Устойчивые интегральные индикаторы с выбором нормирующих функций

Стрижов В.В., Казакова Т.В.

Вычислительный центр РАН

e-mail: `strijov@ccas.ru`

Исследуется задача построения интегрального индикатора «без учителя» [1], устойчивого к изменениям множества описаний объектов. Объекты описаны в линейных шкалах. При построении интегрального индикатора выбираются такое опорное множество объектов и монотонные преобразования, которые доставляют максимум критерию устойчивости.

Задано множество из m объектов, которые описаны набором из n показателей. Задана матрица описаний $A \in \mathbb{R}^{m\times n}$. Элемент матрицы a_{ij} — значение j-го показателя i-го объекта. Вектор $\mathbf{a}_{i\cdot}$ — описание i-го объекта, а вектор $\mathbf{a}_{\cdot j}$ — значения j-го показателя.

Интегральный индикатор объекта — свертка вида $q_i = \sum_{j=1}^{n} w_j g_j(a_{ij})$, где $g_i : \mathbb{R} \longrightarrow [0,1]$ — гладкая монотонная функция. Без ограничения общности будем считать, что выполнено условие монотонности индикатора, такое, что из $a_{ij} \geqslant a_{\xi j}$ следует $q_i \geqslant q_\xi$ для $j \in \{1,...,n\}$. Это влечет неотрицательность значений $w_1,...,w_n$. Так как на практике часто выставляется требование инвариантности интегрального индикатора к линейным преобразованиями, введем еще одно условие, накладываемое на веса: $\sum_{j=1}^{n} w_j^2 = 1$. Считаем также, что значения элементов a_{ij} удовлетворяют равенству $a_{ij} = 1 - |a_{ij} - a_j^{\text{opt}}|(\max(\max \mathbf{a}_{\cdot j} - a_j^{\text{opt}}, a_j^{\text{opt}} - \min \mathbf{a}_{\cdot j}))^{-1}$.

Базовым методом получения интегрального индикатора считаем метод главных компонент [2]. Для нахождения первой главной компоненты требуется найти такие линейные комбинации $Z^T = CA^T$

векторов-строк матрицы A, что векторы-столбцы $\mathbf{z}_{\cdot 1},...,\mathbf{z}_{\cdot n}$ матрицы Z обладали бы наибольшей дисперсией: $\max\sum_{j=1}^{n} D\mathbf{z}_{\cdot j}$ при ограничениях нормировки $\|C\|^2=1$ и $CC^T=I_n$. Значение интегрального индикатора вычисляется как проекция векторов-строк матрицы A на первую главную компоненту, $\mathbf{q}_{\text{PCA}}=A\mathbf{c}_{\cdot 1}$.

Обозначим $\mathcal{S}=\{S_1,...,S_l\}$ — множество всех подмножеств $\{\mathbf{a}_{1\cdot},...,\mathbf{a}_{m\cdot}\}$ и $\mathcal{Q}=\{\mathbf{q}_1,...,\mathbf{q}_l\}$, $\mathcal{W}=\{\mathbf{w}_1,...,\mathbf{w}_l\}$ — множества соответствующих им интегральных индикаторов и весов показателей, $l=2^m$. Алгоритм поиска получает множество S_ξ, вычисляет наиболее информативный линейный предиктор $\mathbf{w}_\xi$ и возвращает индикатор $\mathbf{q}_\xi=A\mathbf{w}_\xi$. Устойчивый индикатор $\hat{\mathbf{q}}$ имеет опорным множество $\hat{S}$ максимальной мощности при минимальном расстоянии от устойчивого центра векторов $\bar{\mathbf{q}}\in\mathcal{Q}$, т.е. $\hat{\mathbf{q}}:\#\hat{S}\longrightarrow\max, \|\bar{\mathbf{q}}-\hat{\mathbf{q}}\|\longrightarrow\min$.

Введем монотонную корекцию всей выборки. Выберем из заданного множества $\{g_1,...,g_r\}$ функции, доставляющие минимум выражению $\varepsilon=\sum_{i=1}^{m}\left(\hat{q}_i-\sum_{j=1}^{n}w_j g_{k(j)}(\theta_{k(j)}a_{ij})\right)^2$. Функция $k(j):\{1,...,r\}\longrightarrow\{1,...,n\}$. Для выбора функции k и параметров $\theta_{k(j)}$ используем двухшаговый алгоритм, включающий метод доверительных интервалов для оптимизации параметров и генетический оптимизационный алгоритм для выбора модельных функций.

Предложенный алгоритм построения устойчивых интегральных индикаторов является альтернативой алгоритмам, включающим регуляризацию: влияние объектов-выбросов на интегральный индикатор исключено. Его уместно использовать, когда описание объекта не имеет модели плотности распределения.

Работа поддержана грантом РФФИ 04-01-00401.

Список литературы

1. *Strijov, V., Shakin, V.* Index construction: the expert-statistical method. Environmental research, engineering and management. 2003. No.4(26), P.51-55.
2. *Jolliffe, I. T.* Principal Component Analysis, 2nd ed., Springer, 2002.