УДК 519.81

Е.М. АРИСТОВА
E.M. ARISTOVA

АЛГОРИТМ ИДЕАЛЬНОЙ ТОЧКИ ДЛЯ ПОСТРОЕНИЯ РЕЙТИНГА ОБЪЕКТОВ
THE «IDEAL POINT» ALGORITHM FOR RATING OBJECT'S CREATION

В статье рассматривается алгоритм идеальной точки, который используется для определения рейтинга объектов на примере выбора наилучшего туристического агентства из представленного списка.

Ключевые слова: рейтинг объектов, весовые коэффициенты, задача кластеризации, ранжирование объектов, алгоритм «идеальной точки» для построения рейтинга объектов, лучшее туристическое агентство.

This article considers the «ideal point» algorithm for creation of object's rating. The problem of the best travel agency choice is considered and solved by this algorithm.

Key words: rating facilities, weights, clustering problem, ranging of objects, the «ideal point» algorithm to determine the rating of objects, the best travel agency.

ВВЕДЕНИЕ

Слово «рейтинг» (rating) означает, дословно, «оценка». Данный термин, в основном, использовался в теории вероятностей и математической статистике. Однако и в сфере экономической деятельности возникает необходимость использования величин (параметров, признаков), имеющих предположительную численную природу, но конкретные значения таких величин не поддаются прямому физическому измерению. В таких случаях прибегают к экспертным оценкам, предусматривающих, что некая группа лиц (экспертов) дает заключение о характере распределения оцениваемой величины по некоторой шкале числовых значений.

Одной из наиболее популярных форм предоставления информации являются рейтинги, которые характеризуют относительную значимость того или иного объекта или явления по сравнению с другими аналогичными объектами или явлениями. Среди тех сфер, где наблюдается стремительное развитие индустрии производства информации, не последнее место занимает рынок туристических услуг.

Город Воронеж представлен большим многообразием туристических агентств. Рассмотрим некоторые из них:

1) «Воронежинтур» (г. Воронеж, ул. Пушкинская, д.28);
2) «Семь морей» (г. Воронеж, ул. Куколкина, д.14);
3) «Анна Тур» (г. Воронеж, ул. Куколкина, д.32);
4) «Ален» (г. Воронеж, ул. Кирова, д.1, пр. Московский, д.82);
5) «Дик Тур» (г. Воронеж, ул. Театральная, 28);
6) «ТВА-экспресс» (г. Воронеж, ул. Фр. Энгельса, д.25б);
7) «Тез Тур» (г. Воронеж, ул. Никитинская, д.8а).

Рассмотрим задачу: Необходимо выбрать лучшее туристическое агентство из представленного списка для покупки тура для путешествия, исходя из предложенных оценок по 10 показателям.

Оценки выставлены экспертами на основе данных полученных от посетителей представленных туристических агентств.

Поставленную задачу будем рассматривать как задачу определения рейтинга объектов. Остановимся на ней более подробно [1].

ПОСТАНОВКА ЗАДАЧИ ОПРЕДЕЛЕНИЯ РЕЙТИНГА ОБЪЕКТОВ

Рейтинг – 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 [1].

Пусть для каждого из m объектов множества $X = \{1,2,..,m\}$ 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.

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 [2,5,8].

Аксиоматическое решение проблемы оценки важности в многокритериальных задачах предложено в [6]. 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 [2,7] 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 $J = \{j_1, j_2, .., j_k\}$ различных показателей этого набора. С формальной точки зрения представленная задача – это задача кластеризации [4] 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 [3].

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

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.

Для определения рейтинга объекта рассмотрим алгоритм идеальной точки [1,2].

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 $x = (x_1, .. x_n)$ соответствует «идеальному», гипотетическому, несуществующему объекту. Для каждого из объектов $i \in X$ определим расстояние до «идеального» объекта по одной из следующих формул:

$$L_i^{(1)} = \sum_{j\in J} |x_j - x_{ij}|, L_i^{(2)} = \sum_{j\in J} (x_j - x_{ij})^2, L_i^{(3)} = \max_{j\in J} |x_j - x_{ij}|.$$

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 по p отрезкам $[l + (i-1)h, l + ih]$, $i = \overline{1,p}$. 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.

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Теперь применим алгоритм идеальной точки для решения задачи о выборе наилучшего туристического агентства.

В роли объектов выступают туристические агентства (множество $S = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$, где S_i – i-ое турагентство ($i = \overline{1,7}$).

Значения критериев (показателей) для всех альтернатив (туристических агентств) приведены в таблице 1 по трехбалльной системе оценок. Исходная шкала у всех показателей общая – {отл, хор, удовл}.

Таблица 1 – Значения показателей для представленных туристических агентств[1]

Турагентство (№ по списку)	**Оценки показателей**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	отл	хор	хор	отл	удовл	удовл	хор	удовл	хор	хор
2	хор	отл	хор	отл	хор	хор	удовл	удовл	хор	хор
3	хор	отл	отл	отл	хор	хор	удовл	удовл	хор	отл
4	отл	хор	отл	хор	хор	хор	хор	удовл	удовл	хор
5	хор	хор	отл	отл	хор	хор	хор	хор	хор	хор
6	отл	хор	хор	отл	хор	хор	хор	хор	удовл	хор
7	хор	отл	хор	отл	отл	отл	отл	отл	хор	хор

Показатели представлены множеством

$$K = \{K_1, K_2, K_3, K_4, K_5, K_6, K_7, K_8, K_9, K_{10}\},$$

где K_1 – месторасположение турагентства; K_2 – внешний вид офиса турагентства; K_3 – комфортность помещения; K_4 – количество консультантов; K_5 – вежливость консультантов; K_6 – стрессоустойчивость консультантов; K_7 – компетентность лица, с которым проводилась

[1] Данные представлены журналом De Facto.

беседа; K_8 – понимание запросов клиента; K_9 – скорость обслуживания; K_{10} – дополнительный сервис.

Оценки показателей для каждого турагентства, записанные в порядке их перечисления, образуют его векторную оценку. Значения показателей $K_j, j = \overline{1,10}$ представлены в таблице 2.

Таблица 2 – Значения показателей

Варианты	**Оценки показателей**									
	K_1	K_2	K_3	K_4	K_5	K_6	K_7	K_8	K_9	K_{10}
S_1	5	4	4	5	3	3	4	3	4	4
S_2	4	5	4	5	4	4	3	3	4	4
S_3	4	5	5	5	4	4	3	3	4	5
S_4	5	4	5	4	4	4	4	3	3	4
S_5	4	4	5	5	4	4	4	4	4	4
S_6	5	4	4	5	4	4	4	4	3	4
S_7	4	5	4	5	5	5	5	5	4	4

Матрица A_{ij} имеет вид:

$$A_{ij} = \begin{pmatrix} 5 & 4 & 4 & 5 & 3 & 3 & 4 & 3 & 4 & 4 \\ 4 & 5 & 4 & 5 & 4 & 4 & 3 & 3 & 4 & 4 \\ 4 & 5 & 5 & 5 & 4 & 4 & 3 & 3 & 4 & 5 \\ 5 & 4 & 5 & 4 & 4 & 4 & 4 & 3 & 3 & 4 \\ 4 & 4 & 5 & 5 & 4 & 4 & 4 & 4 & 4 & 4 \\ 5 & 4 & 4 & 5 & 4 & 4 & 4 & 4 & 3 & 4 \\ 4 & 5 & 4 & 5 & 5 & 5 & 5 & 5 & 4 & 4 \end{pmatrix}.$$

Так как в задаче все показатели равноценны, то $w_j = \frac{1}{10}, j = \overline{1,10}$. Новые значения показателей имеют вид:

$$B_{ij} = \begin{pmatrix} 0.5 & 0.4 & 0.4 & 0.5 & 0.3 & 0.3 & 0.4 & 0.3 & 0.4 & 0.4 \\ 0.4 & 0.5 & 0.4 & 0.5 & 0.4 & 0.4 & 0.3 & 0.3 & 0.4 & 0.4 \\ 0.4 & 0.5 & 0.5 & 0.5 & 0.4 & 0.4 & 0.3 & 0.3 & 0.4 & 0.5 \\ 0.5 & 0.4 & 0.5 & 0.4 & 0.4 & 0.4 & 0.4 & 0.3 & 0.3 & 0.4 \\ 0.4 & 0.4 & 0.5 & 0.5 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 & 0.4 \\ 0.5 & 0.4 & 0.4 & 0.5 & 0.4 & 0.4 & 0.4 & 0.4 & 0.3 & 0.4 \\ 0.4 & 0.5 & 0.4 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.4 & 0.4 \end{pmatrix}.$$

Значения показателей нормируются по формуле $b_j = \max_{1 \le i \le 7} b_{ij}$:

$b_1 = 0.5, b_2 = 0.5, b_3 = 0.5, b_4 = 0.5, b_5 = 0.5, b_6 = 0.5, b_7 = 0.5, b_8 = 0.5, b_9 = 0.4, b_{10} = 0.5.$

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

$$X_{ij} = \begin{pmatrix} 1.0 & 0.8 & 0.8 & 1.0 & 0.6 & 0.6 & 0.8 & 0.6 & 1.00 & 0.8 \\ 0.8 & 1.0 & 0.8 & 1.0 & 0.8 & 0.8 & 0.6 & 0.6 & 1.00 & 0.8 \\ 0.8 & 1.0 & 1.0 & 1.0 & 0.8 & 0.8 & 0.6 & 0.6 & 1.00 & 1.0 \\ 1.0 & 0.8 & 1.0 & 0.8 & 0.8 & 0.8 & 0.8 & 0.6 & 0.75 & 0.8 \\ 0.8 & 0.8 & 1.0 & 1.0 & 0.8 & 0.8 & 0.8 & 0.8 & 1.00 & 0.8 \\ 1.0 & 0.8 & 0.8 & 1.0 & 0.8 & 0.8 & 0.8 & 0.8 & 0.75 & 0.8 \\ 0.8 & 1.0 & 0.8 & 1.0 & 1.0 & 1.0 & 1.0 & 1.0 & 1.00 & 0.8 \end{pmatrix}.$$

Видно, что полученные значения $x_{ij}, i = \overline{1,7}, j = \overline{1,10}$ удовлетворяют условиям $0 \le x_{ij} \le 1$.

Для построения рейтинга объектов используем *алгоритм идеальной точки*. Пусть число множеств p задано и равно семи. В каждом из столбцов матрицы со значениями показателей x_{ij} выберем максимальный элемент $x_j = \max_{1 \le i \le 7} x_{ij}$. Вектор $x = (x_1, .. x_{10}) = (1, \dots, 1)$ соответствует «идеальному», гипотетическому, несуществующему объекту.

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

$$L_1 = 0.4, L_2 = 0.4, L_3 = 0.4, L_4 = 0.4, L_5 = 0.2, L_6 = 0.25, L_7 = 0.2.$$

Согласно алгоритму, $l = \min_i L_i = 0.2, \mathrm{L} = \max_i L_i = 0.4, h = \frac{L-l}{p} = \frac{1}{35}$. После распределения исследуемых объектов по отрезкам, получим следующую принадлежность к кластерам (всего семь множеств-кластеров $I_1, I_2, I_3, I_4, I_5, I_6, I_7$) (таблица 3):

Таблица 3 – Распределение по множествам (кластерам)

Объект	**Множество**
1	I_7
2	I_7
3	I_7
4	I_7
5	I_1
6	I_2
7	I_1

Т.о. первый, второй, третий и четвертый объекты имеют рейтинг, равный 7; пятый и седьмой объекты имеют рейтинг, равный 1; а шестой объект – рейтинг, равный 2.

Следовательно, используя алгоритм идеальной точки с заданным расстоянием, получим, что, по мнению респондентов, лучшими турагентствами из представленных являются пятое и седьмое, а именно – «Дик Тур» и «Тез Тур».

Теперь для каждого из объектов $i \in X$ определим расстояние до «идеального» объекта по формуле: $L_i = \sum_{j \in J} (x_j - x_{ij})^2$:

$$L_1 = 0.64, L_2 = 0.52, L_3 = 0.44, L_4 = 0.4625, L_5 = 0.28, L_6 = 0.3425, L_7 = 0.12.$$

Согласно алгоритму, $l = \min_i L_i = 0.12, \mathrm{L} = \max_i L_i = 0.64, h = \frac{L-l}{p} = \frac{13}{175}$. После распределения исследуемых объектов по отрезкам, получим следующую принадлежность к кластерам (всего семь множеств-кластеров $I_1, I_2, I_3, I_4, I_5, I_6, I_7$) (таблица 4):

Таблица 4 – Распределение по множествам (кластерам)

Объект	**Множество**
1	I_7
2	I_6
3	I_5
4	I_5
5	I_3
6	I_3
7	I_1

Т.о. первый имеет рейтинг, равный 7; второй объект имеет рейтинг, равный 6; третий и четвертый объекты имеют рейтинг, равный 5; пятый и шестой объекты имеют рейтинг, равный 3; а седьмой объект – рейтинг, равный 1.

Следовательно, используя алгоритм идеальной точки с заданным расстоянием, получим, что, по мнению респондентов, лучшим турагентством из представленных является «Тез Тур».

ЗАКЛЮЧЕНИЕ

Разработанный в работе подход может применяться при решении задач *ранжирования объектов* в различных сферах деятельности. В данной статье он использовался для выбора наилучшего туристического агентства города Воронежа из списка представленных турагентств.

Решение динамических задач ранжирования может также быть реализовано по разработанной схеме: задача решается для каждого из дискретных моментов времени.

СПИСОК ЛИТЕРАТУРЫ

1. Аристова Е.М. Определение рейтинга объектов / Е.М. Аристова // Вестник Воронежского государственного университета. Серия Системный анализ и информационные технологии. – Воронеж: ВГУ, 2014. – №2. – С.51-56.

2. Аристова Е.М. Учет взаимодействия между целевыми функциями и их агрегирование в задачах оптимизации // Диссертация на соискание ученой степени кандидата физ.-мат. наук по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ. – Воронеж: ВГУ, 2012. – 152 с.

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

4. Мандель И.Д. Кластерный анализ / И.Д. Мандель. – М.: Финансы и статистика. – 1988. – 176 с.

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

6. 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.

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

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

Аристова Екатерина Михайловна

Воронежский государственный университет, г. Воронеж

К.ф.-м.н., доцент, преподаватель, доцент кафедры «Вычислительная математика и прикладные информационные технологии»

Тел.: +7(904) 210-12-60

E-mail: pmim@yandex.ru