Моделирование и анализ данных
2022. Том 12. № 2. С. 5–19
DOI: https://doi.org/10.17759/mda.2022120201
ISSN: 2219-3758 (печатный)
ISSN: 2311-9454 (online)

Modelling and Data Analysis
2022. Vol. 12, no. 2, pp. 5–19
DOI: https://doi.org/10.17759/mda.2022120201
ISSN: 2219-3758 (print)
ISSN: 2311-9454 (online)

◇◇◇◇◇◇◇◇◇◇◇◇АНАЛИЗ ДАННЫХ◇◇◇◇◇◇◇◇◇◇◇◇

УДК 519.858

Анализ больших данных в задачах многокритериального выбора

Иванов А.А.*
Московский Авиационный Институт (НИУ МАИ)
г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0003-4433-6449
e-mail: ivanov17andrey@gmail.com

Яшина Н.П.**
Московский Авиационный Институт (НИУ МАИ)
г. Москва, Российская Федерация
ORCID: https://orcid.org/0000-0002-8401-0315
e-mail: nina_p_yashina@mail.ru

Рассматривается задача многокритериального выбора с неоднородными шкалами критериев. Описана модель задачи многокритериального выбора, основными элементами которой являются множества альтернатив и критериев качества, а также алгоритмы, позволяющие ранжировать альтернативы без предварительного приведение шкал критериев к однородным. Алгоритмы нахождения агрегированного ранжирования альтернатив основаны на построении матриц предпочтений, содержащих информацию по критериям о степени превосходства одной альтернативы над другой. Доказаны утверждения, позволяющие ранжировать альтернативы с оценками по двум критериям качества. Предложены алгоритмы индексирования альтернатив для их ранжирования при произвольном числе критериев. Наилучшее агрегированное ранжирование определяется по суммарному расстоянию до ранжирований альтернатив по критериям. Все алгоритмы имеют полиномиальную

Для цитаты:
Иванов А.А., Яшина Н.П. Анализ больших данных в задачах многокритериального выбора // Моделирование и анализ данных. 2022. Том 12. № 2. С. 5–19. DOI: https://doi.org/10.17759/mda.2022120201

***Иванов Андрей Алексеевич,** магистрант, Московский авиационный институт (НИУ МАИ), г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0003-4433-6449, e-mail: ivanov17andrey@gmail.com

****Яшина Нина Павловна,** кандидат физико-математических наук, доцент, кафедра 805 «Математическая кибернетика», Московский авиационный институт (НИУ МАИ), г. Москва, Российская Федерация, ORCID: https://orcid.org/0000-0002-8401-0315, e-mail: nina_p_yashina@mail.ru

вычислительную сложность, что позволяет работать с большими массивами исходной информации. Разработана программная система ранжирования альтернатив в задачах с большими данными. Исходная информация хранится в Excel таблицах, что позволяет легко учитывать ограничения по шкалам критериев. Организация работы программной системы продемонстрирована на примере выбора наилучшего варианта дрона для приобретения с целью наблюдения за местностью, ее съемок и передачи информации оператору.

Ключевые слова: принятие решений, критерий качества, анализ больших данных, матрица предпочтений, ранжирование альтернатив.

1. ВВЕДЕНИЕ

До недавнего времени считалось, что проблема выбора из предложенных альтернатив лежит в области психологии, напрямую зависит от индивидуальных способностей человека. В последние годы возможность получения и необходимость обработки большого объема исходной информации не позволяют осуществить выбор, основываясь только на интуиции лица, принимающего решения (ЛПР). Появилась потребность в разработке математически обоснованных методов и алгоритмов, позволяющих доверить процесс выбора оптимальных вариантов программной системе.

При принятии решений большинство задач обычно являются многокритериальными: рассматриваемые альтернативы оцениваются по критериям качества. Методы выбора и упорядочения альтернатив в таких задачах в значительной степени зависят от шкал критериев. Традиционные алгоритмы выбора оптимальных вариантов решений основываются на задании информации в количественных шкалах, т.е. численными оценками альтернатив по критериям качества, при этом обычно требуется, чтобы шкалы критериев были однородными. Но, практические задачи с однородными шкалами встречаются крайне редко, поэтому приходится использовать дополнительные процедуры приведения оценок по критериям к единой шкале [1].

Данная статья основывается на алгоритмах, не требующих предварительного приведения шкал критериев к однородным [2,3]. Все алгоритмы имеют полиномиальную вычислительную сложность, что позволяет анализировать большие объемы исходной информации.

В разработанную программную систему данные от ЛПР поступают через Excel таблицы. Это простой и удобный способ ввода информации для обычных пользователей в случае больших данных [4]. Информация хранится на удаленном сервере, что позволяет зарегистрированному пользователю всегда иметь доступ для её корректировки, причем с различных устройств. Работа программной системы продемонстрирована на примере выбора наиболее предпочтительных для ЛПР моделей дронов.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЗАДАЧИ МНОГОКРИТЕРИАЛЬНОГО ВЫБОРА

Для решения задачи необходимо сформировать математическую модель, полностью отражающую разнообразие информации и методов принятия решений. Будем рассматривать математическую модель многокритериального выбора, содержащую следующие элементы:

$$< t,\ A,\ K,\ X,\ F,\ \varphi,\ I,\ r >,$$

где t – постановка задачи (ранжирование, выбор наилучших альтернатив);

$A = \{a_1,\ a_2, \dots, a_n\}$ – множество рассматриваемых альтернатив;

$K = \{K_1,\ K_2, \dots, K_m\}$ – множество критериев;

$X = \{X_1, X_2, \dots, X_m\}$ – множество шкал критериев;

$F = \{F_1,\ F_2, \dots, F_m\}$ – предпочтения по шкалам критериев;

$\varphi -$ отображение $A \to (X_1 \times\ X_2 \times \dots \times X_m)$

I – информация о важности критериев;

r – решающее правило (последовательность алгоритмов непротиворечивого агрегирования предпочтений).

В зависимости от содержательной постановки задачи t требуется построить агрегированное отношение предпочтения на множестве альтернатив $A = \{a_1,\ a_2, \dots, a_n\}$, провести ранжирование альтернатив и\или осуществить выбор наилучших вариантов альтернатив. Множество альтернатив A задается ЛПР. Альтернативы оцениваются по критериям качества $K = \{K_1,\ K_2, \dots, K_m\}$ с порядковыми или числовыми шкалами $X = \{X_1,\ X_2, \dots, X_m\}$. Таким образом, каждой альтернативе ставится в соответствие векторная оценка из множества $(X_1 \times\ X_2 \times \dots \times X_m\ );\ (\varphi\colon A \to (X_1 \times\ X_2 \times \dots \times X_m))$. Значения оценок по числовым шкалам критериев минимизируются или максимизируются. Так стоимость альтернативы обычно минимизируется, а скорость – максимизируется. Существуют шкалы, у которых наилучшие и худшие оценки зависят от предпочтений ЛПР, например, месторасположение альтернативы. Предпочтения по шкалам критериев $F = \{F_1,\ F_2, \dots, F_m\}$ могут быть заданы численными оценками альтернатив по данному критерию, попарным сравнением (в частности ранжированием альтернатив), т.е. отношениями предпочтения $\rho_1,\ \rho_2, \dots, \rho_m$, выбором множества наилучших альтернатив, вербальными оценками. В случае задания предпочтений бинарными отношениями оценок альтернатив по шкалам критериев не существует.

Информация I о важности критериев $K_1,\ K_2, \dots, K_m$ может быть задана отношением предпочтения на множестве критериев (относительная важность критериев) или весовыми коэффициентами $k = \{k_1,\ k_2, \dots, k_m\}$, численно выражающими важность критериев. Относительная важность критериев может быть дополнена информацией о том, во сколько раз один критерий лучше другого. Коэффициенты важности могут быть заданы ЛПР или вычислены в результате работы одной из специально разработанных процедур. Если критерии обладают одинаковой важностью, то все весовые коэффициенты равны единице.

Под решающим правилом r понимается последовательность алгоритмов, приводящая к решению задачи. В первую очередь решающее правило содержит алгоритмы многокритериального выбора, а также вспомогательные алгоритмы обработки информации. При формировании решающего правила r учитывается тип входной информации и требования ЛПР к выходной информации. В связи с этим требуются процедуры, позволяющие выявить и формализовать в удобном для работы основных алгоритмов виде входную информацию. В ряде методов могут потребоваться дополнительные процедуры приведение шкал критериев к единой шкале, а на заключительном этапе решения задачи алгоритмы для упорядочения альтернатив и выбора наилучших из них. Решающее правило также может содержать алгоритмы нахождения коэффициентов важности критериев.

В данной статье предлагаются алгоритмы многокритериального выбора, основанные на задании численных оценок альтернатив по критериям, но не требующие приведения шкал критериев к однородным. Все алгоритмы полиномиальные, что позволяет анализировать большое число альтернатив.

Пусть альтернативы из множества $A = \{a_1, a_2, \dots, a_n\}$ оцениваются по критериям качества $K = \{K_1, K_2, \dots, K_m\}$ с числовыми шкалами отношений $X = \{X_1, X_2, \dots, X_m\}$. Каждой альтернативе из множества $A = \{a_1, a_2, \dots, a_n\}$ ставится в соответствие вектор оценок по критериям. Требуется ранжировать альтернативы, т.е. упорядочить их по предпочтению.

Каждому критерию K_t $(t = 1, \dots, m)$ поставим в соответствие вектор оценок альтернатив $x^t =< x_1^t, x_2^t, \dots, x_n^t >, x_i^t \in \mathbb{R}^+$.

Определение 1. Матрицей предпочтений $P^t = \|p_{ij}^t\|$ $(t = 1, \dots, m)$ называется квадратная матрица порядка n (n – число альтернатив, $i, j = 1, \dots, n$) с элементами

$$p_{ij}^t = \begin{cases} \dfrac{x_i^t}{x_i^t + x_j^t}, \text{ если значения оценок по шкале } X_t \text{ максимизируются,} \\ \dfrac{x_j^t}{x_i^t + x_j^t}, \text{ если значения оценок по шкале } X_t \text{ минимизируются.} \end{cases} \quad (1)$$

Элемент матрицы p_{ij}^t показывает, что по критерию K_t альтернатива a_i более предпочтительна, чем a_j в p_{ij}^t раз.

Для элементов матрицы предпочтений P^t выполняется:

$$1)\ \frac{p_{ij}^t}{p_{ji}^t} = \begin{cases} \dfrac{x_i^t}{x_j^t}, \text{ если значения оценок по шкале } X_t \text{ максимизируются,} \\ \dfrac{x_j^t}{x_i^t}, \text{ если значения оценок по шкале } X_t \text{ минимизируются,} \end{cases}$$

– сохраняется информация о том, во сколько раз альтернатива a_i предпочтительнее альтернативы a_j;

$$2)\ p_{ij}^t + p_{ji}^t = 1\ (i, j = 1, \dots, n).$$

Эти свойства фактически позволяют заменить процедуру приведения шкал критериев к однородным.

Таким образом, на входе алгоритма, позволяющего построить агрегированное предпочтение, должны быть заданы матрицы предпочтений по критериям $P^1, P^2, \dots, P^m$.

При нахождения агрегированного отношения предпочтения в зависимости от пожеланий ЛПР можно использовать алгоритмы, основывающиеся на суммировании матриц исходных предпочтений. Можно также построить агрегированное отношение, минимально удаленное от исходных. Опишем подробно алгоритмы, рассматриваемые в данной статье.

3. АЛГОРИТМ ПОСТРОЕНИЯ АГРЕГИРОВАННОГО РАНЖИРОВАНИЯ АЛЬТЕРНАТИВ ДЛЯ ДВУХ КРИТЕРИЕВ

Рассмотрим случай, когда оценка альтернатив проводится по двум критериям качества, имеющим равную важность.

Просуммировав матрицы предпочтений по критериям, получим матрицу суммарных предпочтений $P_\Sigma = \left\| p_{ij}^{\Sigma} \right\|$.

Утверждение 1. Для альтернатив, оцениваемых по двум критериям качества, оценки по которым являются положительными и максимизируются, предпочтительнее альтернатива с большим произведением компонент векторной оценки.

Доказательство утверждения 1. Пусть оценки по шкалам критериев K_1 и K_2 максимизируются. Альтернатива a_1 имеет векторную оценку по критериям $< x_1, y_1 >$, а альтернатива a_2 – $< x_2, y_2 >$ ($x_1, y_1, x_2, y_2 > 0$). Тогда матрицы предпочтений по критериям K_1 и K_2 имеют вид:

$$P^1 = \begin{pmatrix} \frac{1}{2} & \frac{x_1}{x_1 + x_2} \\ \frac{x_2}{x_1 + x_2} & \frac{1}{2} \end{pmatrix}, \quad P^2 = \begin{pmatrix} \frac{1}{2} & \frac{y_1}{y_1 + y_2} \\ \frac{y_2}{y_1 + y_2} & \frac{1}{2} \end{pmatrix}$$

Найдем матрицу суммарных предпочтений:

$$P_\Sigma = \begin{pmatrix} 1 & \frac{x_1}{x_1 + x_2} + \frac{y_1}{y_1 + y_2} \\ \frac{x_2}{x_1 + x_2} + \frac{y_2}{y_1 + y_2} & 1 \end{pmatrix}.$$

Сравним числа:

$$\frac{x_1}{x_1 + x_2} + \frac{y_1}{y_1 + y_2} \gtreqless \frac{x_2}{x_1 + x_2} + \frac{y_2}{y_1 + y_2} \Leftrightarrow$$

(так как все оценки положительные)

$$x_1 y_1 + x_1 y_2 + x_1 y_1 + x_2 y_1 \Leftrightarrow \gtreqless x_2 y_1 + x_2 y_2 + x_1 y_2 + x_2 y_2 \ \Leftrightarrow x_1 y_1 \gtreqless x_2 y_2.$$

Получаем, для двух критериев лучшей будет векторная оценка с большим значением произведения ее компонент.

Утверждение 2. Для альтернатив, оцениваемых по двум критериям качества, оценки по которым являются положительными и минимизируются, предпочтительнее альтернатива с меньшим произведением компонент векторной оценки.

Доказательство утверждения 2. Пусть оценки по шкалам критериев K_1 и K_2 минимизируются. Альтернатива a_1 имеет векторную оценку по критериям $< x_1, y_1 >$, а альтернатива a_2 – $< x_2, y_2 >$ $(x_1, y_1, x_2, y_2 > 0)$. Тогда матрицы предпочтений по критериям K_1 и K_2 имеют вид:

$$P^1 = \begin{pmatrix} \frac{1}{2} & \frac{x_2}{x_1 + x_2} \\ \frac{x_1}{x_1 + x_2} & \frac{1}{2} \end{pmatrix}, \quad P^2 = \begin{pmatrix} \frac{1}{2} & \frac{y_2}{y_1 + y_2} \\ \frac{y_1}{y_1 + y_2} & \frac{1}{2} \end{pmatrix}$$

Найдем матрицу суммарных предпочтений:

$$P_\Sigma = \begin{pmatrix} 1 & \frac{x_2}{x_1 + x_2} + \frac{y_2}{y_1 + y_2} \\ \frac{x_1}{x_1 + x_2} + \frac{y_1}{y_1 + y_2} & 1 \end{pmatrix}.$$

Сравним числа:

$$\frac{x_2}{x_1 + x_2} + \frac{y_2}{y_1 + y_2} \gtreqless \frac{x_1}{x_1 + x_2} + \frac{y_1}{y_1 + y_2} \Leftrightarrow$$

$$x_2 y_1 + x_2 y_2 + x_1 y_2 + x_2 y_2 \Leftrightarrow \gtreqless x_1 y_1 + x_1 y_2 + x_1 y_1 + x_2 y_1 \Leftrightarrow x_2 y_2 \gtreqless x_1 y_1.$$

Получаем, для двух критериев, для которых оценки по шкалам минимизируются, лучшей будет векторная оценка с меньшим значением произведения ее компонент.

Утверждение 3. Пусть для альтернатив, оцениваемых по двум критериям качества с положительными шкалами, оценки по критерию K_1 максимизируются, а по K_2, минимизируются. Тогда предпочтительнее альтернатива с большим отношением оценки по критерию K_1 к оценке по критерию K_2.

Доказательство утверждения 3. Пусть альтернатива a_1 имеет вектор оценок по критериям $< x_1, y_1 >$, а альтернатива a_2 – $< x_2, y_2 >$ $(x_1, y_1, x_2, y_2 > 0)$. Найдем матрицы предпочтений по критериям K_1 и K_2 .

По шкале критерия K_1 оценки максимизируются, следовательно,

$$P^1 = \begin{pmatrix} \frac{1}{2} & \frac{x_1}{x_1 + x_2} \\ \frac{x_2}{x_1 + x_2} & \frac{1}{2} \end{pmatrix}.$$

По шкале критерия K_2 оценки минимизируются, следовательно,

$$P^2 = \begin{pmatrix} \frac{1}{2} & \frac{y_2}{y_1 + y_2} \\ \frac{y_1}{y_1 + y_2} & \frac{1}{2} \end{pmatrix}.$$

Матрица суммарных предпочтений имеет вид:

$$P_{\Sigma} = \begin{pmatrix} 1 & \dfrac{x_1}{x_1 + x_2} + \dfrac{y_2}{y_1 + y_2} \\ \dfrac{x_2}{x_1 + x_2} + \dfrac{y_1}{y_1 + y_2} & 1 \end{pmatrix}.$$

Сравним ее элементы:

$$\frac{x_1}{x_1 + x_2} + \frac{y_2}{y_1 + y_2} \gtreqless \frac{x_2}{x_1 + x_2} + \frac{y_1}{y_1 + y_2} \Leftrightarrow$$

$$x_1 y_1 + x_1 y_2 + x_1 y_2 + x_2 y_2 \gtreqless x_2 y_1 + x_2 y_2 + x_1 y_1 + x_2 y_1 \Leftrightarrow$$

$$x_1 y_2 \gtreqless x_2 y_1 \Leftrightarrow \frac{x_1}{y_1} \gtreqless \frac{x_2}{y_2}.$$

Получаем: предпочтительнее альтернатива, у векторной оценки которой большее значение отношения первой компоненты (оценки максимизируются) ко второй (оценки максимизируются). Что и требовалось доказать.

Теорема 1. При сравнении векторных оценок альтернатив по двум критериям качества с положительными шкалами получим ранжирование альтернатив, т.е. все альтернативы попарно сравнимы.

Доказательство теоремы 1 следует из утверждений 1–3, в которых векторным оценкам ставится в соответствие число, что позволяет сравнить все альтернативы по предпочтительности. Заметим, что в случае равенства чисел альтернативы имеют равную предпочтительность.

Пример 1. Пусть множество альтернатив $A = \{a_1, a_2, a_3, a_4\}$ оценивается по двум критериям качества K_1, K_2. Численные оценки альтернатив приведены в таблице.

	K_1	K_2	**Произведение оценок**	**Отношение оценок**
a_1	3	6	18	$\frac{1}{2}$
a_2	4	5	20	$\frac{4}{5}$
a_3	2	7	14	$\frac{2}{7}$
a_4	7	3	21	$\frac{7}{3}$

Если значения оценок по шкалам критериев максимизируются, то по утверждению 1 наилучшая альтернатива a_2.

Если значения оценок по шкалам критериев минимизируются, то по утверждению 2 наилучшая альтернатива a_3.

Если значения оценок по шкале критерия K_1 максимизируются, а по шкале критерия K_2 минимизируются, то по утверждению 3 наилучшая альтернатива a_4.

4. АЛГОРИТМЫ ПОСТРОЕНИЯ АГРЕГИРОВАННОГО РАНЖИРОВАНИЯ АЛЬТЕРНАТИВ ДЛЯ ПРОИЗВОЛЬНОГО ЧИСЛА КРИТЕРИЕВ

Опишем алгоритмы построения агрегированного отношения предпочтения для произвольного числа альтернатив, основывающиеся на суммировании матриц исходных предпочтений.

Алгоритм ранжирования альтернатив на основе орграфа суммарного предпочтения.

1. Формирование матриц предпочтения P^1, P^2, ... P^m по критериям по формулам (1).
2. Построение матрицы суммарных предпочтений.
3. $P_\Sigma = \sum_{t=1}^{m} P^t$.
4. Ранжирование альтернатив.

Предложенный алгоритм допускает введение весовых коэффициентов $k_1, k_2, \dots, k_m$, характеризующих компетентность экспертов или важность критериев качества. В этом случае матрица суммарных предпочтений будет вычисляться по формуле:

$$P_\Sigma = \sum_{t=1}^{m} k_t P^t.$$

Ранжирование альтернатив по одному критерию качества осуществляет Excel таблица. Алгоритмы ранжирования в случае задания двух критериев описаны в предыдущем параграфе. Для ранжирования альтернатив при большем числе критериев воспользуемся алгоритмами, основанными на индексирования альтернатив [5].

Опишем алгоритмы нахождения индексов альтернатив. Будем рассматривать следующие процедуры.

Процедура разности весов: каждой альтернативе $a_i \in A$ ставится в соответствие индекс, равный разности суммы элементов строки матрицы суммарных предпочтений $P_\Sigma = \left\| p_{ij}^\Sigma \right\|$ и элементов столбца этой матрицы.

$$I_i^{DW} = \sum_{j=1}^{m} (p_{ij}^\Sigma - p_{ji}^\Sigma)] \tag{2}$$

Процедура отношения весов: каждой альтернативе $a_i \in A$ ставится в соответствие индекс, равный отношению суммы элементов строки матрицы суммарных предпочтений $P_\Sigma = \left\| p_{ij}^\Sigma \right\|$ и элементов столбца этой матрицы.

$$I_i^{RW} = \frac{\sum_{j=1}^{m} p_{ij}^\Sigma}{\sum_{j=1}^{m} p_{ji}^\Sigma} \tag{3}$$

Если знаменатель дроби равен нулю, то $I_i^{RW} = \infty$. При программной реализации знак бесконечности заменяется числом, заведомо большим индексов всех альтернатив, индекс которых не равен бесконечности. В случае, когда числитель и знаменатель равен нулю можно рекомендовать поставить альтернативу a_i в ранжировании на последнее место, т.к. информация о сравнении ее с другими альтернативами отсутствует. Мы будем решать задачи с полной информацией об альтернативах.

Преобразуем формулу (2) с учетом особенностей матриц предпочтений по критериям и матрицы суммарных предпочтений в рассматриваемом случае. По определению для матрицы предпочтений по критерию K_t выполняется: $p_{ij}^t + p_{ji}^t = 1 \ (i,j = 1, \dots, n; t = 1, \dots, m)$. Следовательно, для элементов матрицы суммарных предпочтений справедливы равенства $p_{ij}^{\Sigma} + p_{ji}^{\Sigma} = m, i,j = 1, \dots, n$. Тогда:

$$I_i^{DW} = \sum_{j=1}^{n}\left(p_{ij}^{\Sigma} - p_{ji}^{\Sigma}\right) = \sum_{j=1}^{n}\left(p_{ij}^{\Sigma} - \left(m - p_{ij}^{\Sigma}\right)\right) = \sum_{j=1}^{n}\left(2p_{ij}^{\Sigma} - m\right) = 2\sum_{j=1}^{n} p_{ij}^{\Sigma} - nm.$$

Таким образом, при ранжировании индекс альтернативы $a_i \in A$ можно определять суммой элементов i —ой строки матрицы суммарных предпочтений.

Проведем аналогичные преобразования для формулы (3).

$$I_i^{RW} = \frac{\sum_{j=1}^{n} p_{ij}^{\Sigma}}{\sum_{j=1}^{n} p_{ji}^{\Sigma}} = \frac{\sum_{j=1}^{n}\left(m - p_{ji}^{\Sigma}\right)}{\sum_{j=1}^{n} p_{ji}^{\Sigma}} = \frac{nm - \sum_{j=1}^{n} p_{ji}^{\Sigma}}{\sum_{j=1}^{n} p_{ji}^{\Sigma}} = \frac{nm}{\sum_{j=1}^{n} p_{ji}^{\Sigma}} - 1.$$

Таким образом, при ранжировании индекс альтернативы $a_i \in A$ можно определять величиной, обратной к сумме элементов i —го столбца матрицы суммарных предпочтений.

Вместо элементов столбца можно подставить элементы строки:

$$\sum_{j=1}^{n} p_{ji}^{\Sigma} = \sum_{j=1}^{n}\left(m - p_{ij}^{\Sigma}\right) = nm - \sum_{j=1}^{n} p_{ij}^{\Sigma}.$$

Заметим, что полученные формулы справедливы только при выполнении соотношений $p_{ij}^{\Sigma} + p_{ji}^{\Sigma} = m; \ i,j = 1, \dots, n$. Но предложенную методику формирования матриц предпочтений можно распространить на весьма частую ситуацию, когда не для каждой альтернативы известны все оценки по критериям. В этом случае при сравнении альтернатив, информация о которых отсутствует по некоторому t – ому критерию $(t = 1, \dots, m)$, целесообразно положить $p_{ij}^t = p_{ji}^t = 0$, а не $p_{ij}^t = p_{ji}^t = \frac{1}{2}$, уменьшив их значимость в матрице суммарных предпочтений.

Пример 2. Пусть множество альтернатив $A = \{a_1, a_2, a_3, a_4\}$ оценивается по трем критериям качества K_1, K_2, K_3. Численные оценки альтернатив приведены в таблице (значения по шкалам оценок максимизируются).

	K_1	K_2	K_3
a_1	2	3	4
a_2	3	4	2
a_3	2	4	4
a_4	3	3	2

Матрицы предпочтений альтернатив для каждого критерия соответственно имеют вид:

$$P^1 = \begin{pmatrix} \frac{1}{2} & \frac{2}{5} & \frac{1}{2} & \frac{2}{5} \\ \frac{3}{5} & \frac{1}{2} & \frac{3}{5} & \frac{1}{2} \\ \frac{1}{2} & \frac{2}{5} & \frac{1}{2} & \frac{2}{5} \\ \frac{3}{5} & \frac{1}{2} & \frac{3}{5} & \frac{1}{2} \end{pmatrix};\ P^2 = \begin{pmatrix} \frac{1}{2} & \frac{3}{7} & \frac{3}{7} & \frac{1}{2} \\ \frac{4}{7} & \frac{1}{2} & \frac{1}{2} & \frac{4}{7} \\ \frac{4}{7} & \frac{1}{2} & \frac{1}{2} & \frac{4}{7} \\ \frac{1}{2} & \frac{3}{7} & \frac{3}{7} & \frac{1}{2} \end{pmatrix};\ P^3 = \begin{pmatrix} \frac{1}{2} & \frac{2}{3} & \frac{1}{2} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & \frac{2}{3} & \frac{1}{2} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}.$$

Вычислим матрицу суммарных предпочтений *P*:

$$P_\Sigma = \begin{pmatrix} \frac{3}{2} & \frac{157}{105} & \frac{10}{7} & \frac{47}{30} \\ \frac{158}{105} & \frac{3}{2} & \frac{43}{30} & \frac{11}{7} \\ \frac{11}{7} & \frac{47}{30} & \frac{3}{2} & \frac{172}{105} \\ \frac{43}{30} & \frac{10}{7} & \frac{143}{105} & \frac{3}{2} \end{pmatrix}.$$

Применив процедуру разности весов (2), получим вектор индексов

$$I^{DW} = < -\frac{2}{105}; \frac{2}{105}; \frac{58}{105}; -\frac{58}{105} >.$$

Получим ранжирование альтернатив $a_3 - a_2 - a_1 - a_4$.

Убедимся, что ранжирование не изменится при суммировании элементов строк. Поскольку диагональные элементы равны будем суммировать без них. Получим вектор

$$< \frac{943}{210}; \frac{947}{210}; \frac{1003}{210}; \frac{887}{210} >.$$

Применив процедуру отношения весов (1) получим вектор индексов (приближенно)

$$I^{RW} = < 0.997;\ 1;\ 1.1;\ 0.912 >.$$

Ранжирования альтернатив $a_4 - a_1 - a_2 - a_3$ не изменилось.

Если полученные разными методами агрегированные ранжирования различны, то выбрать ранжирование наиболее согласованное с ранжированиями по критериям можно следующим способом.

Зададим отношение предпочтения, соответствующее полученному ранжированию бинарной матрицей $R = \|r_{ij}\|$ с элементами

$$r_{ij} = \begin{cases} 1, \text{если } a_i \text{ не менее предпочтительна } a_j \\ 0 \text{ в противном случае} \end{cases}$$

Проранжируем альтернативы по критериям и зададим соответствующие отношения предпочтения бинарными матрицами $R^1, R^2, \dots, R^m$.

Найдем суммарное расстояние от матрицы R до $R^1, R^2, \dots, R^m$ по формуле

$$D(R) = \sum_{t=1}^{m} d(R, R^t) = \sum_{t=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| r_{ij} - r_{ij}^t \right| \qquad (4)$$

Расстояние между матрицами зададим, как расстояние Хэмминга.

Пример 3. На основе оценок по критериям из примера 2 построим бинарные матрицы R^1, R^2, R^3.

	K_1	K_2	K_3
a_1	2	3	4
a_2	3	4	2
a_3	2	4	4
a_4	3	3	2

$$R^1 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}; \ R^2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}; \ R^3 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Для ранжирования $a_3 - a_2 - a_1 - a_4$ бинарная матрица

$$R = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Суммарное расстояние найдем по формуле (4)

$$D(R) = d(R, R^1) + d(R, R^2) + d(R, R^3) = 8 + 2 + 4 = 14.$$

Если мы получим различные агрегированные ранжирования, решая задачу разным методами, то их можно сравнить по суммарному расстоянию до предпочтений по критериям. Наиболее согласованным с ранжированиями по критериям является агрегированное ранжирование с наименьшим суммарным расстоянием.

5. ОРГАНИЗАЦИЯ РАБОТЫ С БОЛЬШИМИ ДАННЫМИ В ПРОГРАММНОЙ СИСТЕМЕ МНОГОКРИТЕРИАЛЬНОГО ВЫБОРА

При решении задач с большим объемом исходных данных у разработчика программной системы возникает проблема выбора наиболее удобного способа хранения информации и для ввода ЛПР, и для считывания программой. Необходимо учесть, что в задачах принятия решений информация может постоянно обновляться в связи с поступлением новых вариантов альтернатив.

В разработанной программной системе выбран наиболее простой и удобный для ЛПР способ ввода данных через Excel таблицы. Программа загружает Excel таблицу в систему, считывает и создаёт свои таблицы, удобные для работы алгоритмов многокритериального выбора. База данных программной системы содержит следующую информацию.

1. Таблица с наименованиями альтернатив и их числовые характеристиками.
2. Информация по критериям: тип шкалы, максимизируются или минимизируются оценки по шкале, весовые коэффициенты важности критериев.

Программная система организована так, что исходные данные хранятся на удаленном сервере, т. о. зарегистрированный пользователь всегда имеет доступ для их корректировки, причем с различных устройств.

Рассмотрим работу программы на примере ранжирования дронов с целью приобретения для наблюдения за местностью, сьемки местности и передачи информации оператору. В настоящее время разработано сотни различных моделей дронов даже в рамках данного предназначения. ЛПР обращается на специальный сайт по продажам дронов, например в интернет-магазин, где ему предоставляется огромный список моделей. Сайт предлагает ЛПР помощь в выборе наиболее предпочтительных вариантов дронов. Разработчики сайта хранят список моделей дронов с их характеристиками по критериям качества в Excel таблице. В таблице также присутствует информация о том максимизируются или минимизируются оценки по критериям.

Будем оценивать модели дронов по следующим критериям: K_1 — стоимость (в рублях, min); K_2 — вес (в кг, min); K_3 — максимальная длительность полета (в часах или мин., max); K_4 — максимальная высота полета (в метрах, max); K_5 — обслуживание (в баллах, max); K_6 — защита от столкновений и погодных явлений (в баллах, max).

Основные этапы работы программной системы ранжирования альтернатив в диалоговом режиме с ЛПР

1. ЛПР (лицо, принимающее решения) вводит интервалы-ограничения для каждого критерия (например, ограничения по стоимости). Таблица может сама упорядочивать модели по одному критерию, и мы легко отсекаем дроны вне заданного интервала.
2. ЛПР выбирает критерии из заданного перечня, по которым хочет ранжировать дроны. Дополнительно, по желанию ЛПР, можно ввести весовые коэффициенты важности критериев.
3. Программа ранжирует дроны и выдает на экран 5 (или более) наилучших аппаратов.

ЛПР имеет возможность изменять наборы критериев и повторно осуществлять выбор наиболее предпочтительных моделей дронов.

ЛПР выбирает одну или несколько понравившихся ему моделей и оставляет на сайте заказ.

6. ЗАКЛЮЧЕНИЕ

Предложены и обоснованы алгоритмы многокритериального выбора, не требующие приведения шкал критериев к однородным. Все алгоритмы имеют полиномиальную вычислительную сложность, что позволяет работать с большими массивами исходной информации. Разработана программная система, предназначенная для решения задач многокритериального принятия решений, а также для проведения практических и лабораторных работ в вузах по дисциплине «Математическая теория принятие решений».

Литература

1. *Петровский А.Б.* Теория принятия решений. М.: Академия. 2009.
2. *Smerchinskaya S.O., Yashina N.P.* On an algorithm for pairwise comparison of alternatives in multi-criteria problems // International Journal of Modeling, Simulation, and Scientific Computing. 2018. Vol. 9, Issue 1. DOI: 10.1142/S179396231850006X.
3. *Смерчинская С.О., Яшина Н.П.* Агрегирование предпочтений в многокритериальных задачах // Вестник Московского авиационного института. 2013. № 2, том 20. С. 219–225.
4. *Shikhman V., Müller D.* Mathematical Foundations of Big Data Analytics. 2021. DOI: 10.1007/978-3-662-62521-7. ISBN: 978–3-662–62520–0
5. *Нефедов В.Н., Смерчинская С.О., Яшина Н.П.* Построение агрегированного отношения, минимально удаленного от экспертных предпочтений. Прикладная дискретная математика. 2018. № 42. С. 120–132. DOI: 10.17223/20710410/42/9

Big Data Analysis in Multi-Criteria Choice Problems

Andrey A. Ivanov*
Moscow Aviation Institute, Moscow, Russia
ORCID: https://orcid.org/0000-0003-4433-6449
e-mail: ivanov17andrey@gmail.com

Nina P. Yashina**
Moscow Aviation Institute, Moscow, Russia
ORCID: https://orcid.org/0000-0002-8401-0315
e-mail: nina_p_yashina@mail.ru

The problem of multi-criteria choice with non-uniform scales of criteria is considered. A model of a multicriteria choice problem is described, the main elements of which are sets of alternatives and quality criteria, as well as algorithms that allow ranking alternatives without prior reduction of the criteria scales to homogeneous ones. Algorithms for constructing aggregated ranking of alternatives are based on the construction of preference matrices by criteria containing information on the degree of superiority of one alternative over another. Propositions are proved that allow ranking alternatives with assessments according to two quality criteria. Algorithms for indexing alternatives are proposed that allow ranking alternatives for an arbitrary number of criteria. The best aggregated ranking is determined by the total distance to the rankings of alternatives by criteria. All algorithms have polynomial computational complexity, which makes it possible to work with large arrays of initial information. A software system for ranking alternatives in problems with big data has been developed. The initial information is stored in Excel tables, which makes it easy to take into account the limitations on the criteria scales. The operation of the software system is demonstrated by the example of choosing the best version of a drone for purchase in order to observe the terrain, shoot it and transmit information to the operator.

Keywords: decision making, quality criterion, big data analysis, preference matrix, ranking of alternatives.

For citation:
Ivanov A.A., Yashina N.P. Big Data Analysis in Multi-Criteria Choice Problems. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2022. Vol. 12, no. 2, pp. 5–19. DOI: https://doi.org/10.17759/mda.2022120201 (In Russ.,abstr. in Engl.).

***Andrey A. Ivanov,** Graduate Student, Moscow Aviation Institute (MAI), Moscow, Russia, ORCID: https://orcid.org/0000-0003-4433-6449, e-mail: ivanov17andrey@gmail.com

****Nina P. Yashina,** Candidate of Physical and Mathematical Sciences, Associate Professor, Department of Mathematical Cybernetics, Moscow Aviation Institute (MAI), Moscow, Russia, ORCID: https://orcid.org/0000-0002-8401-0315, e-mail: nina_p_yashina@mail.ru

References

1. Petrovskiy A.B. Teoriya Prinyatiya Resheniy. M.: Akademiya. 2009
2. Smerchinskaya S.O., Yashina N.P. On an algorithm for pairwise comparison of alternatives in multi-criteria problems // International Journal of Modeling, Simulation, and Scientific Computing. 2018. Vol. 9, Issue 1. DOI: 10.1142/S179396231850006X.
3. Smerchinskaya S.O., Yashina N.P. Agregirovaniye predpochteniy v mnogokriteriallnikh zadachakh // Vestnik Moskovskogo aviacionnogo instituta. 2013. № 2, tom 20. C. 219–225.
4. Shikhman V., Müller D. Mathematical Foundations of Big Data Analytics. 2021. DOI: 10.1007/978-3-662-62521-7. ISBN: 978–3-662–62520–0
5. Nefedov V.N., Smerchinskaya S.O., Yashina N.P Postroeniye agregirovannogo otnosheniya, minimallno udalennogo ot ekspertikh predpochteniy. Prikladnaya diskretnaya matematika. 2018. № 42. C. 120–132. DOI: 10.17223/20710410/42/9

Получена 25.04.2022
Принята в печать 03.06.2022

Received 25.04.2022
Accepted 03.06.2022