

РАНГОВО-ЭКСПЕРТНАЯ ФУНКЦИЯ ОТКЛОНЕНИЙ ДЛЯ КЛАССИФИКАЦИИ СЛОЖНЫХ ОБЪЕКТОВ¹

В.Б. Коробов, А.Г. Тутыгин, А.С. Лохов

Аннотация. Предложена новая функция для классификации природных, социальных и природно-социальных объектов. В ее основу положена сумма ранговых отклонений сравниваемого объекта от эталонного с учетом значимости характеристик (показателей) объекта. Для оценки характеристик применяются весовые коэффициенты, рассчитываемые по результатам экспертных опросов или другими методами. Разработана вербально-числовая шкала оценки близости сравниваемых объектов по численному значению функции. Показано, что предложенная функция не является метрикой в геометрическом смысле, но является функцией близости, применяемой в задачах многомерного шкалирования. Приведены примеры расчета значений функции для геоэкологической задачи сравнения уязвимости территорий к аварийным разливам нефти, а также для задачи выбора альтернатив при покупке недвижимости. Представлена рекомендованная последовательность применения набора процедур для решения задач с использованием функции отклонений.

Ключевые слова: ранг, объект, классификация, вербально-числовая шкала.

ВВЕДЕНИЕ

Классификация представляет собой одну из главных задач науки. Целью классификации является упорядочение объектов таким образом, чтобы объекты, относящиеся к одной группе, можно было считать близкими по своим качествам. Результатом классификации всегда является группировка объектов согласно их свойствам.

Методология классификаций для различных наук разная. Можно провести границу между общественными и естественными науками, в рамках которых сложно выработать единый подход. Так, например, разные авторы выделяют от 8 до 22 типов цивилизаций [1]. Еще сложнее классифицировать смешанные объекты, составленные из социальных, природных, техногенных и прочих компонентов, например такие, как разные типы поселений. Тем не менее, это вполне возможно, если со-

поставлять компоненты объектов в некоторых соизмеримых величинах: баллах, рангах, расстояниях в гиперпространствах и т. д.

Единых универсальных методов классификации нет. Но они активно разрабатываются представителями различных наук. Большой вклад в создание классификаций природных объектов внес Карл Линней [2].

1. ИДЕОЛОГИЯ СРАВНЕНИЯ ОБЪЕКТОВ

В основу предлагаемой авторами методологии положено покомпонентное сравнение объектов. Принадлежность их к одному классу определяется с помощью функции, которая представляет собой суммарную величину отклонений между характеристиками (показателями) соответствующих компонентов. Идея эта не нова и применяется для решения многих прикладных задач. Одной из самых простых таких функций и в то же время достаточно эффективной является хорошо известная метрика Ричарда Хемминга [3], разработанная им в рамках теории кодирования, но затем нашедшая применение во многих областях науки и техники. Кроме того, существуют и ее модифицированные варианты – например, взвешенная метрика Хемминга [4].

¹ Работа выполнена в рамках темы госзадания № FMWE-2021-0006 «Современные и древние донные осадки и взвесь Мирового океана – геологическая летопись изменений среды и климата: рассеянное осадочное вещество и донные осадки морей России, Атлантического, Тихого и Северного Ледовитого океанов – литологические, геохимические и микропалеонтологические исследования; изучение загрязнений, палеообстановок и процессов в маргинальных фильтрах рек».



Остановимся на метрике Хемминга несколько подробнее. В общем виде она записывается следующим образом:

$$d(x_i, x_j) = \sum_{i,j=1}^N |x_i - x_j|,$$

где x_i, x_j – координаты векторов сравниваемых объектов, $i, j = 1, N$; N – количество сравниваемых характеристик объектов. Наиболее простым методом определения близости объектов является представление разностей в виде бинарных отношений: если разница между соответствующими характеристиками менее некоторой заданной величины, допустим 10 %, ей присваивается значение 1, если больше – 0. Затем подсчитывается число единиц, и если их сумма равна или больше некоторого процента от всего количества сравниваемых характеристик, например, 90 %, то такие объекты считаются близкими и их относят к одному классу. Критерии близости и принадлежности – в данном случае 10 % и 90 % соответственно – устанавливаются исследователями произвольно в зависимости от условий задачи. Единых рекомендаций здесь нет и выработать их не представляется возможным, при этом сам подход является довольно простым.

Схожие идеи применяются также в методах многомерного шкалирования, предназначенных для анализа данных путем понижения их размерности, реализуемом при помощи сравнения объектов различными способами. Они широко используются в социологии и психологии, их основоположником является Л. Гуттман [5, 6].

В настоящее время в связи с развитием методов машинного обучения, искусственного интеллекта и кибернетики в целом особое место в науке занимает схожая с классификацией задача кластеризации [7], в которой сами классы должны быть определены алгоритмом.

Тем не менее, различные методы классификации разрабатываются и сейчас. Из них можно отметить метод ATOVIC (Amended fused TOPSIS-VIKOR for classification) [8] – комбинацию двух методов многокритериального принятия решений для задач выбора альтернатив, модифицированную для проведения классификаций. Здесь для каждого класса определяется идеальный объект и так называемый негативный объект с точки зрения принадлежности к классу. Далее при помощи метрик Минковского и Чебышева проверяется близость каждого идеального и негативного объекта к каждому классу, на основе которой принимается ре-

шение о принадлежности каждого из них [9]. Существуют и другие метрические методы, наиболее известный из них – метод k ближайших соседей [10], он имеет большое количество различных модификаций и позволяет использовать разные метрики.

В контексте методов классификации можно также упомянуть наивный байесовский классификатор, основанный на теореме Байеса и использующий для принятия решения различные вероятностные статистики [11].

Практика применения метрики Хемминга, в том числе и авторами, показала ее эффективность, но в тоже время, как это часто бывает, когда методология, выведенная для решения конкретных задач, транслируется на другие объекты, выявила и некоторые ее недостатки [12]. Эти недостатки присущи и другим метрическим методам; главным из них является то, что даже очень высокий процент совпадений пар характеристик при не менее высоком критерии совпадения отдельных пар (который может быть установлен и менее 5 %) не гарантирует, что объекты будут принадлежать к одному классу.

Например, пусть производится геоэкологическая классификация территорий по степени их загрязненности. Для описания объекта обычно выбирают ограниченное число показателей (характеристик), наиболее характерных для данных территорий. При этом почти по всем показателям, кроме одного, могут быть очень близкие результаты, вплоть до практически полного совпадения, что будет давать все формальные основания отнести сравниваемые объекты к одному классу. Но единственное несовпадение может быть настолько большим – на порядки – что объекты нельзя будет отнести к одному типу ни при каких условиях. Именно так и происходит в случаях аварийных разливов и выбросов токсичных веществ в больших объемах, ведущих к экологическим катастрофам. Применение метрики Хемминга в этих случаях неправомерно, поскольку приводит к ошибочным результатам. Этой метрике и подобным ей присущи и некоторые другие недостатки, которые будут рассмотрены нами в ходе дальнейшего изложения.

Несмотря на это, сам принцип сравнения и классификации объектов, основанный на анализе отклонений характеристик объектов, представляется оправданным и методологически верным. На базе этого принципа авторами предложена функция, позволяющая проводить сравнение и классифицировать широкий круг объектов.

2. РАНГОВАЯ ФУНКЦИЯ ОТКЛОНЕНИЙ

Чтобы определить, относится ли данный объект к тому или иному классу, его необходимо сравнить с неким другим объектом, входящим в этот класс. Для проведения таких сравнений введем понятие «эталонного» объекта, под которым будем понимать объект, в наибольшей степени характеризующий данный класс. (Каким образом формализовать объект-эталон, будет рассмотрено авторами в отдельной публикации, поскольку это требует разработки специальной методологии, зависящей от сложности объекта и наличия информации.) Тем не менее, с точки зрения процедур расчета эталон также является объектом, поэтому далее речь пойдет о более общем случае – сравнении двух объектов друг с другом.

Отклонения можно измерять несколькими способами. Наиболее естественный – простая разность, выраженная в процентах или булевыми символами. Можно предварительно перевести все характеристики в баллы путем применения специальных шкал, что позволяет получать разности в унифицированных единицах. Нередко проводятся различные нормировки показателей для приведения их к безразмерному виду. Применяются и другие методы для того, чтобы сделать характеристики объекта сопоставимыми.

В настоящей работе предлагается метод, основанный на ранговых шкалах, обладающих одним важным свойством: вербальные определения не имеют знака, что позволяет отрицательные и положительные значения, причем как количественные, так и качественные, рассматривать как однозначные. Это очень важно для некоторых характеристик. Так, температура воздуха при оценке комфортности климата для проживания оценивается примерно в диапазоне от $-50\text{ }^{\circ}\text{C}$ до $+40\text{ }^{\circ}\text{C}$, а промышленные предприятия по их финансовому результату, который может быть как отрицательным, так и положительным, – по степени их убыточности или доходности.

Перейдем непосредственно к описанию предлагаемой функции. Пусть даны два объекта x и y со следующими наборами характеристик: $\bar{x} = (x_1, \dots, x_N)$, $\bar{y} = (y_1, \dots, y_N)$, $\bar{x}, \bar{y} \in \mathbb{R}^N$, которые принадлежат множеству векторов размерности N , а его элементы – множеству действительных чисел и представлены в своих «естественных» единицах измерения, например физических. Тогда вектор их отклонений $\bar{z} \in \mathbb{R}_+^N \cup \{0\}$ с неотрицательными координатами будет иметь вид

$$\bar{z} = |\bar{x} - \bar{y}| = (z_1, \dots, z_N). \quad (1)$$

Однако сама по себе такая векторная величина малоинформативна и может указывать лишь на близость величин характеристик объектов. Так, если значения одной характеристики у объектов отличаются на 0,01 ед. при прочих равных, означает ли это, что они близки друг к другу? Совсем не обязательно; имея информацию лишь о двух значениях, невозможно оценить характер и диапазон изменения какой-либо характеристики.

Одним из способов решения этой проблемы является ранжирование – упорядоченное расположение значений характеристик в определенной последовательности, в данном случае по возрастанию отклонения их величин. Ранжированию подлежат как количественные, так и качественные характеристики, что позволяет использовать эту процедуру для измерения отклонений по разнице рангов между сравниваемыми объектами. При этом сразу можно получить не только количественную, но и вербальную оценку степени отклонения каждой характеристики – «незначительно», ... «существенно» и т. д., т. е. определить, насколько оно значимо, что может быть полезным в некоторых задачах частной оптимизации, например при анализе альтернатив.

Однако для этого необходимо разрабатывать специальные вербально-числовые шкалы рангов для каждой характеристики, но эти затраты окупаются возможностью их самостоятельного применения для решения других задач.

Введем в рассмотрение неотрицательную матрицу частных ранговых шкал $\mathbf{H} \in \mathbb{M}_{N \times M}(\mathbb{R}_+ \cup \{0\})$, в которой строка с номером $i = \overline{1, N}$ соответствует шкале характеристики с тем же номером:

$$\mathbf{H} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,M} \\ \vdots & \ddots & \vdots \\ h_{N,1} & \cdots & h_{N,M} \end{pmatrix},$$

где $M \in \mathbb{N}$ – максимальное значение ранга, $M \geq 1$, тогда число градаций или интервалов ранговой шкалы будет равно $M + 1$ (при $M = 1$ получаем бинарную шкалу). Как было сказано выше, значения шкалы должны быть ранжированы по возрастанию, от левого столба матрицы \mathbf{H} к правому, это выражается следующим условием: для каждого i выполняется условие $0 \leq h_{i,1} < \dots < h_{i,M}$. При построении ранговых шкал применяются различные методы. Исходя из личного опыта авторов, представляется предпочтительным строить их путем экспертного опроса на основе теории нечетких множеств.



Различные объекты могут описываться разным числом характеристик, при этом один и тот же объект в зависимости от его изученности или условий, предъявляемых к точности классификации, также может иметь различную длину компонентов. В связи с этим возникает вопрос: как быть, если число градаций будет неодинаковым для каждого компонента (что вполне естественно)? В данном случае можно предложить несложную техническую процедуру синхронизации шкал, смысл которой заключается в введении масштабных множителей, приводящих все шкалы к единому числу градаций. Но таких ситуаций предпочитают избегать, что вполне оправдано, и для описания объекта разрабатывают шкалы с одинаковым числом градаций. Так, для классификации потенциала загрязнения атмосферы была разработана единая вербальная шкала из пяти градаций для всех семи характеристик атмосферы [13], используемая на практике уже несколько десятилетий.

Определим вектор ранговых отклонений $\bar{\mathbf{r}} = (r_1, \dots, r_N)$ объектов x и y , $\bar{\mathbf{r}} \in \mathbb{N}_0^N$, такой, что выполнены ограничения $0 \leq r_i \leq M$, $i = \overline{1, N}$. Тогда

$$r_i = \sum_{j=1}^M \theta(z_i - h_{i,j}), \quad (2)$$

где θ – функция Хэвисайда:

$$\theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases} \quad x \in \mathbb{R}.$$

Практически с помощью формулы (2) для каждого i осуществляется подсчет количества значений $h_{i,j}$, строго меньших z_i . Это количество и есть значение ранга отклонения, которое будет изменяться от 0 до M включительно, где M является количеством столбцов матрицы H , а $M + 1$ – количеством градаций частных ранговых шкал. Данное выражение также можно переписать и в векторном виде, заменив функцию Хэвисайда функцией подсчета неотрицательных элементов вектора.

Все характеристики объекта неравноценны по их роли в функционировании объекта или относительно целевого назначения объекта. Чтобы учесть эту неравноценность характеристик, используют весовые коэффициенты, выражающиеся в долях единицы или в процентах. Введем вектор $\bar{\mathbf{k}} = (k_1, \dots, k_N)$, $\bar{\mathbf{k}} \in \mathbb{R}_+^N \cup \{0\}$ весовых коэффициентов, сумма значений компонент которого долж-

ны быть равна единице: $\sum_{i=1}^N k_i = 1$.

Тогда ранговая функция отклонений определяется как отображение из декартова произведения двух множеств векторов размерности N , элементы которых являются неотрицательными действительными числами, во множество этих же чисел $R : (\mathbb{R}_+^N \cup \{0\}) \times (\mathbb{R}_+^N \cup \{0\}) \rightarrow \mathbb{R}_+ \cup \{0\}$ и выражается следующим образом:

$$R(\bar{\mathbf{k}}, \bar{\mathbf{r}}) = \frac{\bar{\mathbf{k}} \cdot \bar{\mathbf{r}}}{M} = \sum_{i=1}^N \frac{k_i r_i}{M}, \quad (3)$$

где символом « \cdot » обозначено скалярное произведение векторов. Здесь также произведена нормировка вектора ранговых отклонений $\bar{\mathbf{r}}$ по максимальному значению ранга, поэтому область значений R является множеством долей единицы, которые для удобства можно представить в процентном выражении.

Описанная функция не является метрикой, так как для нее не выполняются аксиома тождества и неравенство треугольника. Поэтому эта функция не определяет в привычном геометрическом смысле расстояние между объектами в некотором пространстве. Для решения этой проблемы в теории многомерного шкалирования введено определение функции близости [6, с. 39] – «...это функция вида $s(a_i, a_j)$... такая, что для всех i, j, k удовлетворяются соотношения (аксиомы) (a_i, a_j, a_k – векторы характеристик объекта) $s(a_i, a_i) \geq s(a_i, a_j)$; $s(a_i, a_j) = s(a_j, a_i)$; ... для больших значений $s(a_i, a_j)$ и $s(a_j, a_k)$ величина $s(a_i, a_k)$ имеет по крайней мере тот же порядок» [6, с. 39].

Эти соотношения являются по сути ослабленными аксиомами из определения метрики. Несложно проверить, что для описанной авторами функции отклонений они выполняются, поэтому она является функцией близости. Это, в частности, позволяет применять аффинное преобразование, используемое в методах многомерного шкалирования для масштабирования и понижения размерности.

Одним из основных преимуществ этой функции является возможность устанавливать критерии равенства величин. Так, объекты, измеряемые физические характеристики которых отличаются не более чем на ошибку измерения, должны приниматься равными. То есть первая градация частной ранговой шкалы (первый столбец матрицы H) для них должна быть не меньше погрешности измерения. Но в общем случае она может быть равна нулю, что будет означать отсутствие отклонения только при точном равенстве характеристик; это применимо, например, к качественным величинам.

Естественным образом возникает вопрос – почему необходимо использовать частные ранговые шкалы и нельзя взять отношение характеристик в процентах? Например, заменив формулы (1) и (2) следующей:

$$z_i = 1 - \frac{\min(x_i, y_i)}{\max(x_i, y_i)}, \quad i = \overline{1, N}. \quad (4)$$

Ответим на этот вопрос на примере. Пусть объектами являются какие-либо водные массы, у которых имеется характеристика – концентрация взвешенных веществ в воде, измеряемая в мг/л. Она имеет довольно широкий диапазон изменения и может принимать значения как в несколько мг/л, так и в несколько сотых мг/л. Тогда, если она равна 0,01 и 0,02 мг/л у двух сравниваемых объектов, то их различие по данной характеристике с использованием формулы (4) составляет 50 %, тогда как у объектов со значениями 2 и 0,01 мг/л оно равно 99,5 %. В данном случае объекты со значениями 0,01 и 0,02 мг/л могут, с точки зрения конкретной задачи, различаться несущественно либо могут быть приняты одинаковыми, если точность измерений составляла 0,01 мг/л. Поэтому отличие характеристик в 50 % в данном случае некорректно.

3. УНИВЕРСАЛЬНАЯ ШКАЛА КЛАССИФИКАЦИИ ОБЪЕКТОВ

Для оценки близости сравниваемых объектов необходима специальная шкала. Предложим один из возможных ее вариантов исходя из следующих соображений. Максимум функции, как сказано выше, достигается в том случае, когда разница между объектами будет наибольшей, т. е. ранги отклонения каждой характеристики принимают свое максимальное значение M . Минимальное значение функции будет равным нулю, т. е. в этом случае один объект строго соответствует другому. Тогда все отклонения будут равны нулю. Следовательно, все возможные отклонения будут находиться в диапазоне от 0 до 1, или 100 %. Это дает возможность построить оценочную шкалу (табл. 1).

Могут быть и другие варианты данной шкалы, но это уже вопрос дискуссии. (Отметим, что и эту шкалу предпочтительно строить методом нечетких множеств.) В графическом виде эта шкала представлена на рис. 1. Такой же вид имеют и частные ранговые шкалы, образующие матрицу H .

Универсальная шкала оценки близости объектов по значению ранговой функции отклонений

№	Интервал, %	Степень близости
1	0–5	Совпадение
2	6–10	Незначительное расхождение
3	11–20	Небольшое расхождение
4	21–30	Умеренное расхождение
5	31–40	Заметное расхождение
6	41–55	Существенное расхождение
7	56–70	Значительное расхождение
8	71–100	Весьма значительное расхождение

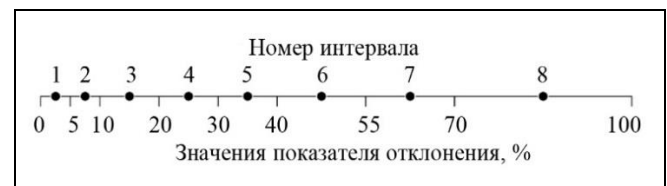


Рис. 1. Универсальная шкала оценки близости объектов. Цифрами обозначены номера интервалов, соответствующие табл. 1

4. ПРИМЕРЫ РАСЧЕТА

Пример 1. Данный пример в первую очередь демонстрирует технические аспекты применения процедур расчета, а методики выбора частных ранговых шкал и весовых коэффициентов выходят за рамки настоящей работы.

Пусть решается задача сравнения уязвимости территорий к аварийным разливам нефти [14]. Объектами в данном случае будут области, для которых определены значения показателей. Рассмотрим три показателя:

- прогнозируемая площадь загрязнения, полученная путем моделирования для конкретного участка местности, m^2 , по величине которой определяется степень ущерба и объем рекультивационных работ;
- средний уклон поверхности, %, от которого зависит скорость распространения загрязнения по земной поверхности и форма пятна;
- показатель распространенности водных объектов, %, при попадании в которые нефть и нефтепродукты потоками воды переносятся на значительные расстояния, увеличивая тем самым площадь загрязнения (берега и прибрежные территории) и ущерб природной среде.

Для них были разработаны частные ранговые шкалы с четырьмя градациями и определены весовые коэффициенты. Исходя из этого, $N = 3$, $M = 4$, необходимые для вычисления функции входные данные представлены в табл. 2.

Отметим, что в табл. 2 все значения, кроме столбцов 1 и 4, представлены в единицах измерения самих показателей. Столбцы 5–8 представляют матрицу частных ранговых шкал H . В табл. 3 приведен пример расчета

Таблица 2

Исходные данные для расчета ранговой функции отклонений

Характеристика	i	\bar{x}	\bar{y}	\bar{k}	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$	$h_{i,4}$
Номер столбца	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Площадь загрязнения, м ²	1	2376,7	1831,6	0,5	10	25	100	200
Уклон поверхности, %	2	1,05	1,13	0,3	0,05	0,25	0,5	1
Распространенность водных объектов, %	3	2	0,24	0,2	0,01	0,1	1	5

Таблица 3

Результаты расчета функции отклонений

i	\bar{z}	$z_i - h_{i,1}$	$z_i - h_{i,2}$	$z_i - h_{i,3}$	$z_i - h_{i,4}$	r_i	$r_i \cdot k_i / M$	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
1	545,1	535,1	520,1	445,1	345,1	4	0,5	
2	0,08	0,03	-0,17	-0,42	-0,92	1	0,075	
3	1,76	1,75	1,66	0,76	-3,24	3	0,15	
Значение ранговой функции отклонений							$R = 0,725$	

вектора отклонений характеристик (формула (1), столбец 2), его преобразование к вектору ранговых отклонений (формула (2), столбцы 3–6 и 7) и расчет скалярного произведения (формула (3), столбец 8). Преобразование к ранговым шкалам происходит путем подсчета положительных значений в столбцах 3–6, что и является значением ранга (столбец 7).

Значение функции отклонений равно 0,725; это означает, что объекты отличаются друг от друга на 72,5 %. Согласно универсальной шкале классификации это соответствует степени близости «весьма значительное расхождение» (см. табл. 1).

Такие таблицы могут строиться для каждого участка объекта или отдельной альтернативы. Тогда мы получим некоторое множество предметов исследования, позволяющих провести их классификацию, что является одной из основных целей изучения как объектов, так и альтернатив, например, в задачах промышленной безопасности и задачах выбора (см. ниже). ♦

Пример 2. В качестве еще одного примера приведем решение классической в области принятия решений задачи выбора альтернатив при покупке квартиры. Пусть имеются три альтернативных варианта (a , b и c) и рассматривается пять их различных характеристик. Обозначим их векторы как \bar{a} , \bar{b} и \bar{c} (табл. 4, столбцы 3–5). Наиболее простой способ сформировать эталон (\bar{e}) – это взять минимально или максимально возможные значения характеристик. Например, для свойства «износ жилого дома, %» допустимое значение эталона

0, следовательно, чем меньше износ, тем лучше (табл. 4, столбец 2). Сформируем вектор весовых коэффициентов (наиболее быстрый метод – простое ранжирование), расположим характеристики в порядке убывания их значимости по индексу i : 1, 2, 3, 4, 5, тогда для данной задачи $k_i = (6-i)/15$ (табл. 4, столбец 6). Сформируем матрицу частых ранговых шкал \mathbf{H} с четырьмя градациями или интервалами для каждой характеристики, тогда $N = 5$, $M = 3$. В данном случае, чтобы не усложнять пример, шкалы сформированы путем опроса одного эксперта, применительно к задаче – покупателя или его представителя. Но их может быть несколько, наиболее подходящим значением авторы считают порядка десяти человек. Также шкала может быть сформирована и другими методами. На этом все необходимые для применения рангово-экспертной функции исходные данные сформированы (см. табл. 4).

Перейдем к необходимым расчетам – последовательному применению функции для сравнения эталона с каждой из альтернатив. Для удобства первые несколько шагов вычислений опущены, они выполнены по аналогии с предыдущим примером (см. табл. 3, столбцы 2–6). В табл. 5 приведены значения рангов (столбцы 2, 4, 6) их нормировка (столбцы 3, 4, 5) и значения функции отклонений для каждой из альтернатив.

Наилучшей в данном случае будет альтернатива b с наименьшим значением рангово-экспертной функции отклонений $R = 0,553$, т. е. наименее отличающаяся от принятого эталона. ♦

Таблица 4

Исходные данные для расчета ранговой функции отклонений

Характеристика	i	\bar{e}	\bar{a}	\bar{b}	\bar{c}	\bar{k}	$h_{i,1}$	$h_{i,2}$	$h_{i,3}$
Номер столбца	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Стоимость, тыс. руб./м ²	1	70	80	100	150	0,33	5	10	50
Расстояние до центра города, км	2	1	15	5	6	0,27	5	10	15
Износ жилого дома, %	3	0	30	60	10	0,2	10	25	50
Количество парковочных мест вблизи дома, шт.	4	1000	500	50	100	0,07	100	500	1000
Площадь квартиры, м ²	5	60	63	78	55	0,13	5	10	25

Таблица 5

Результаты расчета функции отклонений от эталона

Альтернатива	i	\bar{a}		\bar{b}		\bar{c}	
		r_i	$r_i \cdot k_i / M$	r_i	$r_i \cdot k_i / M$	r_i	$r_i \cdot k_i / M$
Номер столбца	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Стоимость тыс. руб./м ²	1	2	0,220	2	0,220	3	0,330
Расстояние до центра города, км	2	2	0,180	0	0,000	1	0,090
Износ жилого дома, %	3	2	0,133	3	0,200	1	0,067
Количество парковочных мест вблизи дома, шт.	4	2	0,047	2	0,047	2	0,047
Площадь квартиры, м ²	5	0	0,000	2	0,087	1	0,043
Показатель отклонения R		0,580		0,553		0,577	

5. ОБЛАСТИ ПРИМЕНЕНИЯ РАНГОВО-ЭКСПЕРТНОЙ ФУНКЦИИ

Представленная в настоящей работе функция позволяет классифицировать объекты, для описания которых используются характеристики (показатели), выраженные самыми разнообразными количественными и качественными величинами. К таким объектам в первую очередь можно отнести социально-экономические и геоэкологические объекты. При их формализации приходится учитывать большое число факторов из разных областей знания: природные, технические, социальные, военные, геополитические и др. Ранговые шкалы позволяют сделать такие характеристики соизмеримыми. При этом можно использовать уже разработанные вербально-числовые шкалы, но только если они соответствуют условиям поставленной задачи.

В целом же ранговая функция отклонений и ее составляющие могут рассматриваться как один из элементов классификационных технологий [15] и

использоваться для таких сложных задач, как географо-экологическое районирование территорий [14] и др. Покажем это на некоторых, намеренно упрощенных примерах, встречающихся в реальной жизни.

Снова обратимся к задаче уязвимости территорий к аварийным разливам нефти вдоль трассы трубопровода (см. пример 1 из § 4). Разобьем всю трассу на одинаковые по протяженности участки. Протяженность и ширина участков принципиального значения не имеют: возможности ГИС-технологий позволяют сделать их сколь угодно малыми. Рассчитанная для каждого участка функция отклонения (см. табл. 4) картографируется (рис. 2), что дает возможность оценить степень уязвимости объекта в целом и разработать научно обоснованные рекомендации для схемы производственного экологического контроля и выбрать места оптимального расположения средств ликвидации аварийных разливов (заградительные боны, сорбенты и т. д.). Например, на одном из участков с наибольшей уязвимостью – в соответствии с рис. 2, на северо-восточном с показателем $R = 0,87$.

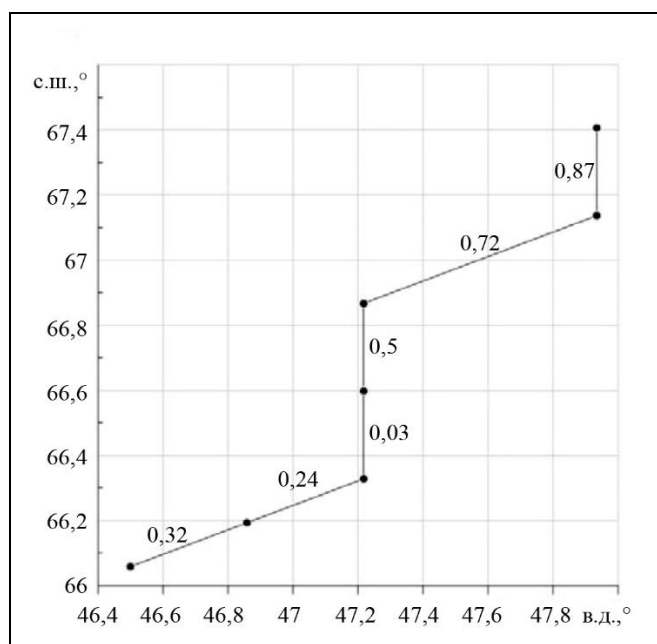


Рис. 2. Схема распределения показателя отклонения для фрагмента линейного объекта (в данном случае – нефтепровода), числами обозначены значения показателя отклонения

ЗАКЛЮЧЕНИЕ

Предложена методика классификации объектов на основе разработанной ранговой функции отклонения, применение которой представляет собой набор процедур, выполняемых в следующей рекомендуемой последовательности.

1) Определяется цель классификации.

2) В соответствии с целью, производится выбор характеристик объекта, позволяющих этот объект формализовать; число его характеристик не ограничено, но должно быть одинаковым для всех объектов.

3) Строятся или используются имеющиеся частные ранговые вербально-числовые шкалы для каждой характеристики объекта; при этом все они должны иметь равное число градаций либо быть предварительно синхронизованы посредством поправочного коэффициента.

4) Для каждого класса строится объект-эталон, характеристики которого являются наиболее подходящими для данного класса.

5) Рассчитываются весовые коэффициенты характеристик.

6) После определения исходных данных последовательность расчета ранговой функции отклонений состоит из трех шагов.

6.1) Вычисляется вектор отклонений характеристик.

6.2) Производится преобразование вектора отклонений характеристик к вектору ранговых отклонений.

6.3) Вектор ранговых отклонений нормируется, затем вычисляется скалярное произведение этого нормированного вектора с вектором весовых коэффициентов.

7) По вербально-числовой шкале определяется степень близости сравниваемого объекта к эталонным; при необходимости производится последовательное сравнение с другими объектами-эталонами.

Данная функция позволяет решать и другие задачи, такие как простое сравнение двух объектов с заданными свойствами при анализе альтернатив, задачи разработки сценариев, оптимизации объектов.

Однако при этом может сложиться ситуация, когда все значения функции отклонения попадут в одну градацию, т. е. объект окажется однородным. Такое иногда встречается на практике. В таких случаях вырабатываются дополнительные правила классификации, позволяющие дифференцировать объект исследования. Часто бывает достаточным добавление одного-двух условий, например, при равенстве альтернатив, отдавать приоритет той, у которой наибольший вклад внес фактор с наибольшим весовым коэффициентом при минимальном отклонении его от эталонного состояния.

ЛИТЕРАТУРА

1. Хантингтон С. Столкновение цивилизаций. – М.: АСТ: Астрель, 2011. – 571 с. [Huntington, S.P. The Clash of Civilizations and the Remaking of World Order. – New York: Simon & Schuster, 2011. – 368 p.]
2. Линней К. Философия ботаники. – М.: Наука, 1989. – 456 с. [Linnaeus, C. Linnaeus' Philosophia Botanica. – Oxford: Oxford University Press, 2005. – 428 p.]
3. Hamming, R.W. Error-Detecting and Error-Correcting Codes // Bell System Technical Journal. – 1950. – Vol. 29, no. 2. – P. 147–160.
4. Deza, M., Deza, E. Encyclopedia of Distances. – Berlin-Heidelberg: Springer-Verlag, 2009. – 590 p.
5. Guttman, L. A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points // Psychometrika. – 1968. – Vol. 23, no. 4. – P. 469–506.
6. Толстова Ю.Н. Основы многомерного шкалирования: учебное пособие. – М.: КДУ, 2006. – 160 с. [Tolstova Yu.N. Osnovy mnogomernogo shkalirovaniya: uchebnoe posobie. – М.: КДУ, 2006. – 160 p. (In Russian)]
7. Everitt, B.S., Landau, S., Leese, M., Stahl, D. Cluster Analysis: Fifth Edition. – Hoboken: John Wiley & Sons, 2011. – 330 p. – DOI: 10.1002/9780470977811.
8. Baccour, L. Amended Fused TOPSIS-VIKOR for Classification (ATOVIC) Applied to Some UCI Data Sets // Expert Sys-

- tems with Applications. – 2018. – Vol. 99. – P. 115–125. – DOI: 10.1016/j.eswa.2018.01.025.
9. Yusuf H., Panoutsos G. Multi-criteria Decision Making Using Fuzzy Logic and ATOVIC with Application to Manufacturing // IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). – Glasgow, UK, 2020. – P. 1–7. – DOI: 10.1109/FUZZ48607.2020.9177772.
10. Taunk, K., De, S., Verma, S., Swetapadma, A. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification // International Conference on Intelligent Computing and Control Systems (ICCS). – Madurai, India, 2019. – P. 1255–1260. – DOI: 10.1109/ICCS45141.2019.9065747.
11. Yang F.J. An Implementation of Naive Bayes Classifier // International Conference on Computational Science and Computational Intelligence (CSCI). – Las Vegas, NV, USA, 2018. – P. 301–306. – DOI: 10.1109/CSCI46756.2018.00065.
12. Коробов В.Б., Тутыгин А.Г. Классификационные методы решения эколого-экономических задач. – Архангельск: Поморский университет, 2010. – 310 с. [Korobov, V.B., Tutygin, A.G. Klassifikatsionnye metody resheniya ehkologo-ehkonomicheskikh zadach. – Arkhangel'sk: Pomorskii universitet, 2010. – 310 p. (In Russian)]
13. Безуглая Э.Ю. Мониторинг состояния атмосферы в городах. – Л.: Гидрометеоздат, 1986. – 200 с. [Bezugaia, E.Yu. Monitoring sostoyaniya atmosfery v gorodakh. – L.: Gidrometeoizdat. – 1986. – 200 p. (In Russian)]
14. Лохов А.С., Губайдуллин М.Г., Коробов В.Б., Тутыгин А.Г. Географо-экологическое районирование трассы нефтепровода по степени опасности воздействия на окружающую среду при аварийных разливах нефти в Арктике // Теоретическая и прикладная экология. – 2020. – № 4. – С. 43–48. – DOI: 10.25750/1995-4301-2020-4-045-050. [Lokhov, A.S., Gubaidullin, M.G., Korobov, V.B., Tutygin, A.G. Geographical and Ecological Land Zoning of Onshore Oil Pipeline Location by Level of Hazard to Environment from Emergency Oil Spills in Arctic Region // Theoretical and Applied Ecology. – 2020. – No. 4. – P. 43–48. (In Russian)]
15. Коробов В.Б., Кочуров Б.И., Тутыгин А.Г. Методология районирования сложных географо-экологических объектов экспертно-статистическими методами // Проблемы региональной экологии. – 2020. – № 5. – С. 42–48. [Korobov, V.B., Kochurov, B.I., Tutygin, A.G. Methodology of Zoning of Complex Geographic and Ecological Objects Using Expert Statistical Methods // Problemy regional'noi ehkologii. – 2020. – No. 5. – P. 42–48. (In Russian)]

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

Поступила в редакцию 26.07.2023,
после доработки 13.09.2023.
Принята к публикации 28.09.2023.

Коробов Владимир Борисович – д-р геогр. наук, Институт океанологии им. П.П. Ширшова РАН, г. Москва,
✉ szoioran@mail.ru,
ORCID iD: <https://orcid.org/0000-0002-3198-9208>,

Тутыгин Андрей Геннадьевич – канд. физ.-мат. наук, Федеральный исследовательский центр комплексного изучения Арктики им. академика Н.П. Лаврова Уральского отделения Российской академии наук, г. Архангельск,
✉ andgt64@yandex.ru,
ORCID iD: <https://orcid.org/0000-0001-9821-651X>,

Лохов Алексей Сергеевич – канд. геогр. наук, Институт океанологии им. П.П. Ширшова РАН, г. Москва,
✉ a.s.lohov@yandex.ru,
ORCID iD: <https://orcid.org/0000-0001-5022-9071>.

© 2023 г. Коробов В.Б., Тутыгин А.Г., Лохов А.С.



Эта статья доступна по [лицензии Creative Commons «Attribution» \(«Атрибуция»\) 4.0 Всемирная](https://creativecommons.org/licenses/by/4.0/).



A RANK-EXPERT DEVIATION FUNCTION TO CLASSIFY COMPLEX OBJECTS

V.B. Korobov¹, A.G. Tutygin², and A.S. Lokhov¹

¹Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia

²Laverov Federal Center for Integrated Arctic Research, Ural Branch, Russian Academy of Sciences, Arkhangelsk, Russia

✉ sziorian@mail.ru, ✉ andgt64@yandex.ru, ✉ a.s.lohov@yandex.ru

Abstract. This paper proposes a novel function for classifying environmental, social, and socio-environmental objects. It is based on the sum of rank deviations between a given object and a reference object considering the significance of the object's characteristics (factors). Characteristics are estimated using weight coefficients, which are provided by expertise or another method. A verbal numerical scale is developed to assess the proximity of objects by the numerical value of the deviation function. As is demonstrated below, this function is not a metric in the geometric sense but a proximity function defined in multidimensional scaling theory. As illustrative examples, the values of the deviation function are calculated for two applications: an environmental problem of comparing the vulnerability of territories to accidental oil spills and an economic problem of choosing real estate objects to purchase. A recommended sequence with a set of procedures based on the deviation function is presented to solve these problems.

Keywords: rank, object, classification, verbal numerical scale.

Funding. This work was performed within state order no. FMWE-2021-0006 “Modern and ancient bottom sediments and suspended sediments of the World Ocean—a geological record of environmental and climate changes: dispersed sedimentary matter and bottom sediments of the seas of Russia, the Atlantic, Pacific and Arctic Oceans—lithological, geochemical, and micro-paleontological studies; the research of pollution, paleoenvironments, and processes of marginal filters of rivers.”