Банк России

МЕТОД ПОСТРОЕНИЯ ОБОБЩАЮЩИХ ПОКАЗАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ПАРНЫХ СРАВНЕНИЙ И ПАРЕТО ПРЕДПОЧТЕНИЙ

Информационно-аналитический материал

Г. Гамбаров

Москва
2023

ОГЛАВЛЕНИЕ

Материал подготовлен Департаментом статистики.

107016, Москва, ул. Неглинная, 12, к. В
Официальный сайт Банка России: www.cbr.ru

Георгий Гамбаров
Банк России, Департамент статистики
д. э. н., доцент
gambarovgm@cbr.ru

Аннотация

Предметом исследования являются методы построения сводного показателя, обобщающего свойства исходных, возможно, несопоставимых показателей. Предложены два относительно простых метода векторной оптимизации, основанных на принципе доминирования по Парето и не использующих предположений о значимости исходных показателей. В качестве регулирующего ограничения в первом методе выступает задание формы связи индикатора с исходными показателями. Второй метод использует предположение о монотонности ранга объекта относительно числа его предпочтений. Предложены индикаторы качества сводного показателя и значимости исходных показателей.

Ключевые слова: обобщающий показатель, векторная оптимизация, регуляризация, значимость показателя, доминирование по Парето, отношение порядка.

Key words: generalizing indicator, vector optimization, regularization, indicator significance, Pareto preference, order relation.

В практике статистических исследований часто возникает ситуация, когда некоторое сложное свойство изучаемого явления описывается несколькими показателями, каждый из которых характеризует его определенную сторону. При этом существует потребность в упорядочении элементов статистической совокупности по степени проявления данного свойства, а также в построении обобщающего индикатора (далее – индикатор), охватывающего все стороны свойства, характеризуемые исходными показателями (далее – показатели). Построение подобных индикаторов осуществляется методами векторной (многокритериальной) оптимизации [4, 5, 6].

Будем считать, что изучаемое свойство описывается m количественными или ранговыми показателями, известными у n объектов статистической совокупности. Требуется упорядочить объекты по степени проявления изучаемого свойства (ранжировать объекты) и построить обобщающий индикатор Y, максимально отражающий информацию, содержащуюся в исходных показателях.

Например, имея информацию о ряде банков, каждый из которых описывается набором показателей, отражающих устойчивость банка, требуется упорядочить банки по степени их устойчивости и построить модель зависимости устойчивости банка (Y) от данных показателей.

Стандартный подход к решению задачи векторной оптимизации – представление значений индикатора Y в виде суммы взвешенных значений показателей:

$$Y_i = w_1 * X_{1i} + \ldots + w_m * X_{mi}, \qquad (1)$$

где $X_{1i}, \ldots, X_{mi}$ – значения показателей i-го объекта;
$w_1, \ldots, w_m$ – веса показателей, заданные экспертным путем.

От точности субъективного задания весов показателей полностью зависит качество индикатора. При этом экспертные оценки осложняются тем, что некоторые показатели невозможно сопоставить, например демографические и затратные показатели.

Задача векторной оптимизации является некорректной и допускает множество решений, поэтому для ее решения требуется определенная регуляризация – введение некоторых дополнительных требований и/или ограничений. В работе предлагаются два подхода к решению данной задачи, не требующих столь сильных предположений, как априорная оценка значимости показателей.

1. Формализация задачи

Будем считать, что все m показателей n объектов статистической совокупности ориентированы таким образом, что с ростом каждого из них усиливается описываемое свойство. Подобной ориентации показателей можно добиться, например, заменой знака на противоположный у показателя, рост которого не приводит к возрастанию обобщающего индикатора.

Естественным представляется следующее требование.

Требование монотонности: рост каждого показателя не должен приводить к уменьшению обобщающего индикатора Y.

Из данного требования следует, что если все значения исходных показателей одного объекта не меньше соответствующих значений другого объекта и хотя бы одно из них больше соответствующего значения переменной второго объекта, то значение обобщающего индикатора Y первого объекта должно быть больше значения обобщающего индикатора Y второго объекта, то есть первый объект Парето предпочтительней второго.

В соответствии с требованием монотонности на множестве n объектов введем отношение нечеткого порядка $R = \{a_{ij}\}$, i, j = 1, …, n. Будем считать, что i-й объект предпочтительней j-го объекта (победил его, лучше него), если значения всех показателей i-го объекта больше соответствующих значений j-го объекта или равно им. При этом хотя бы один из показателей i-го объекта строго больше соответствующего показателя j-го объекта. В этом случае величина a_{ij} = 1.

Если же значения всех показателей i-го объекта меньше соответствующих значений j-го объекта или равно им, при этом хотя бы один из показателей i-го объекта меньше соответствующего показателя j-го объекта, то j-й объект предпочтительней i-го объекта и величина a_{ij} = -1.

Если между i-м и j-м объектами невозможно установить отношение предпочтительности (выявить победителя), то они считаются эквивалентными и величина a_{ij} = 0. В математической форме определение величин a_{ij} имеет вид:

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \geq X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} > X_{\nu j}\text{, то } a_{ij} = 1,$$

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \leq X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} < X_{\nu j}\text{, то } a_{ij} = -1, \qquad (2)$$

$$\text{иначе } a_{ij} = 0,$$

где ν = 1, …, m (m – число показателей);

$X_{\nu i}$ – значение ν-го показателя i-го объекта.

Легко показать, что введенное отношение нечеткого порядка *R* рефлексивно, антисимметрично и транзитивно. Оно описывается матрицей *A*, элементы которой a_{ij} принимают значения 1, (-1), 0:

$$A = \{a_{ij}\}. \qquad (3)$$

Требование монотонности для произвольной пары объектов, играющее ключевую роль в регуляризации по Парето, может быть записано в виде:

$$\text{ЗНАК } (Y_i - Y_j) = a_{ij}. \qquad (4)$$

Число невыполнения требования монотонности для индикатора Y, равное:

$$J = \sum_{i<j} |\text{ЗНАК}(\text{ЗНАК}(Y_i - Y_j) - a_{ij})|, \qquad (5)$$

отражает качество построенного индикатора Y: чем меньше величина J, тем выше качество индикатора. Заметим, что функция «ЗНАК ()» используется здесь дважды, чтобы избежать значения модуля, равного 2.

Коэффициент качества индикатора вычисляется по формуле:

$$K = 1 - \frac{J}{n * (n-1)}, \qquad (6)$$

где *J* – число случаев невыполнения требования монотонности (5);
n – число объектов совокупности.

В случае когда требование монотонности (4) не выполняется ни для одной пары объектов, коэффициент качества индикатора (6) принимает значение 0; когда требование монотонности (4) выполнено для всех пар объектов, коэффициент качества индикатора (6) достигает значения 1.

Заметим также, что значения индикатора Y однозначно задают упорядочение объектов по степени проявления изучаемого свойства.

2. Оптимизационный индикатор

При построении оптимизационного индикатора в качестве регулирующего ограничения выступает задание формы индикатора Y в параметрической форме. При заданной форме зависимости индикатора от показателей построение индикатора может быть сформулировано как оптимизационная задача.

Заметим, что данное предположение тем слабее, чем больше количество объектов и чем выше значимость показателей.

Пусть форма индикатора Y задается *u* параметрами:

$$Y = Y(\beta_1, \dots, \beta_u), \qquad (7)$$

тогда построение индикатора *Y* сводится к поиску параметров $\beta_1, \dots, \beta_u$, минимизирующих величину *J* из (5).

Например, если индикатор *Y* задан в линейном виде, его значения для i-го объекта задаются выражением:

$$Y = \beta_1 * X_{1i} + \dots + \beta_m * X_{mi}, \qquad (8)$$

где m – число исходных показателей;
$X_{1i}, \dots, X_{mi}$ – значения показателей i-го объекта;
$\beta_1, \dots, \beta_m$ – искомые коэффициенты.

В реальных задачах величина *J* из (5) далеко не всегда равна 0. Чем сложнее задача и выше ее размерность (число исходных показателей), тем больше величина *J*.

Оптимизационный индикатор в линейной форме (8) совпадает по форме с индикатором (1), в котором в качестве параметров используются веса, заданные экспертами. Однако первый из них более надежен, поскольку при его построении не использовались экспертные оценки значимости показателей.

Заметим, что экспертные оценки могут быть эффективно использованы при определенном уровне знаний об изучаемом явлении. В частности, они могут быть полезны на последующих этапах построения индикатора после тщательного анализа промежуточных результатов.

3. Балльный индикатор

Во втором способе регуляризации по Парето используется предположение, что значение индикатора *Y* выше у того объекта, у которого больше число чистых предпочтений, то есть разность между числом его предпочтений другим объектам (побед над другими объектами) и числом предпочтений ему других объектов (числом поражений).

Число чистых предпочтений i-го объекта S_i равно сумме элементов i-ой строки матрицы *A* из (3):

$$S_i = \sum_{j=1}^{n} a_{ij}. \qquad (9)$$

Величины $Y_i = S_i$ представляют собой значения обобщающего индикатора Y изучаемого свойства, измеренного в баллах, то есть в порядковой шкале.

При использовании балльного индикатора как количественного показателя необходимо понимать, что это означает принятие **второго ограничивающего предположения**.

Данное предположение тем слабее, чем больше количество объектов и чем выше значимость показателей.

Для построения обобщающего индикатора в количественной шкале необходимо использовать второе ограничивающее предположение, состоящее в том, что в качестве значений обобщающего индикатора Y_i можно использовать его балльные оценки S_i. Данное предположение тем слабее, чем больше количество объектов.

Удобно использовать нормированные величины S_i, изменяющиеся от 0 до 100:

$$Y_i = \frac{S_i - \min(S_i)}{\max(S_i) - min\,(S_i)} * 100, \qquad (10)$$

где $min\,(S_i)$, $\max\,(S_i)$ – минимальное и максимальное значения S_i соответственно.

Построенный таким образом индикатор Y_i не является единственным решением поставленной задачи даже в условиях сделанных предположений. Любое монотонное преобразование величин Y_i из (10) строит индикатор, эквивалентный полученному: их характеристики качества, вычисленные по формуле (5), будут совпадать.

Как показал опыт использования описываемого метода, линейный индикатор целесообразно выбирать при небольшом числе показателей. При большем числе показателей удобнее использовать балльный индикатор. При этом качество балльного индикатора Y_i из (10), как правило, выше качества оптимизационного индикатора Y_i, заданного в линейном виде и построенного с учетом более слабых предположений.

Описанные методы использовались в ряде статистических исследований, в том числе при построении кривой бескупонной доходности и индикатора ставок рынка репо Московской Биржи, определении весов в общей формуле композитного веса, а также при построении обобщающего коэффициента концентрации.

4. Определение значимости показателей

При построении обобщающего индикатора часто возникает потребность в оценке значимости исходных показателей, то есть оценке их вклада в уровень и изменчивость индикатора. Представляется естественным считать, что значимость показателя тем выше, чем выше его информативность, измеряемая суммарным числом предпочтений одного объекта над другим. Коэффициент значимости i-го показателя:

$$Z_i = \frac{2 * E_i}{n * (n-1)}, \qquad (11)$$

где E_i – число предпочтений (число единиц) в матрице A из (3), построенной по только одному i-му показателю.

В соответствии с (11) максимальной значимостью (+1) будет обладать показатель, не имеющий равных значений, а минимальной значимостью (0) будет обладать показатель, все значения которого равны между собой.

Заметим, что значимость, рассчитанная по формуле (11), пропорциональна значимости, оцениваемой как снижение качества обобщающего индикатора (7) в результате исключения i-го показателя. В частности, при двух исходных показателях после удаления одного из них в качестве нового индикатора выступает оставшийся показатель. Если его качество (5) ниже качества первого (удаленного) показателя, это означает, что снижение качества обобщающего индикатора (5) в результате исключения первого показателя будет больше, чем при исключении второго показателя.

Заметим, что чем выше значимость показателя, тем он ближе к количественному показателю, то есть тем в большей степени с его значениями допустимы арифметические операции.

Наряду с коэффициентами значимости (11) можно использовать нормированные коэффициенты значимости, вычисляемые по формуле:

$$Z_{i,norm} = \frac{z_i}{\sum_{j=1}^{m} z_j}. \qquad (12)$$

Заметим, что сумма нормированных коэффициентов значимости равна 1:

$$\sum_{j=1}^{m} Z_{i,norm} = 1.$$

Наличие коэффициентов значимости исходных показателей позволяет применять следующую схему построения индикатора. На первом этапе в рассмотрение включаются все показатели, связанные с изучаемым свойством. На этом этапе важно отобрать все существенные показатели. На втором этапе на основании коэффициентов значимости из рассмотрения исключаются все несущественные показатели. Если после этого осталось до пяти показателей, можно рекомендовать использовать линейный оптимизационный индикатор, иначе – балльный.

5. Сравнение оптимизационного и балльного индикаторов

Сравнение оптимизационного (линейного) и балльного показателей проведем на примере взвешивания ставок рынка межбанковских депозитов.

При расчете средней ставки на рынке межбанковских депозитов RUONIA вес (значимость) ставки тем выше, чем больше объем совершенных по данной ставке сделок и чем больше число участников этих сделок. Построим индикатор значимости ставок по данным двух показателей: объема сделок и числа их участников, на примере данных за 1 февраля 2019 г. (табл. 1).

ЗНАЧЕНИЯ ИСХОДНЫХ ПОКАЗАТЕЛЕЙ ЗНАЧИМОСТИ СТАВОК *Табл. 1*

Номер ставки	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Ставка, %	7	7,2	7,25	7,3	7,5	7,55	7,6	7,65	7,76	7,85
Объем, млрд руб.	12	1	11	2	24,6	6	9,2	35	0,2	13
Число контрагентов	1	2	5	2	13	2	6	4	2	2

В таблице 2 размещены значения элементов матрицы А, вычисленные в соответствии с (1); в последней колонке представлены суммы баллов S_i, рассчитанные по (9).

ЗНАЧЕНИЯ ЭЛЕМЕНТОВ МАТРИЦЫ А И ЧИСЛА БАЛЛОВ *Табл. 2*

Номер ставки	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**S_i**
1	0	0	-1	0	-1	0	0	-1	0	-1	-4
2	0	0	-1	0	-1	-1	-1	-1	0	-1	-6
3	1	1	0	1	-1	1	-1	0	1	1	4
4	0	0	-1	0	-1	-1	-1	-1	0	-1	-6
5	1	1	1	1	0	1	1	0	1	1	8
6	0	1	-1	1	-1	0	-1	-1	1	-1	-2
7	0	1	1	1	-1	1	0	0	1	0	4
8	1	1	0	1	0	1	0	0	1	1	6
9	0	0	-1	0	-1	-1	-1	-1	0	-1	-6
10	1	1	-1	1	-1	1	0	-1	1	0	2

Линейная форма индикатора, полученная в результате минимизации выражения (4) и нормированная так, что максимальное значение индикатора равно 100 и имеет вид:

$$Y_{\text{линейный}} = 0{,}65 * \text{Объем сделок} + 6{,}47 * \text{Число контрагентов}. \qquad (13)$$

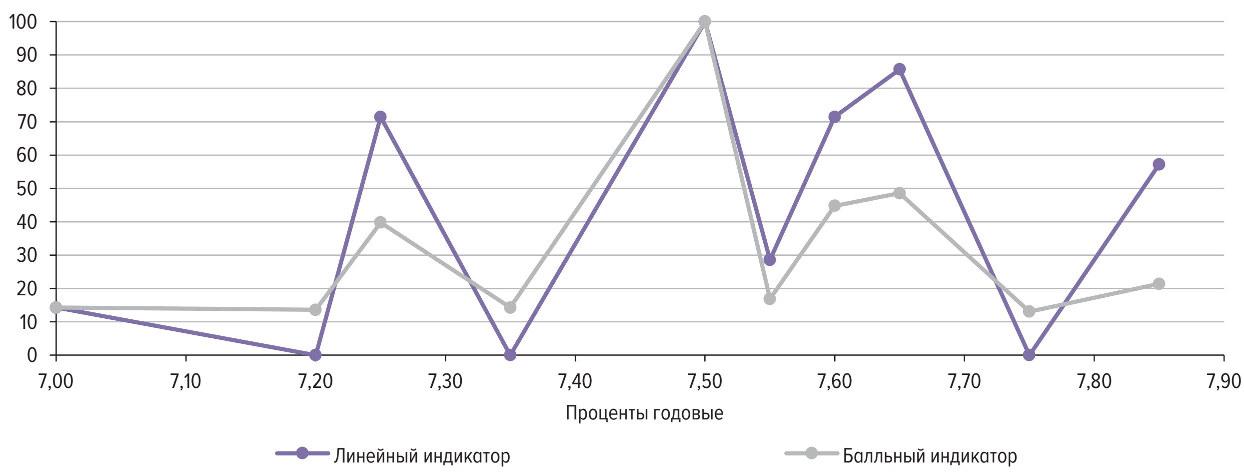
Значения индикатора в линейной оптимизационной форме, рассчитанные по формуле (13), представлены во второй строке таблицы 3. В третьей строке таблицы 3 представлены значения индикатора, рассчитанные на основании баллов по формуле (10).

ЗНАЧЕНИЯ ЛИНЕЙНОГО И БАЛЛЬНОГО ИНДИКАТОРОВ *Табл. 3*

Ставка, %	7,00	7,20	7,25	7,35	7,50	7,55	7,60	7,65	7,75	7,85
Y оптимизационный	14,2	13,6	39,5	14,2	100	16,8	44,8	48,5	13,1	21,4
Y балльный	14,3	0	71,4	0	100	28,6	71,4	85,7	0	57,1

На рисунке 1 представлены значения линейного оптимизационного и балльного индикаторов.

ЗНАЧЕНИЯ БАЛЛЬНОГО И ЛИНЕЙНОГО ОПТИМИЗАЦИОННОГО ИНДИКАТОРОВ *Рис. 1*



Как видно из таблицы 3 и рисунка 1, значения линейного и балльного индикаторов достаточно близки. Качество линейного и балльного индикаторов, рассчитанное по формуле (6), равно 0,76 и 0,78 соответственно. Незначительное превышение качества балльного индикатора над линейным является типичным. В то же время линейный индикатор представляется более удобным для аналитической работы.

Коэффициент значимости первого показателя (объема сделок), вычисленный по формуле (11), равен 1; второго (числа контрагентов) – 0,8. Нормированные коэффициенты значимости (12), сумма которых равна 1, равны 0,56 и 0,44 соответственно, что достаточно близко к среднемесячным значениям начала 2019 г. (0,54 и 0,46).

6. Использование экспертных суждений для повышения качества обобщающего индикатора

В большинстве статистических исследований на этапе интерпретации результатов обнаруживается, что часть результатов не согласуется с содержательными представлениями об изучаемом явлении. При построении обобщающего показателя балльным методом может возникнуть ситуация, когда одинаковый ранг имеют объекты, которые не считаются равнозначными. Чаще всего подобная ситуация возникает при малом числе объектов. В этом случае целесообразно формализовать экспертные суждения, правомерность которых не вызывает существенных сомнений.

Ниже предлагаются три способа использования содержательных знаний экспертов для повышения упорядоченности объектов и качества обобщающего индикатора.

Способ 1. Использование суждений о существенном различии значимости показателей

Данный способ эффективен при условии, что среди исходных показателей можно выделить группу показателей (группа *G*), наиболее важных для изучаемого свойства.

После выделения группы наиболее важных показателей (группа G) определяется группа объектов O с равным числом чистых предпочтений S_i из (9). Для объектов данной группы O с одинаковым числом баллов пересчитываются элементы матрицы A = $\{a_{ij}\}$ из (3) по следующему правилу (правило Парето предпочтений).

Объект выделенной группы *O* считается предпочтительней другого объекта данной группы, если значение какого-либо показателя из группы показателей G первого объекта не меньше аналогичного показателя второго объекта и при этом хотя бы одно из сравниваемых значений первого объекта больше соответствующего значения второго объекта. Более формально значения скорректированных значений a_{ij} вычисляются по следующему правилу:

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \geq X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} > X_{\nu j}\text{, то } a_{ij} = b_1,$$

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \leq X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} < X_{\nu j}\text{, то } a_{ij} = -b_1, \quad (14)$$

$$\text{иначе } a_{ij} = 0,$$

где ν – номер показателя группы *G*;
i, j – объекты группы *O;*
b_1 – положительная величина, меньшая 1;
$X_{\nu i}$ – значение ν-го показателя i-го объекта.

Величина b_1 задается таким образом, чтобы после корректировки матрицы A = $\{a_{ij}\}$ из (3) упорядочение объектов группы *O* относительно остальных объектов сохранялось.

Способ 2. Использование суждений о существенном различии значений показателя

В случае когда эксперты считают, что значительное превышение величины не самого важного исходного показателя (не попавшего в группу *G*) одного объекта над аналогичным показателем другого объекта может служить основанием для принятия решения о предпочтении первого объекта, эффективен следующий подход.

После определения групп объектов, для которых желательно скорректировать результаты парных сравнений объектов, выделяются показатели (группа *U*), по которым будет определяться значимое превышение значений. Один объект (i-ый) из группы объектов *O* полагается предпочтительней другого объекта (j-го) из группы объектов *O* в соответствии со следующим правилом:

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \geq k_v * X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} > k_v * X_{\nu j}\text{, то } a_{ij} = b_{\nu 2},$$

$$\text{если } \forall\ \nu\text{: } X_{\nu i} \leq k_v * X_{\nu j} \text{ и } \exists\ \nu\text{: } X_{\nu i} < k_v * X_{\nu i}\text{, то } a_{ij} = -b_{\nu 2}, \qquad (15)$$

$$\text{иначе } a_{ij} = 0,$$

где ν – номер показателя из группы *U*;
i, j – объекты группы *O;*
k_v – положительная величина, большая 1;
$b_{\nu 2}$ – положительная величина, меньшая 1.

Величины $k_v, b_{\nu 2}$ могут задаваться как равными для всех показателей группы *U*, так и индивидуально, для каждого показателя группы *U*. При этом величина $b_{\nu 2}$ задается таким образом, чтобы после корректировки матрицы A = $\{a_{ij}\}$ из (3) упорядочение объектов группы *O* относительно остальных объектов сохранялось.

Заметим, что к группе *U* могут быть отнесены все показатели.

Способ 3. Экспертная оценка предпочтительности объектов

Данный способ целесообразно применять в случае, когда равные ранги получили объекты, степень проявления изучаемого свойства у которых, по мнению экспертов, существенно различается. С целью формализации данного суждения для каждой выделенной группы объектов *O* с равными рангами проводятся следующие действия.

Среди объектов группы *O* устанавливаются предпочтения части объектов над другими. Для этого в матрице A = $\{a_{ij}\}$ из (3) предпочтение i-го объекта над j-м объектом устанавливается непосредственно путем замены величины $a_{ij} = 0$ на элемент $a_{ij} = b_3$ и величины $a_{ji} = 0$ на элемент $a_{ji} = -b_3$.

Положительная величина b_3 задается таким образом, чтобы после корректировки матрицы A = $\{a_{ij}\}$ из (3) упорядочение объектов группы *O* относительно остальных объектов сохранялось. В частности, величину b_3 можно установить равной $\frac{1}{r}$, где r – максимум из числа предпочтений i-го и j-го объектов, сделанных для i-го и j-го объектов группы *O*.

Данные способы использования содержательных знаний экспертов можно применять как по отдельности, так и совместно. При этом даже в случае использования всех трех способов сделанные в них предположения существенно слабее априорного задания весов показателей.

Заключение

Предложено два способа упорядочения объектов и построения обобщающего индикатора методами регуляризации по Парето: оптимизационный и балльный. В первом из них регулирующим ограничением выступает задание формы связи индикатора с исходными показателями.

Во втором способе для упорядочения объектов делается предположение о том, что значение индикатора Y выше у того объекта, у которого больше число чистых предпочтений, то есть разность между числом его предпочтений другим объектам (побед) и числом предпочтений ему других объектов (поражений).

Для построения количественного обобщающего индикатора дополнительно делается предположение о том, что балльные оценки можно использовать в качестве количественного показателя. Следует отметить, что, несмотря на дополнительное предположение, качество балльного индикатора оказывалось, как правило, выше качества линейного оптимизационного индикатора.

Предложены три способа повышения качества построения обобщающих показателей путем формализации экспертных суждений, полученных на этапе интерпретации промежуточных результатов.

Предложены индикатор качества сводного показателя и коэффициенты значимости исходных показателей, сумма которых равна 1.

Список литературы

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998. – 1022 с.
2. Гамбаров Г.М. Метод регуляризации финансовых показателей по Парето. – М.: Финансы и кредит, 2006. 6 (230).
3. Гороховик В.В. Выпуклые и негладкие задачи векторной оптимизации. – М.: Машиностроение, 2015. – 489 с.
4. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения. – М.: Радио и связь, 1981. – 560 с.
5. Макаров И.В., Виноградская Т.М., Рубчинский А.А., Соколов В.Б. Теория выбора и принятия решений. – М.: Наука, 1982.
6. Машунин Ю.И. Методы и модели векторной оптимизации. – М.: Наука, 1986. – 143 с.
7. Ногин В.Д. Принятие решений в многокритериальной среде: количественный подход. – М.: Физмат, 2002. – 237 с.
8. Coleman T. Branch M.A. Grace A. Optimization Toolbox. For use with MATLAB / The Marth Work, Inc. Printing History: January 1999.
9. Waterson M. Economic theory of the industry. Cambridge University Press, 1984.