УДК 519.81

ОПРЕДЕЛЕНИЕ РЕЙТИНГА ОБЪЕКТОВ

Е. М. Аристова

Воронежский государственный университет

Поступила в редакцию 12.07.2013 г.

Аннотация. В статье рассматривается задача определения рейтинга объектов, и предлагаются три алгоритма многокритериальной оптимизации для построения такого рейтинга. А также рассматривается пример построения рейтинга банков.
Ключевые слова: рейтинг объектов, весовые коэффициенты, плохо формализуемая задача, задача кластеризации, алгоритм линейной свертки показателей, алгоритм «идеальной точки», алгоритм многокритериальной оптимизации для определения рейтинга объектов, ранжирование объектов.
Annotation. This article considers the problem of rating objects definition and offers three multi-objective optimization algorithms for constructing such a rating. And also the example of rating banks is considered.
Keywords: Rating facilities, weights, poorly formalized problems, the task of clustering, algorithm for linear convolution of indicators, the «idealspot» algorithm, multi-objective optimization algorithm to determine the rating of objects, ranging facilities.

Рейтинг (англ. *rating*) — числовой или порядковый показатель, отображающий важность или значимость определенного объекта или явления. В жизни очень часто необходимо принимать решения о ранжировании объектов в различных сферах деятельности.

ПОСТАНОВКА ЗАДАЧИ ОПРЕДЕЛЕНИЯ РЕЙТИНГА ОБЪЕКТОВ

Пусть для каждого из m объектов множества $X = \{1, 2, \ldots, m\}$ заданы n числовых неотрицательных значений показателей (характеристик), т. е. задана матрица $A = (a_{ij})$, $i = \overline{1,m}$, $j = \overline{1,n}$. В этой матрице каждому объекту $i \in X$ соответствует строка с номером i, а каждому показателю – столбец с номером j. В матрице A нет ни одной строки с нулевыми значениями показателей, так как рассмотрение таких объектов не имеет смысла. В матрице A также отсутствуют столбцы, все элементы которых равны нулю. Предполагается, что большие значения показателей более предпочтительны по отношению к меньшим их значениям.

Характеристикам соответствуют весовые коэффициенты $\vartheta_j \geq 0$, $j = \overline{1,n}$, значения которых нормируются. Нормированные значения весовых коэффициентов имеют вид $w_j = \frac{\vartheta_j}{\sum_{j=1}^{n} \vartheta_j}$, при этом $\sum_{j=1}^{n} w_j = 1$. Вычисляются новые значения показателей с учетом весовых коэффициентов, $b_{ij} = a_{ij} w_j$. Значения показателей b_{ij} также нормируются: определяются $b_j = \max_{1 \leq i \leq m} b_{ij}$, нормированные значения x_{ij} показателей b_{ij} с учетом весов w_j имеют вид $x_{ij} = \frac{b_{ij}}{b_j}$, $i = \overline{1,m}$, $j = \overline{1,n}$. Полученные значения x_{ij} удовлетворяют условиям $0 \leq x_{ij} \leq 1$ и не имеют размерности. Отсутствие размерности делает осмысленными математические и логические операции над разными по содержанию показателями.

Понятие рейтинга объекта трактуется по-разному и не всегда ему можно дать формальное определение. Вместо этого понятия можно использовать такие, как престижность, важность, вес, значимость и другие качества объекта, однако никакие из них невозможно определить строго и описать с применением числовых характеристик. Рассматриваемая задача относится к классу так называемых *плохо формализуемых задач*. При этом неизвестно, является ли используемый набор показателей достаточно представительным для определения рейтинга объекта; выявление такого набора является само по себе нетривиальной задачей.

Рассматриваемая задача относится к классу многокритериальных задач, если в качестве критериев рассматривать показатели. При этом многокритериальная задача может быть сформулирована как задача, в которой мы стремимся придать каждому критерию максимально возможное значение. Алгоритмы решения таких задач описаны, например в [1, 2].

Аксиоматическое решение проблемы оценки важности в многокритериальных задачах предложено в [3]. Эта оценка основана на парных сравнениях критериев (показателей). При этом для каждой пары критериев f_i и f_j введен линейный порядок, т.е. известно, что либо $f_i \succ f_j$, либо $f_j \succ f_i$, либо они равноценны ($\succ$ – знак предпочтения).

Остановимся на проблеме определения коэффициентов важности критериев, т. е. *весовых коэффициентов*. В работах [4, 5] рассмотрены различные аспекты этой проблемы, стоит отметить, что рассматриваемая задача относится к задачам выбора вариантов.

Решение задачи определения рейтинга может быть выполнено как для всего набора из n показателей, так и для любого подмножества $J = \{j_1, j_2, \dots, j_k\}$ различных показателей этого набора. С формальной точки зрения представленная задача – это задача кластеризации [6] с некоторыми дополнительными условиями. Задача кластеризации состоит в разделении исследуемого множества объектов на группы «похожих» объектов, называемых *кластерами*. В задаче кластеризации отнесение каждого из объектов данных осуществляется к одному (или нескольким) из заранее неопределенных классов. Разбиение объектов данных по кластерам осуществляется при одновременном их формировании. Определение кластеров и разбиение по ним объектов данных выражается в итоговой модели данных, которая является решением задачи кластеризации [7].

Под *кластером* в рассматриваемой задаче понимается множество объектов, имеющих одинаковый рейтинг. Множество объектов X необходимо разбить на p непересекающихся подмножеств $X_1, \dots X_p$, таких, что $\bigcup_{k=1}^{n} X_k = X$, $X_k \cap X_s = 0$, $s \neq k$, при этом число подмножеств p может быть заранее неизвестно, т. е. определяться в процессе решения задачи. В некоторых из рассматриваемых далее алгоритмов это число задано заранее. Возможны также дополнительные ограничения вида $x_{\min} \leq |X_i| \leq x_{\max}$ для некоторых или всех подмножеств.

Дополнительные условия состоят в том, что разбиение производится для выделенного набора показателей и в соответствии с максимальным значением некоторых функций $g_i(A)$, $i = \overline{1, k}$, вычисленных для каждого из объектов. Из высказанного ранее предположения о предпочтительности больших значений показателей по сравнению с меньшими следует: если для двух объектов с номерами s и k выполнены условия $x_{sj} \geq x_{kj}$, $j = \overline{1, n}$, то рейтинг объекта с номером s не может быть ниже, чем рейтинг объекта с номером k.

В процессе разбиения каждому из подмножеств X_k ставится в соответствие некоторый показатель q_k, называемый *рейтингом объекта*; все объекты из множества X_k неразличимы по этому показателю. Значения показателей q_k – рейтингов объектов – зависят от алгоритмов их вычисления.

АЛГОРИТМЫ РЕШЕНИЯ ЗАДАЧИ

В настоящей работе рассмотрим три алгоритма многокритериальной оптимизации, которые можно применить к задаче определения рейтинга объектов. Рассмотрим их более подробно.

Алгоритм линейной свертки показателей

Каждому объекту $i \in X$ ставится в соответствие *линейная свертка* его показателей

$L_i = \sum_{j \in J} x_{ij}$, вычисляемая с применением значений весовых коэффициентов w_j, $j = \overline{1, n}$. Отметим, что при использовании данного алгоритма число множеств $p > 1$ задано.

Обозначим $l = \min_i L_i$, $L = \max_i L_i$ и вычислим их значения. Далее вычислим шаг $h = \frac{L - l}{p}$ и распределим исследуемые объекты по отрезкам $\left[l + (i-1)h, l + ih \right]$, $i = \overline{1, p}$ в соответствии со значениями сверток L_i. Тогда объект с минимальным значением L_i попадет в первый из отрезков, а с максимальным – в последний отрезок с номером p. Объекты, значения сверток которых находятся в интервале с номером k, имеют рейтинг $q_k = p - k + 1$ и образуют множество X_{p-k+1} (чем выше рейтинг множества объектов, тем меньше его номер).

Алгоритм идеальной точки

В каждом из столбцов матрицы со значениями показателей x_{ij} выберем максимальный элемент $x_j = \max_{1 \le i \le m} x_{ij}$. Вектор $x = (x_1, \ldots, x_n)$ соответствует «идеальному», гипотетическому, несуществующему объекту. Для каждого из объектов $i \in X$ определим расстояние до «идеального» объекта по одной из следующих формул:

$$L_i^{(1)} = \sum_{j \in J} \left| x_j - x_{ij} \right|, \; L_i^{(2)} = \sum_{j \in J} \left(x_j - x_{ij} \right)^2,$$

$$L_i^{(3)} = \max_{j \in J} \left| x_j - x_{ij} \right|.$$

Известно, что каждое из значений $L_i^{(1)}$, $L_i^{(2)}$, $L_i^{(3)}$ определяет расстояние со всеми аксиомами. А также задано число множеств $p > 1$.

Зафиксируем в качестве расстояния одно из значений L_i и вычислим $l = \min_i L_i$, $L = \max_i L_i$ и шаг $h = \frac{L - l}{p}$. Распределим исследуемые объекты поотрезкам $\left[l + (i-1)h, l + ih \right]$, $i = \overline{1, p}$. Тогда в первый отрезок попадет объект, наиболее близкий к «идеальному», а в последний с номером p – наиболее далекий от него. Объекты, расстояния которых до «идеального», находятся в интервале с номером k, имеют рейтинг $q_k = k$ и образуют множество X_k.

Стоит отметить, что при одном и том же значении p алгоритм «идеальной точки» для метрики $L_i^{(1)}$ дает то же решение, что и алгоритм линейной свертки показателей. Действительно, $L_i^{(1)} = \sum_{j \in J} \left| x_j - x_{ij} \right| = \sum_{j \in J} x_j - L_i$, так как $x_j \ge x_{ij}$ для любого фиксированного значения j. Тогда

$$L^{(1)} = \max_{1 \le i \le m} L_i^{(1)} = \sum_{j \in J} x_j - l;$$

$$l^{(1)} = \min_{j \in J} L_i^{(1)} = \sum_{j \in J} x_j - L.$$

Следовательно, $L^{(1)} - l^{(1)} = L - l$. Поскольку длины диапазонов изменения значений L_i и $L_i^{(1)}$ совпадают, то совпадают и решения задач по обоим алгоритмам.

Алгоритм многокритериальной оптимизации

Упорядочим по невозрастанию элементы каждого из столбцов $j_1, \ldots, j_k$ матрицы показателей x_{ij}. В соответствии с этим упорядочением построим целочисленную матрицу z_{ij} номеров объектов; в каждом из указанных столбцов номера объектов будут расположены таким образом, что их характеристики не возрастают. Зададим параметры t и t_1 – максимальное и минимальное число элементов множества, соответственно. Выделим t первых строк матрицы z_{ij} и найдем множество u_1 – пересечение полученных n множеств. Если $t_1 \le |u_1|$, то вычислим для каждого $i \in u_1$ и всех $j \in \{ j_1, \ldots, j_k \}$ значения следующих параметров: $s_1(i) = \max_j x_{ij}$ и $s_2(i) = \min_j x_{ij}$. Определим функцию

$$\varphi_1(t) = \max_i \left\{ \left[s_1(i) - s_2(i) \right] / s_1(i) \right\}.$$

Найдем $\varphi_1(t_0) = \min_t \varphi_1(t)$.

Поиск минимума соответствует уменьшению максимального по всем показателям «расстояния» между объектами из множества u_1 [8]. Он выполняется путем последовательного уменьшения значений параметра t на единицу, затем формируется новое множество u_1 и т. д. Поиск минимума завершается после определения значения t_0 параметра t. Если $|u_1| < t_1$, то уменьшаем t_1 и (или) увеличиваем t и повторяем описанную процедуру. Объекты из множества u_1 образуют множество X_1, и их рейтинг $q_1 = 1$.

Далее удаляем из матрицы x_{ij} все строки, соответствующие объектам из множества X_1; упорядочим по невозрастанию элементы каждого столбца новой матрицы x_{ij}; построим новую целочисленную матрицу z_{ij} по тому же правилу, что и на первом шаге, и перейдем ко второму шагу. Аналогично построим множество $X_2 = u_2$, его рейтинг $q_2 = 2$ и т. д. Процесс завершен, если из матрицы x_{ij} удалены все строки. В процессе решения задачи определяется параметр p, который равен числу шагов алгоритма.

Описанный алгоритм основан на двух основных элементах:

– выбор по каждому из показателей объектов с максимальными их значениями;

– многократная оптимизация функции $\varphi_1(t)$ с целочисленными значениями аргумента.

Если считать, что функция $\varphi_1(t)$ является *нелинейной сверткой критериев*, то в алгоритме реализован подход, в соответствии с которым решение многокритериальной задачи сводится к решению конечной последовательности однокритериальных задач для критериальной свертки.

ПРИМЕР ИСПОЛЬЗОВАНИЯ

Рассмотрим следующую задачу. Пусть требуется выбрать лучший из пяти банков для вложения денежных средств, учитывая их оценки по трем показателям. Значения критериев (показателей) для всех альтернатив (банков) приведены в табл. 1 по четырех балльной системе оценок.

В роли объектов выступают банки (множество $S = (s_1, \ldots, s_5)$, где s_i – i-ый банк). Показателей всего три $(m = 3)$ – это расположение банка, политика банка и его репутация. Оценки поставлены экспертами. Исходная шкала у всех показателей общая – {отл, хор, удовл, неудовл}. Оценки показателей для каждого банка, записанные в порядке их перечисления, образуют его векторную оценку. Значения показателей представлены в табл. 2.

Таблица 1

Значения показателей для банков

Банк (номер по списку)	**Оценки**		
	А	Б	В
1	удовл	отл	отл
2	хор	хор	хор
3	отл	хор	удовл
4	хор	неуд	хор
5	отл	удовл	хор

Таблица 2

Значения показателей

Варианты	**Показатели**		
	K_1	K_2	K_3
s_1	3	5	5
s_2	4	4	4
s_3	5	4	3
s_4	4	2	3
s_5	5	3	4

Матрица A_{ij} имеет вид $A_{ij} = \begin{pmatrix} 3 & 5 & 5 \\ 4 & 4 & 4 \\ 5 & 4 & 3 \\ 4 & 2 & 3 \\ 5 & 3 & 4 \end{pmatrix}$.

Так как в задаче все показатели равноценны, то $w_j = \dfrac{1}{3}$. Новые значения показателей имеют вид:

$$B_{ij} = \begin{pmatrix} 1 & \frac{5}{3} & \frac{5}{3} \\ \frac{4}{3} & \frac{4}{3} & \frac{4}{3} \\ \frac{5}{3} & \frac{4}{3} & 1 \\ \frac{4}{3} & \frac{2}{3} & 1 \\ \frac{5}{3} & 1 & \frac{4}{3} \end{pmatrix}.$$

Значения показателей нормируются по формуле $b_j = \max_{1 \le i \le m} b_{ij}$:

$$b_1 = \frac{5}{3},\ b_2 = \frac{5}{3},\ b_3 = \frac{5}{3}.$$

Нормированные значения показателей b_{ij} с учетом весов w_j представим в виде матрицы

$$X_{ij} = \begin{pmatrix} \frac{3}{5} & 1 & 1 \\ \frac{4}{5} & \frac{4}{5} & \frac{4}{5} \\ 1 & \frac{4}{5} & \frac{3}{5} \\ \frac{4}{5} & \frac{2}{5} & \frac{3}{5} \\ 1 & \frac{3}{5} & \frac{4}{5} \end{pmatrix}.$$

Видно, что полученные значения x_{ij}, $i = \overline{1,5}$, $j = \overline{1,3}$ удовлетворяют условиям $0 \le x_{ij} \le 1$.

Для построения рейтинга объектов будем использовать вышеописанные алгоритмы. Рассмотрим вначале *алгоритм линейной свертки критериев*.

Пусть число множеств p задано и равно пяти. Каждому объекту $i \in X$ ставится в соответствие линейная свертка его показателей $L_i = \sum_{j \in J} x_{ij}$, вычисляемая с применением значений весовых коэффициентов w_j, $j = \overline{1,3}$:

$$L_1 = \frac{13}{5},\ L_2 = \frac{12}{5},\ L_3 = \frac{12}{5},\ L_4 = \frac{9}{5},\ L_5 = \frac{12}{5}.$$

Согласно алгоритму, $l = \min_i L_i = \frac{9}{5}$, $L = \max_i L_i = \frac{13}{5}$, $h = \frac{L - l}{p} = \frac{4}{25}$. После распределения исследуемых объектов по отрезкам, получим следующую принадлежность к кластерам (всего пять множеств-кластеров I_1, I_2, I_3, I_4, I_5):

Таблица 3

Распределение по множествам (кластерам)

Объект	**Множество**
1	I_5
2	I_4
3	I_4
4	I_1
5	I_4

Таким образом, первый объект имеет рейтинг, равный $q_5 = 5 - 5 + 1 = 1$, второй, третий и пятый объекты имеют рейтинг, равный $q_4 = 5 - 4 + 1 - 2$, а четвертый объект – рейтинг, равный $q_1 = 5 - 1 + 1 = 5$. Следовательно, используя алгоритм линейной свертки критериев, получим, что лучшим банком для вложения денежных средств является первый банк.

Теперь рассмотрим *алгоритм идеальной точки*. Пусть число множеств p также задано и равно пяти. В каждом из столбцов матрицы со значениями показателей x_{ij} выберем максимальный элемент $x_j = \max_{1 \le i \le 5} x_{ij}$. Вектор $x = (x_1, \ldots, x_3) = (1,1,1)$ соответствует «идеальному», гипотетическому, несуществующему объекту.

Для каждого из объектов $i \in X$ определим расстояние до «идеального» объекта по одной формуле: $L_i = \max_{j \in J} \left| x_j - x_{ij} \right|$:

$$L_1 = \frac{2}{5},\ L_2 = \frac{1}{5},\ L_3 = \frac{2}{5},\ L_4 = \frac{3}{5},\ L_5 = \frac{2}{5}.$$

Согласно алгоритму, $l = \min_i L_i = \frac{1}{5}$, $L = \max_i L_i = \frac{3}{5}$, $h = \frac{L - l}{p} = \frac{2}{25}$. После распределения исследуемых объектов по отрезкам, получим следующую принадлежность к кластерам (всего пять множеств-кластеров I_1, I_2, I_3, I_4, I_5):

Таблица4

Распределение по множествам (кластерам)

Объект	**Множество**
1	I_3
2	I_1
3	I_3
4	I_5
5	I_3

Таким образом первый, третий и пятый объекты имеют рейтинг, равный 3, четвертый объект имеет рейтинг, равный 5, а второй объект – рейтинг, равный 1. Следовательно, используя алгоритм идеальной точки, получим, что для вложения денежных средств больше подходит второй банк.

ЗАКЛЮЧЕНИЕ

Разработанный в работе подход может применяться при решении задач *ранжирования объектов* в различных сферах деятельности, например, определение рейтинга высших учебных заведений, определение лучшего банка для вложения денежных средств. В статье рассмотрен пример построения рейтинга банков с помощью алгоритма линейной свертки критериев и алгоритма идеальной точки. Решение динамических задач ранжирования может также быть реализовано по разработанной схеме: задача решается для каждого из дискретных моментов времени.

СПИСОК ЛИТЕРАТУРЫ

1. *Ногин В.Д.* Основы теории оптимизации / В.Д. Ногин, И.О. Протодьяконов, И.И. Евлампиев. – М. : Высшая школа, 1986. – 384 с.

2. *Подиновский В.В.* Парето-оптимальные решения многокритериальных задач / В.В. Подиновский, В.Д. Ногин. – М. : Физматлит, 2007. – 256 с.

3. *Подиновский В.В.* Аксиоматическое решение проблемы оценки важности критериев в многокритериальных задачах / В.В. Подиновский // Современное состояние теории исследования операций. – М. : Наука, 1979. – С. 117–149.

4. *Подиновский В.В.* Количественная важность критериев / В.В. Подиновский // Автоматика и телемеханика. – № 5. – 2000. – С. 110–123.

5. *Шахнов И.Ф.* Количественная оценка важности целей / И.Ф. Шахнов // Известия РАН. Теория и системы управления. – № 1. – 2003. – С. 78–86.

6. *Мандель И.Д.* Кластерный анализ / И.Д. Мандель. – М. : Финансы и статистика, 1988.

7. *Барсегян А.А.* Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов и [др.]. – СПб. : БХВ-Петербург, 2009. – 512 с.

8. *Сигал И.Х.* Введение в прикладное дискретное программирование / И.Х. Сигал, А.П. Иванова. – М. : Физматлит, 2007. – 304 с.

Аристова Екатерина Михайловна – к.ф.-м.н., преподаватель факультета ПММ каф. Вычислительной математики и прикладных информационных технологий, Воронежский госуниверситет. Тел.: 8-920-453-17-72.
E-mail: pmim@yandex.ru.

Aristova Ekaterina Michailovna – phd, lecturer of the Faculty of Applied Mathematics, Informatics and Mechanics of the Computer Science and Applied Information Technologies Dept., Voronezh State University. Phone: 8-920-453-17-72.
E-mail: pmim@yandex.ru.