

Построение устойчивых интегральных индикаторов

Казакова Т.В. Стрижов В.В. Шакин В.В.

(Москва)

При построении интегральных индикаторов большого количества объектов появляется проблема их адекватной сравнимости, так как некоторые отдельные объекты имеют значения показателей, существенно отличающиеся от показателей основного числа объектов. Это означает, что в рамках линейной модели зависимости интегральных индикаторов объектов от значений их показателей объекты-выбросы имеют большее влияние на веса показателей, чем прочие объекты. Существуют два способа решения этой проблемы: исключение подобных объектов из выборки и разбиение множества объектов на несколько классов, внутри которых производится сравнение. Часто эти способы неприемлемы.

Предлагается нелинейное монотонное отображение значений измеряемых показателей с целью получения устойчивого интегрального индикатора. Это отображение необходимо при анализе сложных структур данных, например, данных, соответствующих элементам декартова произведения множеств «объекты-показатели-отсчеты времени» или «объекты-показатели-местоположения».

Устойчивые интегральные индикаторы были построены в рамках программы социально-гигиенического мониторинга Минздрава РФ. Минздравом осуществляется анализ данных по содержанию контаминантов в продуктах питания. Цели и задачи исследования: 1) выявление продуктов питания, представляющих наибольшую потенциальную опасность для здоровья населения; 2) выявление контаминантов, являющихся приоритетными загрязнителями продуктов питания; 3) выявление экологически неблагополучных по качеству продаваемых продуктов питания субъектов РФ.

Предоставленные экспертами данные имеют структуру «объекты-показатели-местоположения», то есть задан набор матриц $\mathcal{D} = \{D_1, \dots, D_l \mid D_i \in R^{m \times n}\}$, каждый элемент d_{ijk} которого поставлен в соответствие элементу декартова произведения трех множеств этой структуры. Требуется получить три вектора – интегральных индикатора – для каждого из множеств, используя модель линейной зависимости индикаторов от перекодированных значений показателей.

Для приведения показателей к единой шкале в [1] было предложено отображение отрезка $[\min_i d_{ijk}, \max_i d_{ijk}]$ в отрезок $[0, 1]$. Рассмотрим параметрическое отображение ρ из фиксированного множества

отображений: билинейное отображение и функции распределения равномерного, нормального и экспоненциального распределений.

Выбор конкретного вида функции определяется характером распределения исходных данных.

Для оценки параметров введем функционал качества s . Используем алгоритм скользящего контроля.

Рассмотрим отображение $D \rightarrow A$, где $D = \langle \mathbf{d}_1, \dots, \mathbf{d}_n \rangle$ – матрица исходных данных и $A = \langle \mathbf{a}_1, \dots, \mathbf{a}_n \rangle$ – матрица перекодированных данных.

Вектор-столбец \mathbf{a}_j есть результат отображения $\mathbf{a}_j = \rho_j(\theta, \mathbf{d}_j)$, где $\rho_j \in P$ и θ – вектор параметров функции ρ_j . Параметры оцениваются как $\Theta = \arg \min_{\theta \in B} s(\Theta, Q)$. Вектор Θ есть присоединенные векторы

$\langle \theta_1, \dots, \theta_n \rangle$, Q – матрица интегральных индикаторов, полученная следующим образом. Пусть \mathbf{q}_i есть интегральный индикатор объектов, описанных матрицей A при удалении строки, описывающей i -ый объект. Тогда $Q = \langle \mathbf{q}_1, \dots, \mathbf{q}_n \rangle$. Значения индикаторов удаленных объектов на диагонали матрицы заменяются средним значением по строкам. Функционал качества интегрального индикатора $s = \lambda_1$, где $\lambda_1, \dots, \lambda_r$ – сингулярные числа в разложении $Q = U \Lambda V^T$. Область поиска B параметров задана.

Вычисление интегрального индикатора происходит на основе обобщения метода сингулярного разложения матрицы по следующей схеме: выбираем максимальный по норме вектор \mathbf{a}_1 длины n и вектор \mathbf{b}_1 длины m , записанный в виде строки. i -ая координата вектора \mathbf{c}_1 вычисляется по

формуле $c_1^{(i)} = \frac{(A_i, \mathbf{b}_1^T \mathbf{a}_1)}{(\mathbf{b}_1^T \mathbf{a}_1, \mathbf{b}_1^T \mathbf{a}_1)}$, где $A_i \in R^{m \times n}$, $i = \overline{1, l}$ – матрицы в

пространстве объекты-показатели. Определим скалярное произведение матриц следующим образом $X = \{x_{ij}\}_{i,j=1}^{m,n}$, $Y = \{y_{ij}\}_{i,j=1}^{m,n}$,

$(X, Y) = \sum_{i,j=1}^{m,n} x_{ij} y_{ij}$. Далее

$$a_k^{(i)} = \frac{(A_i, \mathbf{b}_{k-1}^T \mathbf{c}_{k-1})}{(\mathbf{b}_{k-1}^T \mathbf{c}_{k-1}, \mathbf{b}_{k-1}^T \mathbf{c}_{k-1})},$$

$A_i \in R^{m \times l}, i = \overline{1, n}$ – матрицы в пространстве объекты-местоположения;

$$b_k^{(i)} = \frac{(A_i, \mathbf{c}_{k-1}^T \mathbf{a}_k)}{(\mathbf{c}_{k-1}^T \mathbf{a}_k, \mathbf{c}_{k-1}^T \mathbf{a}_k)},$$

$A_i \in R^{l \times n}, i = \overline{1, m}$ – матрицы в пространстве местоположения-показатели.

Векторы $\mathbf{a}, \mathbf{b}, \mathbf{c}$ определяются как пределы соответствующих последовательностей $\mathbf{a} = \lim \mathbf{a}_k, \mathbf{b} = \lim \mathbf{b}_k, \mathbf{c} = \lim \mathbf{c}_k$. В качестве интегральных индикаторов множеств признаки, объекты, местоположения используются векторы $\mathbf{u}, \mathbf{v}, \mathbf{w}$:

$$\mathbf{u} = \frac{\mathbf{a}}{\|\mathbf{a}\|}, \mathbf{v} = \frac{\mathbf{b}}{\|\mathbf{b}\|}, \mathbf{w} = \frac{\mathbf{c}}{\|\mathbf{c}\|}.$$

Предлагаемый метод предварительной обработки данных, учитывая распределения исходных объектов по показателям, уменьшает влияние объектов с существенно отличными значениями показателей. Данное преобразование позволяет получить индикаторы объектов, устойчивые к появлению объектов-выбросов. Обобщение метода сингулярного разложения матрицы дает возможность строить интегральные индикаторы для данных со структурой «объекты-показатели-отсчеты времени» или «объекты-показатели-местоположения».

Данная работа поддержана грантом РФФИ 04-01-00401-а.

Литература

1. Айвазян С.А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении и межрегиональных сопоставлениях. – М.: ЦЭМИ РАН, 2000. – С. 56.
2. Strijov, V., Shakin, V. Index construction: the expert-statistical method. Environmental research, engineering and management. 2003. No. 4(26). P.51-55. ISSN 1392-1649.