УДК 519.81 DOi 10.30837/ bi.2019.2(93).06

К.Э. Петров[1], Т.С. Чайникова[2], И.В. Кобзев[3], В.Г. Демчук[4]

[1]Доктор технических наук, профессор кафедры искусственного интеллекта,
Харьковский национальный университет радиоэлектроники,
kostiantyn.petrov@nure.ua, ORCID iD: 0000-0003-1973-711X

[2]Менеджер по контролю качества, Компания Certent, г. Киев,
chaynikova@gmail.com, ORCID iD: 0000-0002-8923-8829

[3]Кандидат технических наук, доцент кафедры информационных технологий и систем управления,
Харьковский региональный институт государственного управления
Национальной академии государственного управления при Президенте Украины,
ikobzev12@gmail.com, ORCID iD: 0000-0002-7182-5814

[4]Аспирант кафедры электронных вычислительных машин,
Харьковский национальный университет радиоэлектроники,
vadym.demchuk@nure.ua, ORCID iD: 0000-0003-3700-2344

КОМПАРАТОРНАЯ ИДЕНТИФИКАЦИЯ МОДЕЛИ МНОГОФАКТОРНОГО ОЦЕНИВАНИЯ АЛЬТЕРНАТИВ С ИСПОЛЬЗОВАНИЕМ МЕТОДА БЭГГИНГА

Предложен один из подходов к решению задачи построения математической модели многофакторного оценивания альтернатив, который базируется на идеях теории компараторной идентификации. Рассмотрены методы ее структурной и параметрической идентификации. Структура модели определяется в рамках полинома Колмогорова-Габора, что позволяет синтезировать как традиционные аддитивные и мультипликативные, так и модели многофакторного оценивания альтернатив произвольной сложности. Для повышения точности модели предлагается использовать ансамбль моделей, агрегирование которых осуществляется на основе метода бэггинга. Приведены результаты компьютерного моделирования, подтверждающие эффективность описанного подхода. Предложенный подход открывает перспективы повышения эффективности и объективности идентификации моделей экспертного оценивания альтернатив.

МНОГОКРИТЕРИАЛЬНАЯ ОПТИМИЗАЦИЯ, ФУНКЦИЯ ПОЛЕЗНОСТИ, ОТНОШЕНИЕ ПОРЯДКА, РАНЖИРОВАНИЕ АЛЬТЕРНАТИВ, АНСАМБЛЬ МОДЕЛЕЙ

Запропоновано один з підходів до розв'язання задачі побудови математичної моделі багатофакторного оцінювання альтернатив, який базується на ідеях теорії компараторної ідентифікації. Розглянуто методи її структурної і параметричної ідентифікації. Структура моделі визначається в рамках поліному Колмогорова-Габора, що дозволяє синтезувати як традиційні адитивні і мультиплікативні, так і моделі багатофакторного оцінювання альтернатив довільної складності. Для підвищення точності моделі пропонується використовувати ансамбль моделей, агрегування яких здійснюється на основі методу беггінга. Наведено результати комп'ютерного моделювання, що підтверджують ефективність описаного підходу. Запропонований підхід відкриває перспективи підвищення ефективності та об'єктивності ідентифікації моделей експертного оцінювання альтернатив.

БАГАТОКРИТЕРІАЛЬНА ОПТИМІЗАЦІЯ, ФУНКЦІЯ КОРИСНОСТІ, ВІДНОШЕННЯ ПОРЯДКУ, РАНЖУВАННЯ АЛЬТЕРНАТИВ, АНСАМБЛЬ МОДЕЛЕЙ

One of the approaches to solving the problem of constructing a mathematical model of multifactor estimation of alternatives, which is based on the ideas of the comparative identification theory, is proposed. The methods of its structural and parametric identification are considered. A model structure is determined within the fragments of Kolmogorov-Gabor polynomial, that allows to synthesize both traditional additive and multiplicative and models of multifactor estimations of alternatives, which can have of any complication. To improve the accuracy of the model, it is proposed to use an ensemble of models, the aggregation of which is carried out on the basis of the bagging method. The results of computer modeling confirming the effectiveness of the described approach are presented. The proposed approach opens up prospects for improving the efficiency and objectivity of the identification of models for expert estimation of alternatives.

MULTICRITERIA OPTIMIZATION, FUNCTION OF UTILITY, RELATIONSHIP OF ORDER, RANKING OF ALTERNATIVES, ENSEMBLE OF MODELS

Введение

Процесс принятия решений можно определить как особый вид интеллектуальной человеческой деятельности, состоящий в выборе одного или нескольких вариантов решений (альтернатив) из некоторого имеющегося множества [1]. Одним из ключевых этапов этого процесса является оценивание альтернатив экспертом, т. к именно от него зависит эффективность принятого решения.

Альтернативы описываются в форме множеств разнородных частных характеристик, проанализировав которые эксперт должен представить свое суждение о наиболее предпочтительной или предпочтительных из них для решения имеющейся проблемы.

Основная гипотеза теории рационального поведения заключается в том, что индивидуум выбирает альтернативу, которая дает наилучший результат. Теория же полезности предполагает существование некоторой обобщенной оценки альтернативы, на основе которой и происходит этот выбор.

Таким образом, процесс оценивания, это не что иное, как присвоение альтернативам некоторых качественных или количественных оценок с целью их ранжирования по степени предпочтительности для эксперта.

Формализация процесса получения таких оценок сталкивается с огромными трудностями, так как в этом случае и объектом и субъектом исследования является эксперт, что существенно влияет на результат принятия решения. И все же, несмотря на целый ряд недостатков, неточностей и субъективизма, информация, полученная от экспертов, является практически единственным источником исходных данных.

Для получения любой информации от экспертов необходимо провести с ними серию активных или пассивных экспериментов. Этот этап присутствует в большинстве подходов, применяемых в настоящее время для получения многофакторных оценок альтернатив.

Ранжирование допустимого множества и выбор экстремального, т. е. наилучшего с точки зрения качества, единственного решения так или иначе сводится к решению задачи многокритериальной оптимизации.

Существующие в настоящее время подходы к решению такого рода задач, например, такие как метод анализа иерархий [2] и его обобщение — анализ сетей [3], методы ELECTRE [4], PROMETHEE [5], TOPSIS [6], STEM [7], методы последовательной оптимизации [7, 8], функционально-стоимостной анализ [7], метод формирования обобщенного скалярного критерия (свертки) [9] и т. п. требуют от эксперта в явном виде предоставить информацию об относительной важности частных критериев, характеризующих альтернативы или о значениях их «весовых» коэффициентов.

В отличие от этих методов, в работе предлагается один из подходов к определению относительных многофакторных оценок альтернатив в случае, когда информация, полученная от экспертов, имеет нечисловой характер, т. е. имеются данные только о наиболее предпочтительной альтернативе или установлено отношение частичного порядка на множестве альтернатив, который базируется на применении метода компараторной идентификации [1].

Для повышения точности решения задачи определения скалярных многофакторных оценок альтернатив, вместо использования одиночной модели оценивания предлагается формировать ансамбль моделей на основе применения метода бэггинга (сокр. от **b**ootstrap **agg**regat**ing**) [10].

1. Постановка задачи

Пусть имеется некоторое конечное множество альтернатив $X = \{x_1, x_2, ..., x_n\}$. Каждая из них может быть описана кортежем разнородных частных характеристик (факторов) $K(x_i) = \langle k_1(x_i), k_2(x_i), ..., k_m(x_i) \rangle$, $i = \overline{1, n}$, которые могут быть измерены в количественных шкалах.

Эти альтернативы (варианты действий) предъявляется эксперту, который оценивает их с точки зрения полезности для последующего выбора наиболее предпочтительных альтернатив при принятии решения в каждой конкретной ситуации.

Такое оценивание возможно осуществить в рамках теории полезности, основная гипотеза которой заключается в том, что предполагается существование некоторой скалярной многофакторной оценки обобщенной полезности $P(x_i)$, $i = \overline{1, n}$ для каждой из альтернатив $x_i \in X$. Эту оценку (функцию полезности) можно представить в таком виде:

$$P(x_i) = F\left[A, K(x_i)\right], \ x_i \in X, \ i = \overline{1, n}, \qquad (1)$$

где $A = \langle a_1, a_2, ..., a_t \rangle$ — кортеж значений коэффициентов, выражающих относительную важность («вес») для эксперта частных факторов $k_j(x_i)$, $j = \overline{1, m}$, которые характеризуют альтернативы $x_i \in X$.

Задача состоит в структурной, т. е. в определении вида оператора F и параметрической идентификации — нахождении значений кортежа A математической модели многофакторного оценивания (1) на основании информации только о наиболее предпочтительной альтернативе либо их частичного упорядочения, полученного в ходе анализа экспертом множества имеющихся альтернатив.

Решение этой задачи позволит провести ранжирование альтернатив исходя из их значимости для эксперта с учетом того, что по определению для функции полезности $P(x_i)$ выполняются следующие соотношения:

$$x_i \succ x_j \ \Leftrightarrow \ P(x_i) > P(x_j);$$

и $$x_i \sim x_j \ \Leftrightarrow \ P(x_i) = P(x_j), \ \forall x_i, x_j \in X, \ i \neq j. \qquad (2)$$

2. Структурная и параметрическая идентификация модели многофакторного оценивания

Без потери общности рассмотрим ситуацию когда эксперт на основании анализа представленных альтернатив произвел их упорядочение по степени предпочтительности. В общем случае, исходя из (2) получим систему из $n-1$ ограничений, где n — количество альтернатив x_i.

Например, для случая $X = \{x_1, x_2, ..., x_4\}$ и установленного отношения порядка $x_2 \succ x_3 \sim x_1 \succ x_4$ получим такую систему ограничений:

$$\begin{cases} P(x_2) > P(x_3) \\ P(x_3) = P(x_1) \\ P(x_1) > P(x_4) \end{cases} \Leftrightarrow \begin{cases} P(x_2) - P(x_3) > 0 \\ P(x_3) - P(x_1) = 0 \\ P(x_1) - P(x_4) > 0 \end{cases}. \qquad (3)$$

В качестве универсальной функции полезности $P(x_i)$ в работе предлагается использовать полином Колмогорова-Габора [11], который в принятых выше обозначениях имеет вид:

$$P(x_i) = a_0 + \sum_{j=1}^{m} a_j k_j(x_i) + \sum_{j=1}^{m}\sum_{q=1}^{m} a_{jq} k_j(x_i) k_q(x_i) + \\ + \sum_{j=1}^{m}\sum_{q=1}^{m}\sum_{r=1}^{m} a_{jqr} k_j(x_i) k_q(x_i) k_r(x_i) + \ldots . \quad (4)$$

Выбор именно полинома такого вида обусловлен следующими причинами:

– как показано в работе Колмогорова [11] полином (4) позволяет точно аппроксимировать любую функцию многих переменных;

– полином содержит в своем составе как аддитивные так и мультипликативные линейные по характеризующим факторам $K(x_i)$ составляющие и позволяет формировать на их основе полиномы любого вида;

– частные полиномы, синтезируемые на основе (4), имеют традиционную форму (аддитивную, мультипликативную и т. п.), а составляющие имеют ясную интерпретацию, как "вклад" тех или иных факторов $K(x_i)$ или их комплексов в обобщенную многофакторную оценку альтернативы.

Кроме того, следует отметить еще одно важное свойство полинома Колмогорова-Габора (4): любой полином, синтезированный как фрагмент (4) является линейным по параметрам A, но в общем случае существенно нелинейным по множеству факторов $K(x_i)$, т. е. по входу.

Для целей нашего исследования выберем фрагмент полинома (4), в котором примем $a_0 = 0$ (при нулевых значениях характеристик $k_j(x_i)$ "полезность" любой альтернативы равна нулю) и ограничимся учетом членов только первого порядка. Как показали компьютерные эксперименты [1], использование такой простейшей аддитивной модели является в данном случае оправданным из-за ограниченного объема информации полученной от эксперта в процессе выбора альтернатив. Однако это не исключает использования более сложных моделей с членами второго, третьего и более высоких порядков, учитывающих взаимовлияние частных характеристик, в случае, если удается получить больше информации от эксперта. Вопросы структурной и параметрической идентификации таких моделей на основе использования генетических алгоритмов и метода группового учета аргументов подробно рассмотрены в [12].

Таким образом, значения a_j и $P(x_i)$ будем определять в рамках фрагмента полинома (4), вида

$$P(x_i) = \sum_{j=1}^{m} a_j k_j^H(x_i), \quad (5)$$

где $K^H(x_i) = \left\langle k_1^H(x_i), k_2^H(x_i), \ldots, k_m^H(x_i) \right\rangle$ – нормированные значения частных характеристик альтернатив; a_j – безразмерные коэффициенты относительной важности нормированных частных характеристик $k_j^H(x_i)$, которые удовлетворяют условиям:

$$a_j \in [0,1], \; j = \overline{1,m}; \; \sum_{j=1}^{m} a_j = 1 . \quad (6)$$

Нормирование частных характеристик необходимо, так как, в общем случае, они имеют различные размерность, интервал изменений и направление доминирования. Это нормирование можно осуществить по формуле:

$$k_j^H(x_i) = \left(\frac{k_j(x_i) - k_j^-(x_i)}{k_j^+(x_i) - k_j^-(x_i)} \right)^{\alpha_j}, \; j = \overline{1,m}, \; i = \overline{1,n}, \quad (7)$$

где $k_j(x_i)$ – действительное (абсолютное) значение j-й частной характеристики; $k_j^-(x_i)$ и $k_j^+(x_i)$ – соответственно ее "наихудшее" и "наилучшее" значения в зависимости от направления доминирования $k_j(x_i)$; α_j – коэффициент нелинейности ($\alpha_j > 0$).

Тогда с учетом (2), (5) и (7) для случая $x_r \succ x_s$ можно записать:

$$P(x_r) > P(x_s) \Leftrightarrow P(x_r) - P(x_s) > 0 \;\; \forall x_r, x_s \in X, \; r \neq s,$$

т. е.

$$P(x_r) - P(x_s) \equiv \left(\sum_{j=1}^{m} a_j k_j^H(x_r) \right) - \left(\sum_{j=1}^{m} a_j k_j^H(x_s) \right) > 0 \quad (8)$$

или

$$\sum_{j=1}^{m} a_j \left[k_j^H(x_r) - k_j^H(x_s) \right] > 0, \quad (9)$$

где a_j, $j = \overline{1,m}$ удовлетворяют условиям (6).

Таким образом, полученная от эксперта информация в виде отношения порядка на множестве альтернатив X (например, (3)), может быть формализована в виде системы линейных ограничений вида (9), которая в свою очередь определяет выпуклый многогранник на гиперплоскости (6) любая точка которого является допустимым решением. Следовательно, задача определения коэффициентов относительной важности частных характеристик a_j является некорректной по Тихонову [13], так как в исходном виде не имеет единственного решения. Для получения такого решения исходную задачу необходимо регуляризировать [13] путем добавления в нее некоторого дополнительного условия.

В связи с отсутствием информации, позволяющей выдвинуть объективную гипотезу, определяющую вид регуляризирующей функции, в качестве рабочей примем эвристическую гипотезу, что точечная оценка параметров модели многофакторного оценивания должна находиться в центральной области многогранника допустимых значений коэффициентов, определяемых системой ограничений (6) и (9).

Для регуляризации задачи, в качестве единственного решения предлагается выбрать чебышевскую точку [14] для системы линейных ограничений (6) и (9). Экспериментальная проверка показала, что данное решение обладает высокой устойчивостью к изменению границ многогранника.

Чебышевская точка, по сути, представляет собой точку, находящуюся внутри области допустимых значений Ω кортежа A (для совместной системы линейных ограничений). Эта точка равно уклонена на некоторую величину $L<0$ от $n+1$, гиперплоскостей, ограничивающий n-мерный симплекс и уклоняется не более чем на L от остальных $m-n-1$ гиперплоскостей, где n — число переменных, а m — количество ограничений.

Поскольку все ограничения вида (6), (9), входящие в систему, являются линейными, задачу определения чебышевской точки можно свести к задаче линейного программирования (ЛП) [14].

Для примера (3) формальная запись такой задачи ЛП будет выглядеть следующим образом:

$$f(a_1,a_2,\ldots,a_m,L)\equiv L\to\min$$

$$\begin{cases}\sum\limits_{j=1}^{m}a_j\left[k_j^H(x_2)-k_j^H(x_3)\right]+L\ge 0\\ \sum\limits_{j=1}^{m}a_j\left[k_j^H(x_3)-k_j^H(x_1)\right]+L\ge 0\\ \sum\limits_{j=1}^{m}a_j\left[-k_j^H(x_3)+k_j^H(x_1)\right]+L\ge 0\\ \sum\limits_{j=1}^{m}a_j\left[k_j^H(x_1)-k_j^H(x_4)\right]+L\ge 0\\ a_j+L\ge 0,\ j=\overline{1,m}\\ \sum\limits_{j=1}^{m}a_j=1\end{cases}\qquad(10)$$

В системе линейных ограничений (10) знаки «>» изменены на «$\ge$» и равенство $P(x_3)-P(x_1)=0$ представлено в виде двух неравенств $P(x_3)-P(x_1)\ge 0$ и $-P(x_3)+P(x_1)\ge 0$. Замена знаков ограничений не влияет на решение так как оно лежит внутри многогранника описываемого системой (10), а не на его границах.

В результате решения задачи (10), например симплекс-методом, получим точечные значения кортежа «весовых» коэффициентов частных характеристик A, а также относительных многофакторных оценок альтернатив $P(x_i)$, $i=\overline{1,n}$, вычисленных на основе A. Это позволяет провести ранжирование альтернатив исходя из полученных значений $P(x_i)$, $i=\overline{1,n}$.

3. Использование бэггинга для повышения точности модели многофакторного оценивания альтернатив

Точность модели является одним из важнейших ее свойств. Повысить точность модели можно при помощи построения ансамбля моделей, т. е. набора моделей, «работающих» над решением одной и той же задачи. В частности, использование метода построения ансамбля моделей основанного на применении бэггинга дает существенно более точный результат, чем применение одиночной модели на имеющемся наборе данных.

Рассмотрим метод бэггинга более подробно. Подобно тому, как усреднение нескольких наблюдений снижает оценку дисперсии данных, так и возможное снижение дисперсии прогноза можно достичь путем извлечения большого количества порций данных (обучающих выборок) из генеральной совокупности, построения предсказательной модели по каждой обучающей выборке и последующее усреднение полученных предсказаний. Если вместо отдельных обучающих выборок (которых, как правило, всегда не хватает) применить процедуру бутстрепа (bootstrap) [15], которая заключается в многократной генерации выборок методом Монте-Карло на базе имеющейся выборки и на основе сгенерированных псевдовыборок построить N моделей регрессии $f_i(x)$, $i=\overline{1,N}$, то средний коллективный прогноз $f_{\text{ср}}(x)=(f_1(x)+f_2(x)+\ldots+f_N(x))/N$ будет обладать более низкой дисперсией. Эта процедура и называется бэггингом [15].

Для повышения точности модели оценивания предлагается использовать следующий алгоритм на основе использования метода бэггинга.

1. Пусть имеется множество $X=\{x_1,x_2,\ldots,x_n\}$ состоящее из n альтернатив и задано отношение порядка на этом множестве. Из множества X случайным образом отбирается несколько подмножеств $X_1\in X$, $X_2\in X$, … , $X_p\in X$, количество элементов в которых $m<n$ (общее количество этих подмножеств равно C_n^m). На каждом из подмножеств $X_q\in X$, $q=\overline{1,p}$ сохраняется отношение порядка альтернатив установленного для исходного множества X.

2. На основе каждого подмножества $X_q\in X$, $q=\overline{1,p}$ строится математическая модель многофакторного оценивания альтернатив (1) в соответствии с подходом описанным выше. В результате получаем значения функций полезности (выходы модели) $P_q(x_i)$, для каждой альтернативы $x_i\in X$, $i=\overline{1,n}$ по каждой из моделей построенной исходя из $X_q\in X$, $q=\overline{1,p}$. Таким образом получаем ансамбль моделей.

3. Находим среднее значение функции полезности $P_{\text{ср}}(x_i)$ для каждой альтернативы $x_i\in X$, $i=\overline{1,n}$ по всем моделям соответствующим $X_q\in X$, $q=\overline{1,p}$ (метод бэггинга) по формуле:

$$P_{\text{ср}}(x_i)=(P_1(x_i)+P_2(x_i)+\ldots+P_p(x_i))/p,\ i=\overline{1,n}.\qquad(11)$$

4. Проводим ранжирование альтернатив в соответствии с полученными средними значениями их функций полезности $P_{\text{ср}}(x_i)$, $i=\overline{1,n}$.

Ожидается, что результат ранжирования полученный путем агрегации моделей (построения ансамбля моделей) будет намного точнее результата ранжирования по одиночной модели при тех же исходных данных.

4. Имитационный эксперимент

Для иллюстрации работоспособности предлагаемого метода рассмотрим следующий абстрактный пример.

Пусть имеется десять альтернатив, каждая из которых характеризуется семью частными характеристиками.

Сгенерируем значения нормализованных частных характеристик $K^H(x_i) = \langle k_j^H(x_i) \rangle$, $j = \overline{1,7}$ альтернатив $X = \{x_i\}$, $i = \overline{1,10}$ с помощью датчика случайных чисел. Все эти данные представлены в таблице 1.

Также случайным образом сгенерируем порядок следования альтернатив по убыванию их полезности для эксперта.

Пусть получено следующее отношение линейного порядка:

$$x_9 \succ x_{10} \succ x_4 \succ x_7 \succ x_5 \succ x_3 \succ x_1 \succ x_6 \succ x_2 \succ x_8 . \quad (12)$$

В качестве функции полезности альтернатив x_i, $i = \overline{1,10}$ выберем полином вида:

$$P(x_i) = a_1 k_1^H(x_i) + a_2 k_2^H(x_i) + \ldots + a_7 k_7^H(x_i), \quad (13)$$

где a_j, $j = \overline{1,7}$ удовлетворяют условиям (6).

В соответствии с (2), (12) и (13) получим следующую систему линейных неравенств:

$$\begin{cases} P(x_9) - P(x_{10}) \ge 0 \\ P(x_{10}) - P(x_4) \ge 0 \\ P(x_4) - P(x_7) \ge 0 \\ P(x_7) - P(x_5) \ge 0 \\ P(x_5) - P(x_3) \ge 0 \\ P(x_3) - P(x_1) \ge 0 \\ P(x_1) - P(x_6) \ge 0 \\ P(x_6) - P(x_2) \ge 0 \\ P(x_2) - P(x_8) \ge 0 \end{cases} \quad (14)$$

Убедимся, что существует такой полином вида (13), который «аппроксимирует» отношение порядка (12) установленного на множестве альтернатив $X = \{x_i\}$, $i = \overline{1,10}$.

Для этого находим чебышевскую точку для системы ограничений (14), т. е. значения $A^* = \langle a_j^* \rangle$, $j = \overline{1,7}$ в соответствии с методом описанным выше (см. (10)). Все вычисления проводились с использованием программного средства Mathcad v.14.0.

Были получены следующие значения: $A^* = \langle 0{,}10; 0{,}19; 0{,}14; 0{,}16; 0{,}17; 0{,}13; 0{,}11 \rangle$ и $L = -0.02$.

Далее, в соответствии с найденными значениями $\langle a_j^* \rangle$, $j = \overline{1,7}$, по формуле (13) вычислим модельные значения функции полезности для каждой из альтернатив $x_i \in X$, т. е. $P^*(x_i)$, $i = \overline{1,10}$. Таким образом, мы получили относительные обобщенные скалярные оценки альтернатив (столбец $P^*(x_i)$ в табл. 1), на основании которых можно произвести их ранжирование по убыванию полезности (столбец R^* в табл. 1).

Следовательно, полученный полином $P^*(x_i)$ со значениями «весовых» коэффициентов A^*, который «аппроксимирует» отношение порядка (12) существует.

Разделим множество альтернатив $X = \{x_i\}$, $i = \overline{1,10}$ на обучающее и проверочное. Пусть обучающее подмножество сдержит альтернативы x_1, x_5, x_6, x_7, x_8, x_9, x_{10}, а проверочное – x_2, x_3, x_4.

В этом случае в соответствии с (12) имеем следующую информацию для проверочного подмножества:

$$Z0: x_9 \succ x_{10} \succ x_7 \succ x_5 \succ x_1 \succ x_6 \succ x_8 . \quad (15)$$

На основе этой информации определяем значения коэффициентов полинома $P_0(x_i)$, который «аппроксимирует» отношение порядка (15) аналогично методу описанному выше. Значения функции полезности для каждой из альтернатив $x_i \in X$, $i = \overline{1,10}$, в том числе и для проверочного подмножества, представлены в таблице 2 (столбец $P_0(x_i)$, а их соответствующие ранги – в таблице 3 (столбец R_0).

Для оценки точности полученного ранжирования альтернатив по отношению к исходному R^* используем коэффициент ранговой корреляции Спирмена [15]:

$$\rho_t = 1 - \frac{6 \cdot \sum_{i=1}^{n} (R_i^* - R_{ti})^2}{n \cdot (n^2 - 1)}, \quad (16)$$

где R_i^* и R_{ti} – ранги (места) в исходном отношении порядка альтернатив и в соответствующих установленных модельных отношениях порядка $Z0$, $Z1$, ... Zt; n – количество альтернатив $x_i \in X$.

Таким образом, анализируя полученное в соответствии с отношением порядка $Z0$ (15) ранжирование

Таблица 1

Значения нормализованных частных характеристик альтернатив

	$k_1^H(x_i)$	$k_2^H(x_i)$	$k_3^H(x_i)$	$k_4^H(x_i)$	$k_5^H(x_i)$	$k_6^H(x_i)$	$k_7^H(x_i)$	$P^*(x_i)$	R^*
x_1	0,00	0,29	0,82	0,16	0,46	1,00	0,62	0,4719	***7***
x_2	0,18	0,71	0,64	0,32	0,25	0,51	0,24	0,4289	***9***
x_3	0,49	0,00	0,70	0,15	0,79	0,63	1,00	0,4972	***6***
x_4	0,39	0,60	0,62	0,58	0,48	0,46	0,83	0,5653	***3***
x_5	0,13	0,84	0,00	0,73	0,73	0,42	0,51	0,5242	***5***
x_6	0,52	0,50	0,94	0,18	0,58	0,37	0,00	0,4541	***8***
x_7	0,68	0,59	0,79	1,00	0,00	0,53	0,24	0,546	***4***
x_8	0,35	1,00	0,08	0,21	0,57	0,00	0,25	0,3942	***10***
x_9	1,00	0,53	0,81	0,62	0,77	0,40	0,42	0,6424	***1***
x_{10}	0,60	0,45	1,00	0,00	1,00	0,48	0,64	0,5883	***2***

альтернатив R_0 можно заметить значительное расхождение с исходным ранжированием R^* (см табл. 3 столбцы R^* и R_0). Об этом говорит и значение коэффициента Спирмена $\rho_0 = 0,87$ (см. табл. 3 строка ρ_t столбца R_0).

Воспользуемся методом бэггинга для повышения точности предлагаемой модели многофакторного оценивания.

Как было отмечено выше основная идея бэггинга – усреднить некоторое множество «шумных» и несмещённых моделей для создания модели с низким расхождением в рамках аппроксимации.

Сгенерируем случайным образом на основе отношения порядка (15) шесть неполных отношений, содержащих только пять альтернатив (общее количество таких отношений равно $C_7^5 = 21$):

$$Z1: x_9 \succ x_7 \succ x_1 \succ x_6 \succ x_8;$$

$$Z2: x_{10} \succ x_7 \succ x_5 \succ x_1 \succ x_8;$$

$$Z3: x_7 \succ x_5 \succ x_1 \succ x_6 \succ x_8;$$

$$Z4: x_9 \succ x_{10} \succ x_7 \succ x_5 \succ x_6;$$

$$Z5: x_9 \succ x_{10} \succ x_7 \succ x_5 \succ x_1;$$

$$Z6: x_{10} \succ x_7 \succ x_5 \succ x_1 \succ x_6.$$

Затем, с помощью метода описанного выше, проведем параметрическую идентификацию моделей многофакторного оценивания для каждого из отношений $Z1$, $Z2$, ..., $Z6$, вычислим значения относительных полезностей альтернатив исходя из полученных моделей (см. табл. 2 столбцы $P_1(x_i)$, $P_2(x_i)$, ... , $P_6(x_i)$) и проведем их ранжирование в порядке убывания этих значений (см. табл. 3 столбцы, R_1, R_2, ... , R_6). В строке ρ_t таблицы 3 приведены соответствующие значения коэффициентов ранговой корреляции для исходного порядка альтернатив R^* и соответствующих ранжирований R_1, R_2, ... , R_6.

Далее усредняем полученные значения относительных полезностей альтернатив по формуле

$$P_{\text{ср}}(x_i) = \frac{\sum_{q=1}^{6} P_q(x_i)}{6}$$

и проводим ранжирование альтернатив исходя из этих усредненных величин.

Полученные результаты представлены в соответствующих столбцах $P_{\text{ср}}(x_i)$ таблицы 2 и $R_{\text{ср}}$ табл. 3.

Таблица 3

Значения рангов альтернатив

	R^*	R_0	R_1	R_2	R_3	R_4	R_5	R_6	$R_{\text{ср}}$
x_1	***7***	7	6	8	5	7	10	7	7
x_2	***9***	9	9	9	8	10	9	10	9
x_3	***6***	2	2	5	6	2	5	2	5
x_4	***3***	3	3	4	3	4	4	3	3
x_5	***5***	6	7	7	4	6	7	6	6
x_6	***8***	8	8	6	9	8	6	8	8
x_7	***4***	5	5	3	1	5	3	5	4
x_8	***10***	10	10	10	10	9	8	9	10
x_9	***1***	1	1	1	2	1	1	1	1
x_{10}	***2***	4	4	2	7	3	2	4	2
ρ_t	-	**0,87**	0,84	0,93	0,75	0,87	0,86	0,86	0,99

Таким образом, исходя из полученных результатов, можно сделать вывод, что ансамбль более простых моделей с использованием метода бэггинга дает существенное улучшение результатов по сравнению с более «полной» и сложной моделью многофакторного оценивания построенной на основе отношения (15).

Заключение

Результаты компьютерного имитационного моделирования показали, что при использовании одинаковой информации о предпочтительности альтернатив для эксперта метод с использованием ансамбля моделей на основе бэггинга более эффективен при построении модели многофакторного оценивания

Таблица 2

Значения функций полезности альтернатив

	$P_0(x_i)$	$P_1(x_i)$	$P_2(x_i)$	$P_3(x_i)$	$P_4(x_i)$	$P_5(x_i)$	$P_6(x_i)$	$P_{\text{ср}}(x_i)$
x_1	0,4236	0,4715	0,4135	0,4531	0,4076	0,3557	0,3888	0,4150
x_2	0,3357	0,3691	0,3649	0,3958	0,3262	0,3605	0,3058	0,3537
x_3	0,5966	0,5677	0,5384	0,4467	0,6108	0,5143	0,6187	0,5494
x_4	0,5936	0,5612	0,5396	0,5578	0,5858	0,5205	0,5945	0,5599
x_5	0,4807	0,4475	0,4818	0,5384	0,438	0,4411	0,4500	0,4661
x_6	0,3670	0,3865	0,4825	0,3656	0,3596	0,4741	0,3288	0,3995
x_7	0,5297	0,5533	0,5472	0,663	0,5168	0,5365	0,5158	0,5554
x_8	0,3183	0,2989	0,3464	0,276	0,3296	0,4075	0,3218	0,3300
x_9	0,6594	0,6425	0,7168	0,6147	0,6682	0,7329	0,6646	0,6733
x_{10}	0,5778	0,5593	0,6219	0,4281	0,5896	0,6203	0,5702	0,5649

чем метод прямого построения модели, учитывающий всю информацию об отношении порядка установленного на множестве альтернатив. Сравнение отношений порядка полученных с помощью этих моделей – R_0 и $R_{ср}$ с исходным R^* показывает, что в первом случае количество инверсий рангов альтернатив $x_i \in X$, $i = \overline{1,10}$ равно четырем ($\rho_0 = 0,87$), а во втором – двум ($\rho_{ср} = 0,99$).

Следует также отметить, что использование в ансамбле малого числа моделей ухудшает итоговый результат. В частности, для данного примера при использовании четырех и пяти моделей $\rho = 0,93$, а для трех – $\rho = 0,87$.

При построении ансамбля состоящего из моделей, учитывающих отношения предпочтения из трех альтернатив (общее количество таких отношений для приведенного примера равно $C_7^3 = 35$), аналогичный результат ($\rho = 0,99$) был достигнут при использовании восьми случайно отобранных моделей. Для ансамбля из семи моделей $\rho = 0,98$, для шести – $\rho = 0,96$.

В перспективе необходимо провести исследование зависимости точности итоговой модели от количества участвующих в ансамбле моделей и их «простоты» с целью выработки возможных рекомендаций относительно выбора этих двух параметров или их адаптивной настройки в зависимости от имеющихся исходных данных, т. е. количества альтернатив и факторов, которые их характеризуют.

Предложенный в работе метод целесообразно использовать в случае когда объем исходных данных невелик, что чаще всего характеризует ситуации экспертного оценивания или принятия решений. В этом случае применение аппарата искусственных нейронных сетей для решения такого рода задач будет неэффективным из-за существенной ограниченности количества исходных данных, которые могут быть использованы для их обучения.

Также представленный подход можно использовать для выработки некоторого согласованного мнения экспертов, т. к. ансамбль моделей можно строить и для отношений порядка, которые получены от экспертов, различной длины («полноты»), т. е. один эксперт – одна модель в ансамбле, а результат работы ансамбля – их согласованное мнение (ранжирование альтернатив в соответствии с их усредненной полезностью).

Список литературы:

[1] *Петров К.Э.* Компараторная структурно-параметрическая идентификация моделей скалярного многофакторного оценивания / К.Э. Петров, В.В. Крючковский. – Херсон: Олди-плюс, 2009. – 294 с.

[2] *Саати Т.* Принятие решений. Метод анализа иерархий / Т. Саати. – М.: Радио и связь, 1993. – 278 с.

[3] *Саати Т.Л.* Принятие решений при зависимостях и обратных связях. Аналитические сети / Т. Л. Саати. – М.: Книжный дом «ЛИБРОКОМ», 2011. – 360 с.

[4] *Figueira J., Mousseau V., Roy B.* ELECTRE methods / J. Figueira, S. Greco, M. Ehrgott (eds.), Multiple Criteria Decision Analysis: The State of the Art Surveys. – New York: Springer Science+Business Media, Inc., 2005. – P. 133–162.

[5] *Corrente S.* Multiple criteria hierarchy process with ELECTRE and PROMETHEE / S. Corrente, S. Greco, R. Slowinski // Omega. – 2013. – Vol. 41, Issue 5. – P. 820–846.

[6] *Bilbao-Terol A.* Using TOPSIS for assessing the sustainability of government bond funds / A. Bilbao-Terol, M. Arenas-Parra, V. Canal-Fernandez, J. Antomil-Ibias. // Omega. – 2014. – 49. – P. 1–17.

[7] *Подиновский В.В.* Парето-оптимальные решения многокритериальных задач / В.В. Подиновский, В.Д. Ногин. – М. : Наука, 1982. – 254 с.

[8] *Ларичев О.И.* Теория и методы принятия решений, а также хроника событий в волшебной стране / О.И. Ларичев. – М. : Логос, 2000. – 294 с.

[9] *Кини Р. Л.* Принятие решений при многих критериях: предпочтения и замещения / Р. Л. Кини, Х. Райфа. – М.: Радио и связь, 1981. – 560 с.

[10] *Breiman L.* Bagging predictors / L. Breiman // Machine Learning. – 1996. – Vol. 24, , Issue 2. – P. 123–140.

[11] *Колмогоров А.Н.* О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения / А.Н. Колмогоров // Доклады АН СССР. – 1957. – Т. 5(114). – С. 953–956.

[12] *Овезгельдыев А.О.* Построение модели индивидуального многофакторного оценивания с применением элементов МГУА и генетических алгоритмов / А.О. Овезгельдыев, К.Э. Петров // Кибернетика и системный анализ. – 2007. – №1. – С. 151–159.

[13] *Тихонов А.Н.* Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. – М.: Наука, 1986. – 288 с.

[14] *Зуховицкий С.И.* Линейное и выпуклое программирование / С.И. Зуховицкий, Л.И. Авдеева. – М.: Наука, 1967. – 460 с.

[15] *Брюс П.* Практическая статистика для специалистов Data Science / П. Брюс, Э. Брюс. – С.-Пб.: БХВ-Петербург, 2016. – 304 с.

Поступила в редколлегию 16.09.2019