задачи, когда участие эксперта предполагается в слабой форме: эксперт не должен задавать численные значения компонент вектора w (оценки значимостей признаков), а требуется лишь задать булеву *матрицу* $B(M \times M)$ сравнительной силы признаков. Будем считать, что эксперт задает матрицу B в следующей форме

$$b_{kl} := \begin{cases} 1, & \text{если признак } k \text{ "сильнее" } l; \\ 0, & \text{если признак } k \text{ "слабее" } l; \\ 1/2, & \text{если признак } k, l \text{ "равносильны",} \end{cases}$$

$$k, l \in [1, M]; \tag{23}$$

отметим, что при таком подходе $b_{kl} + b_{lk} = 1$ $\forall\, k, l$, что содержательно означает, что засчитывается каждая пара сравнений.

Роль матрицы B в данной постановке состоит в том, что в итеративной схеме двухуровневой задачи построение оценок признаков, описанное в п. 2.3, заканчивается не вектором с компонентами (22), а вектором с компонентами

$$w_k := \sum_{l=1}^{M} b_{kl}\psi(\delta_l), \quad k \in [1, M], \tag{24}$$

во всем остальном эндогенная процедура сохраняется. Такую постановку мы называем комбинированной.

Примечание. Сравнивая (22) и (24), видим, что в векторно-матричной форме связь между эндогенным и комбинированным векторами оценок дается равенством

$$w_{\text{комбин.}} = Bw_{\text{эндоген.}}.$$

4. ПРИМЕНЕНИЕ: РАНЖИРОВАНИЕ СУБЪЕКТОВ РФ ПО ДОСТУПНОСТИ ЖИЛЬЯ

Описанную выше процедуру покажем на примере ранжирования субъектов РФ по уровню доступности жилья. Исходя из имеющейся статистической информации[5] были выбраны следующие признаки, по которым проводилось парное сравнение указанных объектов: 1 – среднедушевой доход; 2 – ввод жилья; 3 – среднедушевой жилищный фонд; 4 – величина прожиточного минимума; 5 –

[5] Авторы с благодарностью получили ее в пользование от Н.Н. Ноздриной (ИНП РАН), одного из ведущих в стране специалистов в области исследования жилищных проблем, и обладающего опытом межрегиональных сравнений по доступности жилья (см., например, (Ноздрина, Шнейдерман, 2012)).

степень дифференциации доходов; 6 – цена жилья на первичном рынке; 7 – доля ветхого и аварийного фонда[6].

Признаки 5, 6 и 7 негативны, их более высокие значения соответствуют понижению уровня доступности жилья. Регионы, для которых отсутствуют данные по одному или нескольким выбранным признакам, не включены в рейтинг. Расчет весов объектов (первый уровень итеративного круга) проводился по варианту 4 (см. п. 1.4.4); значение константы $K_{мин}$ = 2000 было найдено экспериментально; расчет оценок значимости признаков (второй уровень итеративного круга) проводился по эндогенному варианту.

В табл. 4 показан итоговый порядок субъектов РФ по возрастанию занятых ими мест в рейтинге. В табл. 5 приведены нормированные оценки значимости признаков *w* (среднее значение равно единице). Подробный анализ представленных результатов не входит в задачи настоящей работы и заслуживает отдельного рассмотрения.

В дальнейшем прежде всего предстоит сравнить эти результаты с имеющимся опытом межрегиональных сравнений доступности жилья, использующим другие способы ранжирования. Здесь мы сделали это применительно к одному из таких способов, когда регионы упорядочиваются по среднему месту, занимаемому каждым из них по всему набору признаков (определяется место региона по каждому признаку отдельно, а затем берется среднее значение); в табл. 6 дана сравнительная по двум методам расчета (по среднему месту и по нашему) выборка из десяти ведущих и десяти замыкающих регионов.

Как видим, различия достаточно существенны. Среди лидеров не совпадают 3 региона, в нижней десятке таких несовпадений 4. К тому же различается и порядок тех, кто входит в составы десяток при применении обоих методов ранжирования. Более подробный анализ будет дан в последующих публикациях.

Таблица 4

Рейтинг-лист субъектов РФ по доступности жилья

Место	Субъект РФ	Место	Субъект РФ
1	Курская обл.	40	Ростовская обл.
2	Липецкая обл.	41	Мари Эл
3	Чувашия	42	Кировская обл.
4	Белгородская обл.	43	Тульская обл.
5	Сев. Осетия	44	Еврейская авт.обл.
6	Татарстан	45	Санкт-Петербург
7	Саратовская обл.	46	Астраханская обл.
8	Тамбовская обл.	47	Карелия
9	Адыгея	48	Псковская обл.
10	Калининградская обл.	49	Оренбургская обл.
11	Брянская обл.	50	Ярославская обл.
12	Пензенская обл.	51	Свердловская обл.
13	Мордовия	52	Новосибирская обл.
14	Челябинская обл.	53	Алтайский кр.
15	Воронежская обл.	54	Хакасия
16	Новгородская обл.	55	Томская обл.
17	Смоленская обл.	56	Архангельская обл.
18	Омская обл.	57	Ненецкий АО
19	Московская обл.	58	Камчатский АО
20	Рязанская обл.	59	Тюменская обл.
21	Орловская обл.	60	Алтай
22	Ульяновская обл.	61	Курганская обл.
23	Башкортостан	62	Самарская обл.
24	Ставропольский кр.	63	Магаданская обл.
25	Вологодская обл.	64	Красноярский кр.
26	Краснодарский кр.	65	Приморский кр.
27	Карачаево-Черкесия	66	Хабаровский кр.
28	Калмыкия	67	Сахалинская обл.
29	Костромская обл.	68	Бурятия
30	Нижегородская обл.	69	Пермский кр.
31	Тверская обл.	70	Забайкальский кр.
32	Калужская обл.	71	Коми
33	Кемеровская обл.	72	Амурская обл.
34	Ленинградская обл.	73	Москва
35	Владимирская обл.	74	Ханты-Мансийский АО
36	Дагестан	75	Тыва
37	Волгоградская обл.	76	Иркутская обл.
38	Удмуртия	77	Якутия
39	Ивановская обл.	78	Ямало-Ненецкий АО

[6] Единицы измерения признаков не влияют на результат, поэтому мы их не указываем.

Что касается табл. 2, то обращают на себя внимание контрастирующие оценки значимости позитивных признаков «среднедушевые доходы» и «среднедушевой жилфонд», выпадающие из ряда близких друг к другу оценок значимости других признаков.

Литература

Беленький В.З., Гребенников В.Г. Некоторые методы ранжирования объектов по результатам их парных сравнений // Анализ и моделирование экономических процессов. Вып. 9. М.: ЦЭМИ РАН, 2012.

Миркин Б.Г. Проблема группового выбора. М.: Наука, 1974.

Ноздрина Н.Н., Шнейдерман И.М. Социально-экономические проблемы обеспечения населения жильем // Народонаселение. 2012. № 3.

Саати Т., Кернс К. Аналитическое планирование. Организация систем. М.: Радио и связь, 1991.

Рукопись поступила в редакцию 15.08.2013 г.

Таблица 5

Оценка значимости признаков *w*

Признак	Оценка
Среднедушевой доход	0,353
Годовой ввод жилья	0,700
Среднедушевой жилой фонд	2,145
Прожиточный минимум	1,084
Степень дифференциации доходов	0,977
Цена жилья на первичном рынке	0,905
Доля ветхого и аварийного фонда	0,836

Таблица 6

Сравнение методов ранжирования
(слева – на основе средних мест объектов, справа – парных сравнений)

Верхняя десятка		Нижняя десятка	
Субъект РФ	Субъект РФ	Субъект РФ	Субъект РФ
Московская обл.	Курская обл.	Амурская обл.	Пермский кр.
Липецкая обл.	Липецкая обл.	Магаданская обл.	Забайкальский кр.
Белгородская обл.	Чувашия	Ханты-Мансийский АО	Коми
Чувашия	Белгородская обл.	Коми	Амурская обл.
Ямало-Ненецкий АО	Северная Осетия	Санкт-Петербург	Москва
Курская обл.	Татарстан	Ямало-Ненецкий АО	Ханты-Мансийский АО
Северная Осетия	Саратовская обл.	Дагестан	Тыва
Саратовская обл.	Тамбовская обл.	Якутия	Иркутская обл.
Калининградская. обл.	Адыгея	Тыва	Якутия
Тамбовская обл.	Калининградская. обл.	Москва	Ямало-Ненецкий АО