

ОЦЕНКА ДОСТОВЕРНОСТИ КЛАСТЕРИЗАЦИИ СУБЪЕКТОВ РОССИЙСКОЙ ФЕДЕРАЦИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МОНТЕ-КАРЛО

Кирилюк И.Л., Сенько О.В.¹

Институт экономики РАН, Россия, 117218, Москва, Нахимовский проспект 32,
+7(926)557-92-75, igokir@rambler.ru

¹Федеральный исследовательский центр "Информатика и управление" РАН, Россия,
119333, Москва, Вавилова 40, +7(499)135-24-89, senkoov@mail.ru

Кластерный анализ - распространённый способ выявления закономерностей в наборах данных. В частности, этими данными могут быть временные ряды [1]. Для них существуют свои особенности кластеризации, связанные с тем, как определять расстояния между временными рядами, или на основе каких характеристик временных рядов выявлять кластеры. После выбора анализируемых характеристик, таких, как средние значения, дисперсии, тренды, или выбора метрик, характеризующих расстояния между временными рядами, к ним могут быть применены стандартные методы кластеризации, например, иерархическая кластеризация.

В экономике представляет интерес выделение кластеров из регионов, сходных по каким либо признакам (например, кластеризация стран мира, или субъектов Российской Федерации). Если часть регионов априори относится к какому либо классу, то задача классификации остальных регионов может быть выполнена с помощью методов классификации, например, метода оптимально достоверных разбиений [2]. Если же априорная информация о классах отсутствует, то применим обычный кластерный анализ. Рядом авторов предложены варианты классификации российских регионов на основе имеющейся для них статистической информации [3].

Важно верифицировать, в какой степени кластеризация не является случайным артефактом. Для оценки качества кластеризации предложен ряд методов, таких, как индекс Данна, силуэты и т. п. Нами используются и разрабатываются оценки, основанные на применении в задачах кластеризации методов непараметрической статистики, таких, как бутстрапы и перестановочные тесты. Сферой применения нами этих методов является кластеризация регионов России по эффективности использования факторов производства в создании валового регионального продукта, то есть, по особенностям их производственных функций.

Литература.

1. Aghabozorgi, S.; Shirkhorshidi, A. S.; Wah, Teh Ying. Time-series clustering – A decade review. // Information Systems, v.53, p.16-38, 2015.
2. Сенько О. В. Перестановочный тест в методе оптимальных разбиений // Ж. вычисл. матем. матем. физ., 2003. Т. 43. № 9. С. 1422–1431.
3. Бахитова Р. Х., Ахметшина Г. А., Лакман И. А. Панельное моделирование объема выпуска продукции для регионов России // Управление большими системами, 2014. Вып.50. С. 99–109.