

МНОГОКРИТЕРИАЛЬНАЯ КЛАССИФИКАЦИЯ ПО ПАРЕТО КАК МЕХАНИЗМ ГРУППИРОВКИ ПРИЗНАКОВ В ПОСТРОЕНИИ ИНТЕГРАЛЬНЫХ ИНДИКАТОРОВ КАЧЕСТВА ЖИЗНИ НАСЕЛЕНИЯ.

Мироненков А.А. (Москва)ⁱ

Работа с данными при построении интегральных индикаторов качества жизни населения в подходах «без обучения», как правило, состоит из следующих этапов:

- Этап 1. Определение числа интегральных индикаторов, необходимого для характеристики анализируемой синтетической категории;
- Этап 2. Разбиение анализируемого набора частных критериев на непересекающиеся блоки сильно связанных переменных;
- Этап 3. Построение блочных интегральных индикаторов отдельно по частным критериям в каждом из блоков;
- Этап 4. Построение сводного ИИ, характеризующего анализируемую синтетическую категорию КЖ.

Задачи второго этапа (разбиение переменных на блоки) производятся либо экспертно, руководствуясь содержательными соображениями (например, Жгун Т.В., 2017), либо с использованием процедуры экстремальной группировки признаков (Айвазян, Мхитарян, 2001, п.13.4.2). При этом первый путь требует привлечения экспертов и плохо соотносится с подходом «без обучения». Второй путь предполагает максимизацию по разбиению S_1, S_2, \dots, S_k функционала

$$F = \sum_{x \in S_1} r^2(x, f^1) + \sum_{x \in S_2} r^2(x, f^2) + \dots + \sum_{x \in S_k} r^2(x, f^k),$$

где r – это коэффициент корреляции, а f^i – первая главная компонента i -того блока (Айвазян, 2012, стр. 99). Вычисление главных компонент предполагает поиск собственных векторов, что само по себе является ресурсоёмкой задачей. Нам требуется перебор всех возможных разбиений переменных и для каждого вычислить главную компоненту. При увеличении числа наблюдений при достаточно большом количестве переменных исследователь неизбежно сталкивается с невозможностью вычисления функционала F ввиду ограниченности вычислительных ресурсов.

В статье (Мироненков, 2020) отмечено два полезных свойства алгоритма классификации по Парето: он нетребователен к вычислительным ресурсам и количество получаемых Парето-классов прямо зависит от степени «согласованности» переменных. Рассмотрим случай двух переменных. Если они сильно коррелированы, то нарезка классов получается дробная, и количество классов сравнивается с количеством наблюдений. Если же переменные противоположно направлены, то среди наблюдений нельзя указать

доминирующие, и в этом случае все наблюдения будут отнесены в единственный Парето-класс.

Таким образом, автором предлагается решить поставленную на Этапе 2 задачу распределения переменных по блокам путём максимизации по всем возможным разбиениям S_1, S_2, \dots, S_k функционала

$$F = N(S_1) + N(S_2) + \dots + N(S_k),$$

где $N(S_i)$ число Парето-классов, вычисленных по переменным i -того блока.

Литература

Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Т.1: Теория вероятностей и прикладная статистика. М.: ЮНИТИ-ДАНА, 2001.

Айвазян С.А. Анализ качества и образа жизни населения: эконометрический подход. М.: Наука, 2012.

Жгун Т.В. Построение интегральной характеристики качества жизни субъектов Российской Федерации с помощью метода главных компонент / Т.В. Жгун // Экономические и социальные перемены: факты, тенденции, прогноз. 2017. Т. 10. № 2. С. 214–235. – DOI: 10.15838/esc/2017.2.50.12.

Мироненков А.А. Иерархическая Парето-классификация регионов России по показателям качества жизни населения. Отправлена в редакцию. 2020.

ⁱ **Мироненков Алексей Алексеевич** – МГУ имени М.В. Ломоносова,
Alexey.mironenkov@gmail.com