

# Динамическая кластеризация в поддержке принятия решений при управлении социально-экономическим развитием регионов

М.А. Анфёров, e-mail: anfyorov@inbox.ru

«МИРЭА - Российский технологический университет»

***Аннотация.** Рассматривается метод динамического кластерного анализа, реализуемый в условиях неопределенности используемых данных, вызванной их искажением и изменением во времени. На примере поддержки принятия решений при управлении социально-экономическим развитием регионов продемонстрирована технология данного анализа позволяющая выявлять регионы со скрытыми проблемами в их социально-экономической системе.*

***Ключевые слова:** принятие решений, сети Кохонена, динамическая кластеризация, государственное регулирование.*

## Введение

Принятие решений в условиях повышенной неопределенности информации, связанной с ее нечеткостью и стохастичностью, а также недостаточностью данных, требует использование на этапе ее анализа использование современных интеллектуальных информационных технологий [1]. В настоящей работе для поддержки принятия решений в таких условиях рассматривается кластерный анализ, реализуемый с использованием нейронных сетей Кохонена [2]. Применение кластерного анализа отличает от других используемых в принятии решений методов группирования (например, ABC-анализа [3]) учет совокупного влияния признаков, характеризующих анализируемые объекты. Что касается использования двумерных карт Кохонена при отображении результатов кластеризации, то они позволяют расставить приоритетность формируемых кластеров, необходимую при принятии решений.

В отличие от традиционного подхода, предполагающего статичность значений признаков исследуемых объектов, близких по используемой метрике в многомерном пространстве, содержащем кластеры, в предлагаемом методе вышеназванные признаки рассматриваются в различные срезы времени (дни, недели, месяцы и т.д.), что позволяет оценить динамику результатов кластеризации и использовать данную оценку в анализе в процессе принятия решений.

В этом случае для каждого кластеризуемого объекта под номером  $k$  используется вектор из  $n$  признаков  $\mathbf{p}_k^t = (p_{k1}^t, p_{k2}^t, \dots, p_{kn}^t)$ , зависящий от дискретного параметра времени  $t$ .

Динамика во времени формируемых кластеров может вызваться с определенной вероятностью рядом обстоятельств: во-первых, нестабильностью функционирования исследуемых объектов и, во-вторых, вынужденным или умышленным искажением анализируемой информации, например при налоговом администрировании [4]. При этом как в первом, так и во втором случае появляется возможность снизить риски в принятии решений.

Сам по себе динамический подход положительно зарекомендовал себя в принятии решений, например, в маркетинге в виде XYZ-анализа, учитывающего при принятии решений стабильность продаж и колебание спроса на товары с использованием доминантного признака (коэффициента вариации). Однако XYZ-анализ лишь отдаленно напоминает предлагаемый подход, так как при динамической кластеризации учитывается вся совокупность признаков и, что касается динамики, отслеживается изменение самой структуры формируемых кластеров.

Динамический кластерный анализ может использоваться для поддержки принятия решений в различных предметных областях. В данной работе показана технология его применения в рамках государственного регулирования социально-экономического развития регионов Российской Федерации, осуществляемого путем реализации соответствующей административно-правовой, экономической, финансовой и социальной политики.

## 1. Постановка задачи

Состояние социально-экономической системы субъектов Российской Федерации характеризуется основными показателями, значение которых в различные периоды времени приведены в статистическом сборнике [5]. В рамках данной задачи из представленного перечня рассмотрены региональные структуры, характеризующие значениями вышеназванных показателей (признаками кластеризации) в период с 2014 года по 2020 год. Соответственно из перечня исключены данные по г. Москве и Московской области, г. Санкт-Петербургу и в целом по федеральным округам, а также по структурам, входящим в состав областей, например, по Ямало-Ненецкому автономному округу в составе Тюменской области. Кроме этого не учитывались данные по Республике Крым и г. Севастополю из учета того, что рассматриваемый период времени является переходным

в плане стабилизации их социально-экономической системы. В результате в кластерном анализе были использованы данные по 77 региональным социально-экономическим системам субъектов Российской Федерации, представленные значениями следующих признаков (показателей):

1. среднедушевых ежемесячных денежных доходов, руб.;
2. валового регионального продукта в текущих ценах, млрд руб.;
3. основных фондов в экономике, млрд руб.;
4. объема отгруженных товаров собственного обрабатывающего производства, млн руб.;
5. объема продукции сельского хозяйства, млн руб.;
6. объема вводимой в действие общей площади жилых помещений, тыс. м<sup>2</sup>;
7. оборота розничной торговли, млрд руб.;
8. сальдированного финансового результата в экономике, млн руб.;
9. индекса потребительских цен, %;
10. инвестиций в основной капитал, млрд руб.

Для учета различных размеров регионов значения объемных показателей были приведены к относительному виду путем деления на величину среднегодовой численности занятого населения, тыс. человек.

Кластеризация регионов выполнялась на платформе Deductor BaseGroup Labs с выявлением трех кластеров. По результатам анализа карт Кохонена определялся кластер, содержащий регионы с наилучшими экономическими показателями, которому присваивалась метка «1». Кластеру, содержащему регионы с низкими показателями, присваивалась метка «3». Кластеру, выступающему в роли переходного между двумя вышеназванными, присваивалась метка «2». На предварительном этапе по результатам анализа значимости используемых признаков был исключен из дальнейшего рассмотрения показатель среднедушевых ежемесячных денежных доходов.

## 2. Кластерный анализ

Динамический кластерный анализ включает использование трех основных инструментов.

Во-первых, значения признаков объектов кластеризации заменяются их дисперсией на множестве  $w$  значений в рассматриваемых временных срезах (в нашем случае – это годы отчетности), то есть

$$D_{kj} = \frac{1}{w-1} \sum_{i=1}^w (p_{kj}^i - \bar{p}_{kj})^2; \quad \bar{p}_{kj} = \frac{1}{w} \sum_{i=1}^w p_{kj}^i. \quad (1)$$

По результатам кластеризации в пространстве новых признаков (дисперсий) выявляются кластеры объектов с высокой неустойчивостью значений наиболее значимых признаков. В нашем случае в данную категорию попадают регионы 69, 23, 27, 45, 57. Для выявленных объектов на данном этапе нельзя сделать однозначный вывод о причинах и характере изменений в социально-экономической сфере данных регионов для принятия необходимых управленческих решений. Высокая дисперсия характеризующих их параметров может быть вызвана как положительным, так и отрицательным трендом, который определяет соответствующую экономическую тенденцию, а также высокой стохастичностью их значений, говорящей о неустойчивых процессах в экономике региона. Требуется дополнительный регрессионный анализ наличия тренда. На рис. 1а показана эмпирическая модель, характеризующая тренд показателя относительного объема валового регионального продукта  $v = v(t)$  для 69 региона.

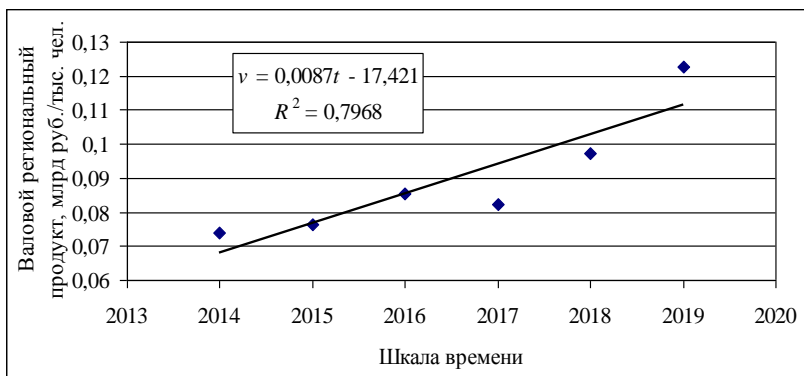
Из полученной зависимости видна значительная устойчивая динамика роста данного экономического показателя, что позволяет судить о положительной тенденции изменений в социально-экономической сфере в данном регионе. Анализ наличия тренда выполнен и по другим используемым при кластеризации признакам, например, относительного объема отгруженных товаров собственного обрабатывающего производства  $s = s(t)$  (см. рис. 1б). Здесь следует заметить, что возможна ситуация, когда по одним экономическим показателям наблюдается устойчивый тренд, а по другим имеет место высокая стохастичность их значений, указывающая на наличие фактора, отрицательно влияющего на социально-экономическую систему. Выявление такого фактора позволяет принимать правильные решения в задействовании необходимых механизмов государственного регулирования данной системы.

Второй инструмент анализа позволяет в отличие от предыдущего выявлять неустойчивость самой кластеризации. Это дает, как отмечалось выше, возможность судить либо о нестабильности функционирования кластеризуемых объектов, либо о сознательном искажении предоставленной информации.

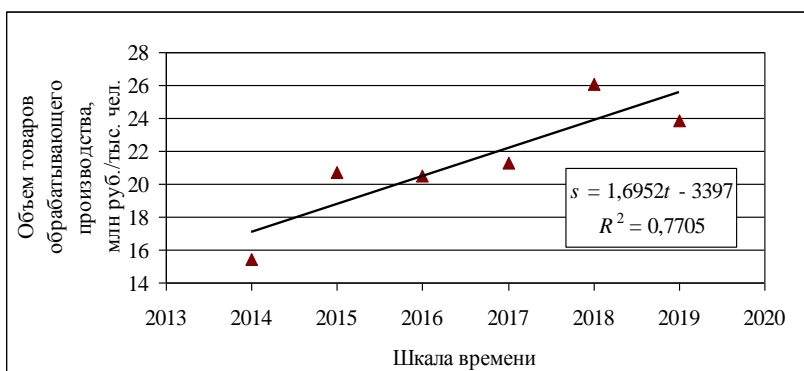
Процедура реализации данного этапа анализа сводится к следующим этапам.

Во-первых, в каждом срезе времени под номером  $r$  выполняется кластеризация анализируемых объектов. При этом для каждого объекта

под номером  $k$  рассчитывается его удаленность (расстояние)  $d_k^r$  до центра включающего его кластера.



а



б

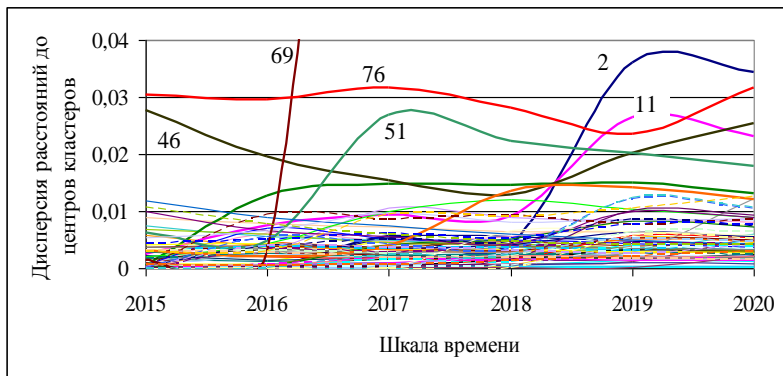
*а – тенденция относительного валового регионального продукт,  
б – тенденция относительного объема отгруженных товаров  
собственного обрабатывающего производства*

Рис. 1. Динамика экономических показателей 69 региона

Во-вторых, для каждого среза времени  $r$  рассчитывается дисперсия значений этих расстояний по всем предыдущим временным срезам, включая рассматриваемый, т.е.

$$D_k^r = \frac{1}{r-1} \sum_{t=1}^r (d_k^t - \bar{d}_k^r)^2; \bar{d}_k^r = \frac{1}{r} \sum_{t=1}^r d_k^t; r = 2, \dots, w. \quad (2)$$

В-третьих, строится диаграмма, отображающая зависимости рассчитанных дисперсий от координаты времени по каждому кластеризуемому объекту (см. рис. 2). По данной диаграмме выявляются объекты с неустойчивостью в плане кластеризации.

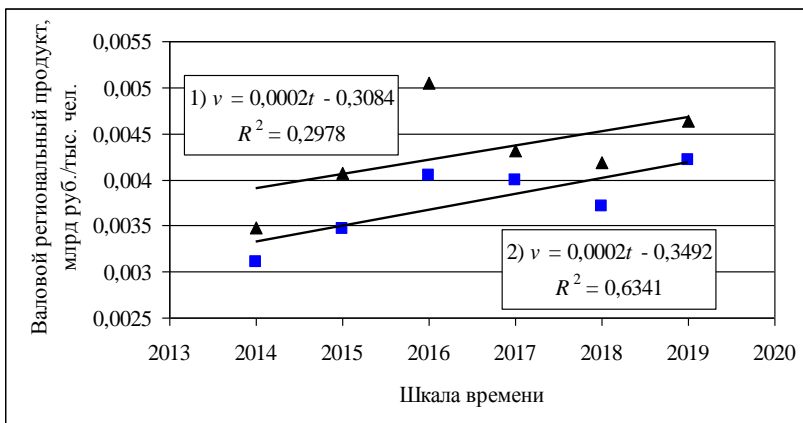


*Числами на графиках обозначены номера регионов с высокой дисперсией*

Рис. 2. Распределение дисперсий расстояний до центров кластеров

Результат такого анализа для рассматриваемого случая показан на рис. 2, из которого видно, что максимальной динамикой положения в кластере характеризуется уже отмеченный ранее регион 69. Причины такого поведения нами уже выявлены. Однако к выявленной ранее группе регионов, характеризующихся неустойчивостью значений признаков, добавился еще ряд регионов: 2, 76, 51, 46, 11. Для выявления причин данной динамики и принятия необходимых управленческих решений также как и в предыдущем случае проведен анализ наличия тренда  $v = v(t)$ . Результат построения эмпирической модели на примере региона 2 показан на рис. 3, из которого видно отсутствие значимой тенденции и наличие высокой стохастичности анализируемого показателя, что позволяет судить о вероятной неустойчивости функционирования социально-экономической системы региона.

Это требует принятия решения в задействовании механизмов государственного регулирования для стабилизации функционирования социально-экономической системы данного региона.



1 – регион 65; 2 – регион 2

Рис. 3. Стохастичность экономических показателей ряда регионов

Третий инструмент анализа основан на выявлении изменения принадлежности тому или иному кластеру, что опять же позволяет судить о стабильности функционирования и возможного искажения предоставляемой информации.

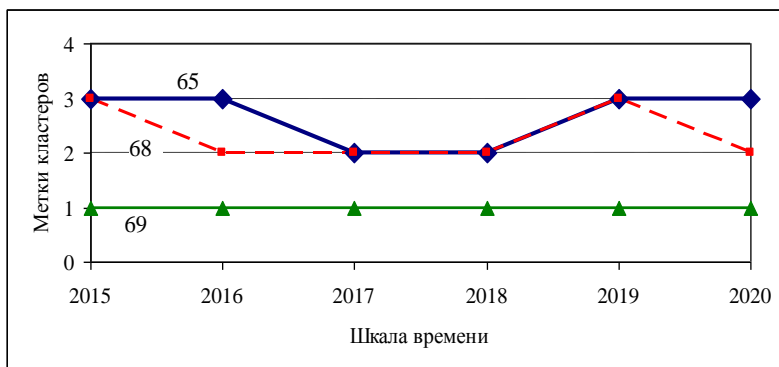


Рис. 4. Изменение принадлежности кластерам

Для этого для каждого кластеризуемого объекта строится диаграмма, отображающая метку кластера, в который попадает кластеризуемый объект в каждом временном срезе (см. рис. 4). Это

позволяет выявить объекты, часто меняющие свою принадлежность кластерам и, тем самым, попадающие в группу риска.

В нашем случае стабильность кластеризации для 69 региона является предсказуемой, т.к. мы уже выявили наличие устойчивой тенденции роста его экономических показателей. Полученный результат только лишний раз подтвердил ранее полученные заключения.

Отмеченная нестабильность 65 региона, дополненная выявленной высокой стохастичностью экономических показателей (см. рис. 3) говорит о необходимости принятия мер в рамках государственного регулирования. Что касается, например, 68 региона, то изменение принадлежности с 3 на 2 кластер говорит о положительной динамике. Это подтверждается и анализом тенденции экономических показателей.

### **Заключение**

Полученные в данной работе результаты подтверждают работоспособность инструментов динамической кластеризации при анализе неопределенной информации с выявлением нестабильных и искаженных данных об анализируемых системах. Это расширяет аналитический инструментарий, используемый при поддержке принятия управленческих решений.

Эффективность динамического кластерного анализа продемонстрирована на примере аналитической поддержки принятия решений при управлении социально-экономическим развитием регионов Российской Федерации за счет использования механизмов государственного регулирования в отношении регионов с успешной и проблемной экономической системой. Причем проблемность функционирования такой системы может быть не вполне очевидной при статичном рассмотрении социально-экономических показателей. Использование же динамического подхода позволяет при анализе данных выявить глубинные воздействующие на экономическую систему негативные факторы.

### **Список литературы**

1. Беляков, С. Л. Метод интеллектуального управления процессом интерактивного изучения пространственных данных для принятия решений / С. Л. Беляков, А. В. Боженюк, С. А. Зубков // Информационные технологии – 2020 – Т. 26. – № 7. – С. 394-402.
2. Kohonen, T. Self-Organizing Maps / T. Kohonen. – 3-rd edition. – Berlin - New York : Springer-Verlag, 2001. – 521 p.
3. Ксенофонтова, О. Л. ABC-XYZ-анализ как средство управления товарным ассортиментом торгового предприятия /



О. Л. Ксенофонтова, Н. А. Новосельская // Современные наукоемкие технологии. Региональное приложение. – 2013.– № 2 (34). – С. 70-66.

4. Анфёров, М. А. SADT моделирование системы налогообложения в Российской Федерации / М. А. Анфёров, О. Б. Рашитова // Экономика и управление: научно-практический журнал. – 2015. – № 2 – С. 94-101.

5. Россия в цифрах: краткий статистический сборник. – М.: Росстат, 2020. – 550 с.