

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

**АНАЛИЗ ДАННЫХ НАУКИ,
ОБРАЗОВАНИЯ И ИННОВАЦИОННОЙ
ДЕЯТЕЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ
МЕТОДОВ АНАЛИЗА ПАТТЕРНОВ**

Препринт WP7/2012/07

Серия WP7

Математические методы
анализа решений в экономике,
бизнесе и политике

Москва
2012

Редакторы серии WP7
«Математические методы анализа решений в экономике,
бизнесе и политике»

Ф.Т. Алескеров, В.В. Подиновский, Б.Г. Миркин

Анализ данных науки, образования и инновационной деятельности с использованием методов анализа паттернов : препринт WP7/2012/07 [Текст] / Ф.Т. Алескеров, Л.М. Гохберг, Л.Г. Егорова, А.Л. Мячин, Г.С. Сагиева ; Нац. исслед. ун-т «Высшая школа экономики». – М. : Изд. дом Высшей школы экономики, 2012.

Анализ данных науки, образования и инновационной деятельности является очень сложной и комплексной проблемой. В данной работе будет описан метод анализа паттернов и результаты его применения к проблеме анализа развития науки, образования и успешности инновационной деятельности в регионах Российской Федерации.

В результате исследования были изучены характеристики регионов России по таким показателям, как уровень социально-экономических условий, потенциал и результативность науки, образования и инновационной деятельности в динамике за 4 года с 2007 по 2010 г. Получена классификация регионов по схожести внутренней структуры указанных показателей, также построены траектории развития регионов с течением времени и найдены группы регионов, придерживающихся одинаковой стратегии развития указанных показателей.

Ключевые слова: паттерн, кластерный анализ, регионы РФ, наука, образование, инновационная деятельность.

Алескеров Ф.Т. – Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ).

Гохберг Л.М. – НИУ ВШЭ.

Егорова Л.Г. – НИУ ВШЭ.

Мячин А.Л. – НИУ ВШЭ.

Сагиева Г.С. – НИУ ВШЭ.

Эта работа является частью более общей задачи по анализу данных науки, образования и инноваций, которая выполняется НИУ ВШЭ по заказу Министерства образования и науки РФ в рамках государственного контракта № 07.514.11.4144 «Разработка экспериментального образца программного обеспечения анализа статистических данных науки, образования и инновационной деятельности с использованием перспективных методов: анализа паттернов данных и онтологического моделирования», шифр «2012-1.4-07-514-0041».

Препринты Национального исследовательского университета
«Высшая школа экономики» размещаются по адресу: <http://www.hse.ru/org/hse/wp>

© Алескеров Ф. Т., 2012

© Гохберг Л. М., 2012

© Егорова Л. Г., 2012

© Мячин А. Л., 2012

© Сагиева Г. С., 2012

© Оформление. Издательский дом
Высшей школы экономики, 2012

Введение

Проблема анализа данных науки, образования и инновационной деятельности является очень сложной и комплексной проблемой и проводить исследования, связанные с этими сферами человеческой деятельности, можно различными способами, с использованием разных методик и разных методических подходов. В данной работе будет описан метод анализа паттернов и результаты его применения к проблеме анализа развития науки, образования и успешности инновационной деятельности в регионах Российской Федерации.

Анализ паттернов – это новая область анализа данных, связанная с поиском взаимосвязей исследуемых объектов, построением их классификации и исследованием развития объектов во времени. Задачей анализа паттернов является разбиение заданной выборки объектов на подмножества, называемые паттернами, так, чтобы каждый паттерн состоял из схожих объектов, а объекты разных паттернов существенно отличались. То есть любой паттерн отражает такие существенные характеристики класса объектов, которые выделяют объекты этого паттерна среди всей остальной совокупности. В этом плане методы анализа паттернов решают такую же задачу, что и классические методы кластеризации. Отличием методов анализа паттернов является работа с совокупностью всех признаков и умение соотносить в один паттерн объекты, имеющие разные количественные характеристики, но одинаковую внутреннюю структуру показателей и взаимосвязей между ними. Более того, с учетом этой особенности анализ паттернов позволяет находить скрытое взаимовлияние показателей, выявлять тренды изменения таких показателей и находить показатели, сигнализирующие о нетипичной динамике объекта.

Ранее оба метода применялись для конкретных прикладных задач, например, для исследования банков по выбранной бизнес-модели [5, 9], стран по характеру зависимости макроэкономических показателей [1, 2], характера конкуренции между политическими партиями по избирательным округам [5] или для поддержки решения менеджера о распределении персонала по филиалам [3, 4, 5].

Есть два подхода к анализу паттернов данных – первый рассматривает паттерны в «статическом виде», анализируя разбиение объектов выборки по классам/типам/кластерам, второй подход обобщает и развивает эту идею, проводя анализ паттернов, выбираемых объектами, с течением времени. В динамическом анализе паттернов проводится анализ смены паттернов объектами с течением времени путем конструирования траекторий, показывающих, какой паттерн был у объекта в определенные моменты времени. Объекты, обладающие сходными траекториями паттернов, формируют динамические группы. Как правило, в плане частоты смены паттернов объектами отдельно рассматриваются группы объектов по степени

приверженности паттернам. В этом аспекте выделяется ряд групп по типу устойчивости: от абсолютно устойчивых до абсолютно неустойчивых. В целом, данный метод позволяет выделить сходства и различия в динамике функционирования и изменения характеристик объектов, выявить общие траектории их развития, а также определить те объекты, которые ведут себя нетипичным образом и требуют более пристального внимания со стороны экспертов.

Анализ паттернов данных выполняет следующие основные задачи:

- разработка типологии/классификации объектов на основе выделенных паттернов (статический анализ паттернов);
- порождение гипотез на основе полученных паттернов данных;
- анализ динамики поведения объектов, выражающейся в потенциальной смене объектом паттерна с течением времени;
- разработка типологии/классификации объектов на основе динамического анализа паттернов.

Оба метода могут быть использованы для анализа данных науки, образования и инновационной деятельности. Объектами анализа являются регионы, целью – построение системы индикаторов, адекватно отражающих положение региона в части развития в нем науки, инновационной деятельности и образования. Стоит отметить, что мы не ставим целью построение рейтинга или вообще «выставления оценок» регионам, нас интересует именно положение дел как оно есть, исследование скрытых взаимосвязей между наукой, образованием и инновационной деятельностью в разрезе регионов, нахождение «близких» по структуре и соотношениям между указанными областями регионов.

1. Описание методов анализа паттернов

Методы анализа паттернов состоят из трех этапов: предварительные этапы связаны с поиском информации, первичной статистической обработкой найденных данных и выбором базовой системы показателей, основной этап статического метода анализа паттернов связан с переходом в признаковое пространство объектов и проведение кластеризации для поиска паттернов данных, затем при необходимости исследования поведения объектов в динамике может быть проведен динамический анализ паттернов.

Общая схема работы методов анализа паттернов представлена на рис. 1.



Рис. 1. Схема применения методов анализа паттернов

Таким образом, применение методов анализа паттернов данных предполагает следующие шаги:

1. Сбор данных, проверка качества найденных данных. Этот шаг предусматривает выбор объектов для анализа и выделение множества показателей, по которым могут быть оценены объекты в выборке, т.е. признакового пространства.
2. Корреляционный анализ исходных показателей. Этот шаг необходим для поиска связей и анализа взаимовлияния между найденными показателями.
3. Выбор базовой системы независимых показателей. На основании предыдущих этапов необходимо сформировать систему показателей, которая будет использоваться в дальнейшем при кластеризации объектов. Предполагается, что количество показателей в базовой системе не должно превышать 10–15 показателей, иначе паттерны становятся слишком чувствительными к отдельным показателям и их количество получается слишком большим для исследования и построения разумной классификации.

4. Кластеризация объектов, представленных кусочно-линейными функциями в признаковом пространстве. На этом этапе необходимо представить исследуемые объекты в виде ломаных линий в пространстве признаков (обычно для этого используется параллельная система координат), определить метрику в признаковом пространстве для вычисления меры сходства (либо различия) между объектами, и провести кластеризацию для получения паттернов.
5. Интерпретация полученных паттернов, как правило, включает в себя создание классификации паттернов данных и описание полученных паттернов.
6. Динамический анализ паттернов состоит в анализе смены паттернов объектами с течением времени путем конструирования траекторий, показывающих какой паттерн был выбран объектом в определенные моменты времени. Объекты, обладающие сходными траекториями смены паттернов, формируют динамические группы (создание классификации для задачи динамического анализа паттернов).

Опишем вкратце суть метода анализа паттернов.

Пусть X – множество объектов, Y – множество номеров (имён, меток) кластеров.

Для формального описания понятия близости объектов задается функция расстояния $\rho(x, x')$ между объектами $x, x' \in X$. Для рассматриваемой конечной выборки объектов $X^m = \{x_1, \dots, x_m\} \subset X$ требуется разбить ее на непересекающиеся подмножества, называемые паттернами, так, чтобы каждый паттерн состоял из объектов, близких по метрике ρ , а объекты разных паттернов существенно отличались. При этом каждому объекту $x_i \in X^m$ приписывается номер паттерна u_i .

Будем понимать под паттерном набор значений системы показателей, описывающих какую-либо группу объектов, а также саму группу объектов, имеющих такие же или почти такие же значения показателей данной системы.

Для визуального отображения паттернов в n -мерном пространстве используется система параллельных координат [8], состоящая из n параллельных линий, как правило, вертикальных и равномерно распределенных. Каждая из этих параллельных линий отражает один из выбранных для кластеризации признаков. Каждый объект из X (точка в n -мерном пространстве признаков) представляется в виде ломаной с вершинами на параллельных осях (названиях показателей); положение вершин на i -й оси соответствует i -й координате точки, т.е. значению i -го показателя. Таким образом, для каждого объекта мы получаем кусочно-линейную функцию, отражающую структуру взаимосвязей между его показателями.

Итак, каждому объекту $x \in X$ поставлен в соответствие n -мерный вектор $Z_x = \{z_1^x, \dots, z_n^x\}$, характеризующий объект x по n признакам, т.е. его признаковое описание.

Построим по каждому вектору Z_x кривую, которая проходит через точки z_1^x, \dots, z_n^x , т.е. функцию $f^x: R \rightarrow R$, такую что $f^x(i) = z_i^x$ для всех $i = 1, \dots, n$. Для простоты удобно брать в качестве функции f^x кусочно-линейную функцию, в таком случае

$$\begin{aligned} f^x(t) &= \{k_i^x t + a_i^x, \text{ если } i \leq t \leq i + 1\}, \\ k_i^x i + a_i^x &= z_i^x, \\ k_i^x(i + 1) + a_i^x &= z_{i+1}^x, i = 1, \dots, n - 1. \end{aligned} \quad (1)$$

Теперь каждому интервалу $[i, i + 1]$ поставим в соответствие пару $V_i^x = (k_i^x, a_i^x)$, и каждой функции f^x – вектор $V^x = (V_1^x, \dots, V_n^x)$. Введем теперь функцию $E(V^{x_1}, V^{x_2})$, зависящую от расстояния $r(V^{x_1}, V^{x_2})$ между объектами x_1 и x_2 в векторном пространстве $\{V^x\}_{x \in X}$ следующим образом

$$E(V^{x_1}, V^{x_2}) = \exp(-\alpha \cdot r(V^{x_1}, V^{x_2})). \quad (2)$$

Далее одним из методов кластерного анализа проводится разбиение множества объектов $\{X_i\}$ на множестве векторов $\{V^x\}_{x \in X}$ так, что

$$\sum_i (\sum_{V \in X_i} E(V, X_i) / \sum_{V \in X \setminus X_i} E(V, X_i)) \rightarrow \max_{\{X_i\}}, \quad (3)$$

где $E(V, X_i) = \sum_{V^y \in X_i} E(V, V^y)$.

Данная целевая функция (3) характеризует качество классификации объектов: большие ее значения соответствуют случаю, когда кривые внутри кластера похожи, а кривые в разных кластерах находятся «далеко» друг от друга в терминах функции $E(V^x, V^y)$. Для нахождения разбиения $\{X_i\}$, максимизирующего этот функционал, можно использовать различные методы кластерного анализа [13].

Для проведения динамического анализа паттернов должна быть доступна информация о количественных характеристиках объектов, измеренная в последовательные моменты времени $t = 1, 2, \dots, k$. Тогда каждому объекту выборки $x_i \in X$ ставится в соответствие последовательность номеров кластеров (траектория) $\{y_i^1, y_i^2, \dots, y_i^t, \dots, y_i^k\}$, к которым данный объект x_i принадлежал в момент времени t . На основании полученных «траекторий развития» можно выделять динамические группы объектов, демонстрирующих одинаковое поведение и сходные характеристики в динамике.

2. Выбор системы показателей

2.1. Постановка задачи

Для использования методики анализа паттернов необходимо в первую очередь сформировать базовую систему индикаторов, которая отражала бы характеристики, описывающий положение региона в части развития науки, образования и ведения инновационной деятельности. В качестве показателей можно использовать как абсолютные показатели различ-

ных статистических сборников, так и индикаторы или агрегированные показатели, выбранные из сборников или составленные для конкретной исследовательской задачи. На основе выбранной системы базовых индикаторов затем строятся кусочно-линейные функции и проводится разбиение всей совокупности объектов на схожие объекты. Каждое такое подмножество исходных объектов, имеющее схожие значения или соотношения в индикаторах, и называется паттерном данных.

Учитывая, что в рамках данной задачи в качестве объекта исследования выступают регионы Российской Федерации, то для выбора показателей в базовую систему индикаторов прежде всего отбираются такие показатели, которые имеют данные по всем регионам. Кроме того, требуется также, чтобы все показатели имели данные хотя бы за последние несколько лет, иначе строить временные ряды и проводить динамический анализ паттернов будет невозможно.

Кроме того, важный этап – проведение первичной статистической обработки данных для изучения взаимосвязей между показателями. Этот этап позволяет выделить связи и зависимости между выбранными показателями и при необходимости сократить число переменных. Возможность редукции числа переменных очень важна с учетом возрастающей сложности и временных затрат на работу алгоритма при увеличении размерности входных показателей. Также возможно удаление выбросов, которые с большой вероятностью образуют отдельный кластер и не внесут ясности в структуру взаимосвязей между объектами в силу своей уникальности.

При необходимости значения исходных показателей могут быть преобразованы, например, приведены к относительным показателям и выражены в процентах, долях, либо могут быть введены дополнительные показатели, представляющие собой агрегированные величины существующих показателей. Такие агрегаты могут использоваться в дальнейшем как полноправные показатели. Также возможно проведение нормализации данных делением на размах или стандартное отклонение.

В результате должна получиться система показателей, характеризующая регион с точки зрения развития в нем науки, образования и успешности инновационной деятельности.

2.2. Исходные данные

Используемые показатели для проведения анализа паттернов объектов данных должны удовлетворять следующим критериям:

- Показатели должны быть измеримы и чувствительны к изменению ситуации в регионе, если это изменение связано с наукой, образованием или инновационной деятельностью или затрагивает их.

- Каждый показатель должен быть ясно определен – должно быть однозначно понятно, что этот показатель измеряет и как он вычисляется.
- Данные, используемые для вычисления показателя, должны быть достоверными.
- Данные должны быть измерены с минимально возможной погрешностью. Любая погрешность увеличивает вероятность неадекватного результата разбиения объектов по паттернам. Под адекватностью результатов подразумевается получение кластеров, соответствующих реальной структуре данных.
- Значение показателя должно собираться с регулярностью: частота его измерения должна быть сопоставима с частотой изменения объектов и не должна ухудшать точность измерения.
- Действия с выбросами в данных зависят от специфики решаемой задачи. Как правило, такие объекты образуют уникальный кластер, что важно в ряде задач для определения аномального поведения объектов, особенно при принятии управленческих решений в отношении подобных объектов. С другой стороны, если задачей является прояснение основной структуры данных, то целесообразнее исключить выбросы при проведении корреляционного анализа.
- Наличие информации о всей системе показателей, выбранной для анализа объектов. Не допускается наличие пропусков и пробелов значений показателей. Если они присутствуют в ограниченном объеме, то необходимо экстраполировать предшествующие значения, чтобы заполнить недостающую информацию, в противном случае – показатель будет исключен.

Для сбора информации и выбора показателей для построения базовой системы индикаторов первоначально предполагалось использовать данные статистических сборников НИУ ВШЭ: «Индикаторы науки: 2010» [10], «Индикаторы образования: 2010» [11] и «Индикаторы инновационной деятельности» [12]. Однако в выбранных сборниках отсутствовали региональные данные для таких отраслей, как наука и образование. Данные по всем регионам были доступны только для показателей инновационной деятельности. Поэтому было решено использовать систему показателей, разработанную для построения рейтинга инновационного развития регионов коллективом ИСИЭЗ НИУ ВШЭ [14].

Эта система показателей является многоуровневой и включает в себя 4 блока, характеризующие:

- 1) социально-экономические условия инновационной деятельности (основные макроэкономические показатели, характеристики образовательного потенциала населения и развития информационного общества);

- 2) научно-технический потенциал инновационной активности (включая кадровый и финансовый потенциал, публикационную и патентную активность, торговлю технологиями);
- 3) инновационную деятельность (активность в сфере технологических и нетехнологических инноваций, развитие малого инновационного бизнеса, затраты на технологические инновации и результативность инновационной деятельности);
- 4) качество проводимой в регионах инновационной политики (с учетом качества нормативно-правовой базы и организационного обеспечения инновационной политики, удельных объемов затрат консолидированного бюджета регионов) [14].

Общая схема иерархии показателей в российском региональном инновационном индексе приведена на Рис. 2.

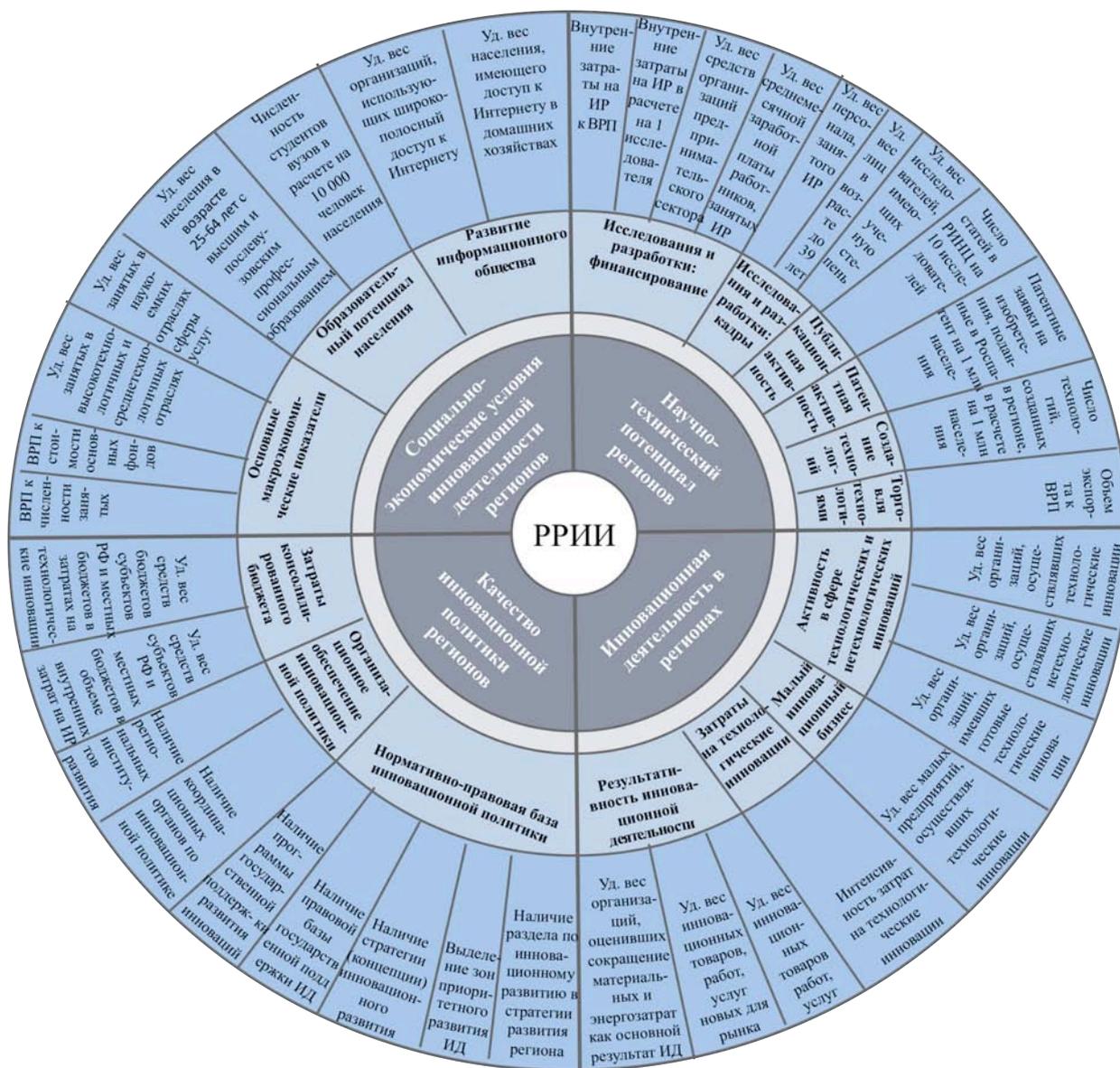


Рис. 2. Схема российского регионального инновационного индекса [14]

Процедура формирования рейтинга в соответствии с разработанной моделью представляет собой следующую последовательность действий (описание приведено в соответствии с [14]).

На первом этапе анализируется текущий состав показателей, включенных в проблемно-содержательные блоки. Для того, чтобы обеспечить устойчивость модели рейтинга и избежать ее перегрузки избыточным числом показателей проводится анализ корреляционных связей между показателями одного тематического блока. Если коэффициент корреляции между двумя показателями свидетельствует об их тесной взаимосвязи, принимается решение об исключении одного из показателей.

На втором этапе рассчитываются значения четырех субиндексов по каждому проблемно-содержательному блоку, включенному в состав интегрального индекса, в том числе:

- индекс «Социально-экономические условия инновационной деятельности» (ИСЭУ);
- индекс «Научно-технический потенциал» (ИНТП);
- индекс «Инновационная деятельность» (ИИД);
- индекс «Качество инновационной политики» (ИКИП).

Однородность и сопоставимость показателей достигается за счет перехода от абсолютных величин к взвешенным (нормированным) значениям, а также при необходимости сглаживается влияние экстремальных (выделяющихся) значений показателей с помощью степенных преобразований. Затем определяются нормированные значения показателей по каждому региону как отношение разницы между значением показателя в регионе и минимального значения показателя по всем регионам к разнице между максимальным и минимальным значениями данного показателя по всем регионам. Диапазон значений нормированных показателей ограничивается интервалом от 0 (у региона с минимальным значением показателя) до 1 (у региона с максимальным значением данного показателя).

Субиндекс рассчитывается как среднее арифметическое нормированных значений показателей в данном проблемно-содержательном блоке.

На третьем этапе производится расчет значения интегрального индекса на основе линейной комбинации входящих в его состав индексов (субиндексов). Так как проблемно-содержательные блоки могут включать различное число показателей, для выравнивания их вклада в итоговую оценку вводятся весовые коэффициенты. При этом показатели, входящие в один тематический блок, имеют равную значимость. Значения весовых коэффициентов субиндексов принимаются равными доле числа показателей, используемых в расчете каждо-

го субиндекса, в общем числе отобранных показателей. Таким образом, сумма весовых коэффициентов субиндексов будет равна единице.

Применительно к сформированной системе показателей формула расчета интегрального индекса имеет вид:

$$PРИИ_r = \frac{(8/35 * ИСЭУ_r + 10/35 * ИНТП_r + 8/35 * ИИД_r + 9/35 * ИКИП_r)}{4} \quad (4)$$

На четвертом этапе выполняется ранжирование регионов в порядке убывания величины интегрального индекса и субиндексов, присвоение региону соответствующих рангов (мест региона в рейтинге), кластеризация регионов по значению индекса.

На основе методологии исследования [14] и используемой в нем системы показателей для решения поставленной задачи нами была построена другая система базовых показателей, описанная ниже.

При выборе показателей учитывалась доступность информации, представимость для всех регионов, а также наличие ежегодных данных по всем показателям. Все используемые в данном исследовании показатели взяты из статистических сборников Федеральной службы государственной статистики [15, 16, 17] и НИУ ВШЭ [12].

2.3. Базовая система показателей

Первоначально для создания базовой системы индикаторов были выделены следующие блоки показателей, выбранные на основе российского регионального инновационного индекса:

- 1) социально-экономические условия;
- 2) образовательный потенциал;
- 3) потенциал научно-технической деятельности;
- 4) результативность исследований и разработок;
- 5) потенциал инновационной деятельности;
- 6) результативность инновационной деятельности.

Каждый из этих блоков также является агрегированным и содержит несколько базовых индикаторов (табл. 1).

Таблица 1. Система показателей для анализа паттернов

Показатель		Доступные годы	Источник
Блок 1. Социально-экономические условия			
1.1	ВРП к величине экономически активного населения в экономике региона, рублей	2002–2009	Статистический сборник Федеральной службы государственной статистики «Российский статистический ежегодник»
1.2	ВРП к стоимости основных фондов	2002–2009	Статистический сборник Федеральной службы государственной статистики «Российский статистический ежегодник»
Блок 2. Образовательный потенциал			
2.1	Удельный вес экономически активного населения, имеющего высшее и послевузовское профессиональное образование, в общей численности населения, %	2006–2011	Статистический сборник Федеральной службы государственной статистики «Обследование населения по проблемам занятости»
2.2	Удельный вес экономически активного населения, имеющего среднее профессиональное образование, в общей численности населения, %	2006–2011	Статистический сборник Федеральной службы государственной статистики «Обследование населения по проблемам занятости»
2.3	Численность студентов образовательных учреждений высшего профессионального образования в расчете на 10 тыс. человек населения	1997–2011	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
2.4	Численность студентов государственных и муниципальных образовательных учреждений среднего профессионального образования на 10 тыс. человек населения	1995–2011	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
2.5	Выпуск квалифицированных рабочих и служащих из образовательных учреждений начального профессионального образования на 10 тыс. человек занятого населения, человек	2005–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»

Блок 3. Потенциал научно-технической деятельности

3.1	Внутренние затраты на исследования и разработки в процентах к ВРП, %	2002–2009	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
3.2	Внутренние затраты на исследования и разработки в расчете на 1 исследователя	2002–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
3.3	Удельный вес персонала, занятого исследованиями и разработками, в общей численности экономически активного населения в экономике региона, %	2000–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
3.4	Удельный вес лиц, имеющих ученую степень в численности исследователей, %	1998–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»

Блок 4. Результативность исследований и разработок

4.1	Число патентных заявок на изобретения, поданных в Роспатент национальными заявителями в расчете на 10 тыс. человек экономически активного населения региона	2002–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
4.2	Число выданных охранных документов в расчете на 10 тыс. человек экономически активного населения региона	2002–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»
4.3	Число созданных передовых производственных технологий в расчете на 10 тыс. человек экономически активного населения региона	2000–2010	Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели»

4.4	Объем поступлений от экспорта инновационных товаров, работ и услуг для добывающего и обрабатывающего производства, производства и распределения электроэнергии, газа и воды, в процентах к общему объему поступлений от экспорта технологий, %	2007–2010	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»
Блок 5. Потенциал инновационной деятельности			
5.1	Удельный вес организаций, приобретавших новые технологии, в общем числе организаций, осуществлявших технологические инновации (добывающие, обрабатывающие производства, производство и распределение электроэнергии, газа и воды), %	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»
5.2	Удельный организаций, передававших новые технологии, в общем числе организаций, осуществлявших технологические инновации (добывающие, обрабатывающие производства, производство и распределение электроэнергии, газа и воды), %	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»
5.3	Затраты на технологические инновации, в процентах от общего объема отгруженных товаров, выполненных работ, услуг (добывающие, обрабатывающие производства, производство и распределение электроэнергии, газа и воды)	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»
5.4	Затраты на технологические инновации, в процентах от общего объема отгруженных товаров, выполненных работ, услуг (связь, деятельность, связанная с использованием вычислительной техники и информационных технологий)	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»

Блок 6. Результативность инновационной деятельности			
6.1	Объём инновационных товаров, работ, услуг, в процентах от общего объема отгруженных товаров, выполненных работ, услуг (добывающие, обрабатывающие производства, производство и распределение электроэнергии, газа и воды)	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»
6.2	Объём инновационных товаров, работ, услуг, в процентах от общего объема отгруженных товаров, работ, услуг (связь, деятельность, связанная с использованием вычислительной техники и информационных технологий)	2007–2011	Статистический сборник НИУ ВШЭ «Индикаторы инновационной деятельности»

С учетом того, что для динамического анализа паттернов необходимо иметь данные по всем показателям за выбранный промежуток времени, решено использовать данные лишь по 4 годам: 2007–2010 гг. Далее все показатели z_i^x (индекс x относится к объектам-регионам, индекс i пробегает все номера из табл. 1) были преобразованы аналогично [14] и получены нормированные показатели по всем блокам согласно формуле

$$\tilde{z}_i^x = \frac{z_i^x - \min(z_i^x)}{\max(z_i^x) - \min(z_i^x)},$$

на основе этих показателей были рассчитаны значения Блоков 1–6 как среднее арифметическое всех нормированных показателей, входящих в этот блок.

2.4. Корреляционный анализ

При выборе базовой системы показателей необходимо помнить о том, что в признаковых описаниях объекта должны быть отражены все существенные стороны объекта. Как правило, чтобы не упустить важных для анализа данных показателей, на первом этапе составляется расширенный список показателей и затем из него исключаются менее значимые признаки. В случае присутствия «лишних», незначимых признаков увеличивается время работы метода и расход ресурсов для обработки данных. Кроме того, в этом случае также возникает проблема адекватности полученных результатов.

Корреляционный анализ позволяет выбирать наименее взаимозависимые коэффициенты, что повышает адекватность представления объектов паттернами данных и увеличивает

точность разбиения по классам. Наличие корреляции между показателями говорит о статистической взаимосвязи между ними, т.е. изменения значений одного показателя сопровождаются изменением значений другого показателя. Таким образом, наличие сильно коррелированных индикаторов не приносит никакой существенной пользы с точки зрения описания объектов, повышает размерность входных данных для анализа паттернов и приводит к неустойчивости паттернов, полученных на основе такой системы показателей.

Для обнаружения статистической взаимосвязи показателей существует несколько видов несколько коэффициентов корреляции: ковариация и линейный коэффициент корреляции, коэффициент ранговой корреляции Кендалла, коэффициент ранговой корреляции Спирмена, коэффициент корреляции знаков Фехнера и проч. Имеющиеся данные являются количественными, поэтому мы использовали линейный коэффициент корреляции при анализе. Линейный коэффициент корреляции представляется собой нормированную на стандартные отклонения ковариацию, поэтому он лишен ее главного недостатка – наличия размерности.

Как оказалось, корреляция между полученными агрегированными показателями является значительной лишь для одной пары показателей, причем значимая корреляция присутствует для всего временного интервала. Блок 3 «Потенциал научно-технической деятельности» и Блок 4 «Результативность научной деятельности» имеют коэффициент корреляции на уровне 0,6–0,7, поэтому было решено исключить из рассмотрения Блок 3, ограничившись для нашего исследования лишь данными по результативности научной деятельности. Остальные блоки можно считать некоррелированными (табл. 2). Корреляционные матрицы для 2008–2010 гг. приведены в Приложении 1.

Таблица 2. Корреляция между блоками показателей для 2007 г.

2007	Блок 1	Блок 2	Блок 3	Блок 4	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1					
Блок 2. Образовательный потенциал	-0,2	1				
Блок 3. Потенциал научно-технической деятельности	0,16	-0,1	1			
Блок 4. Результативность исследований и разработок	0,11	0,11	0,7	1		
Блок 5. Потенциал инновационной деятельности	-0,2	0,14	0,22	0,33	1	
Блок 6. Результативность инновационной деятельности	-0,1	0,05	0,05	0,13	0,21	1

Подводя итоги данного раздела, можно сказать, что на основании выбранной базовой системы показателей получена комплексная оценка развития и результативности науки, образования и инновационной деятельности в каждом регионе Российской Федерации, причем в динамике за 4 года. Каждому из исследуемых объектов-регионов поставлен в соответствие вектор признаков описаний, составленный из 5 агрегированных показателей (Блок 1, Блок 2, Блок 4, Блок 5 и Блок 6). Каждый из этих показателей представляет собой число от 0 до 1, все показатели являются некоррелированными.

3. Анализ паттернов

3.1. Полученные данные

Согласно методу анализа паттернов, имеющиеся признаковые описания объектов нужно представить в системе параллельных координат, заменив точки в пятимерном пространстве признаков на ломаные, построенные следующим образом: на оси абсцисс откладываются номера показателей, которые характеризуют структуру объекта, ось ординат представляет собой ось значений этих показателей. Для каждого объекта мы имеем набор точек, соответствующий его структуре. Кусочно-линейная функция паттерна строится путем соединения этих точек прямыми линиями. Такая процедура производится для каждого объекта анализа.

Примеры полученных кусочно-линейных функции приведены на рис. 3 и рис. 4. Стоит отметить, что здесь и на всех последующих рисунках справа могут быть отмечены не все, а лишь часть регионов, представленных на рисунке.

У читателя может возникнуть вопрос относительно максимального уровня показателя по первому блоку, который характеризует социально-экономические условия региона, для Ненецкого автономного округа в 2007 г. (рис. 3). Наличие единицы является следствием того, что в 2007 г. максимум по обоим показателям первого блока был именно у этого региона, что связано с небольшой численностью населения и достаточно высоким уровнем таких показателей, как ВРП и стоимость основных фондов. Ни в одном из последующих лет Ненецкий автономный округ уже не имеет 1 в этом блоке, как, впрочем, и ни один регион не имеет 1 ни в одном из пяти выбранных блоков за все остальные три года.

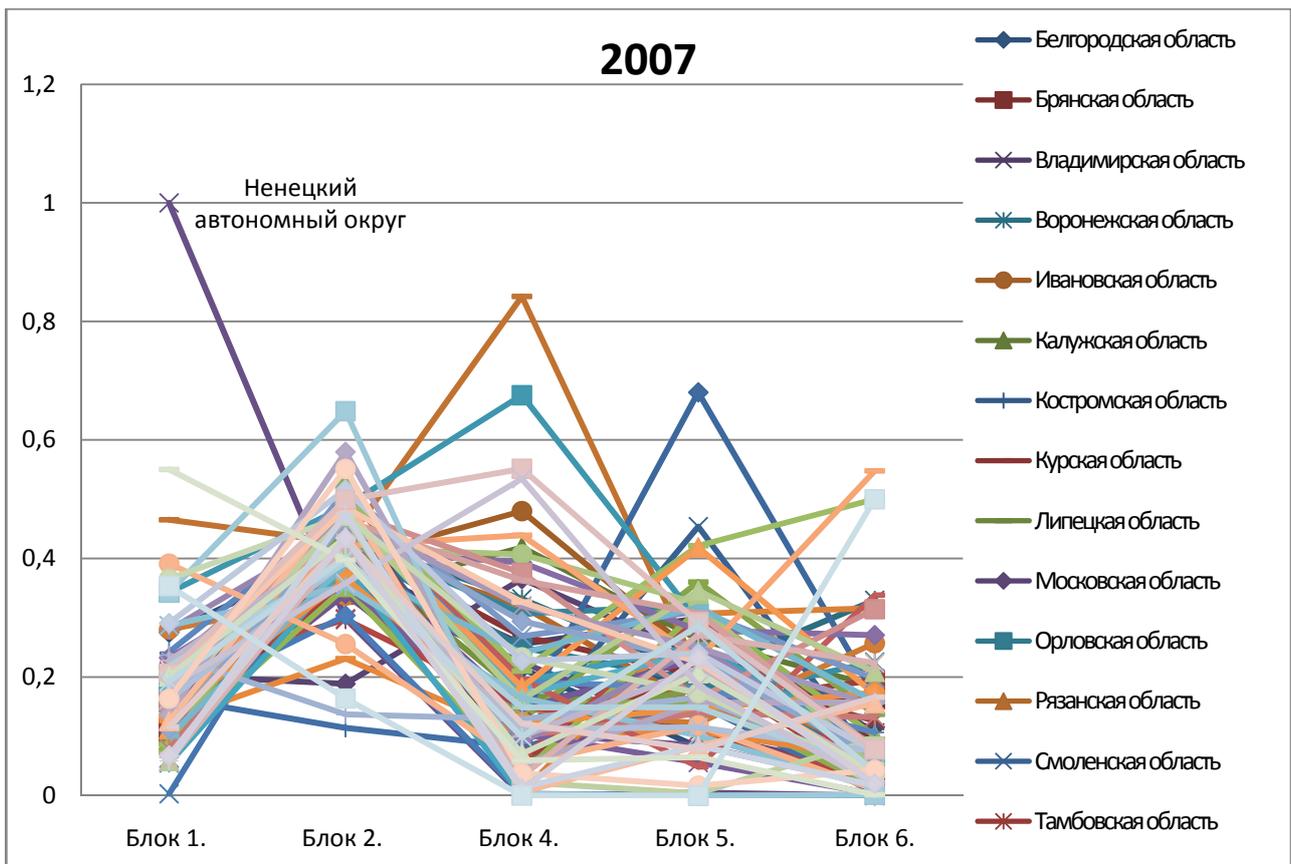


Рис. 3. Кусочно-линейные функции, описывающие объекты, для 2007 г.

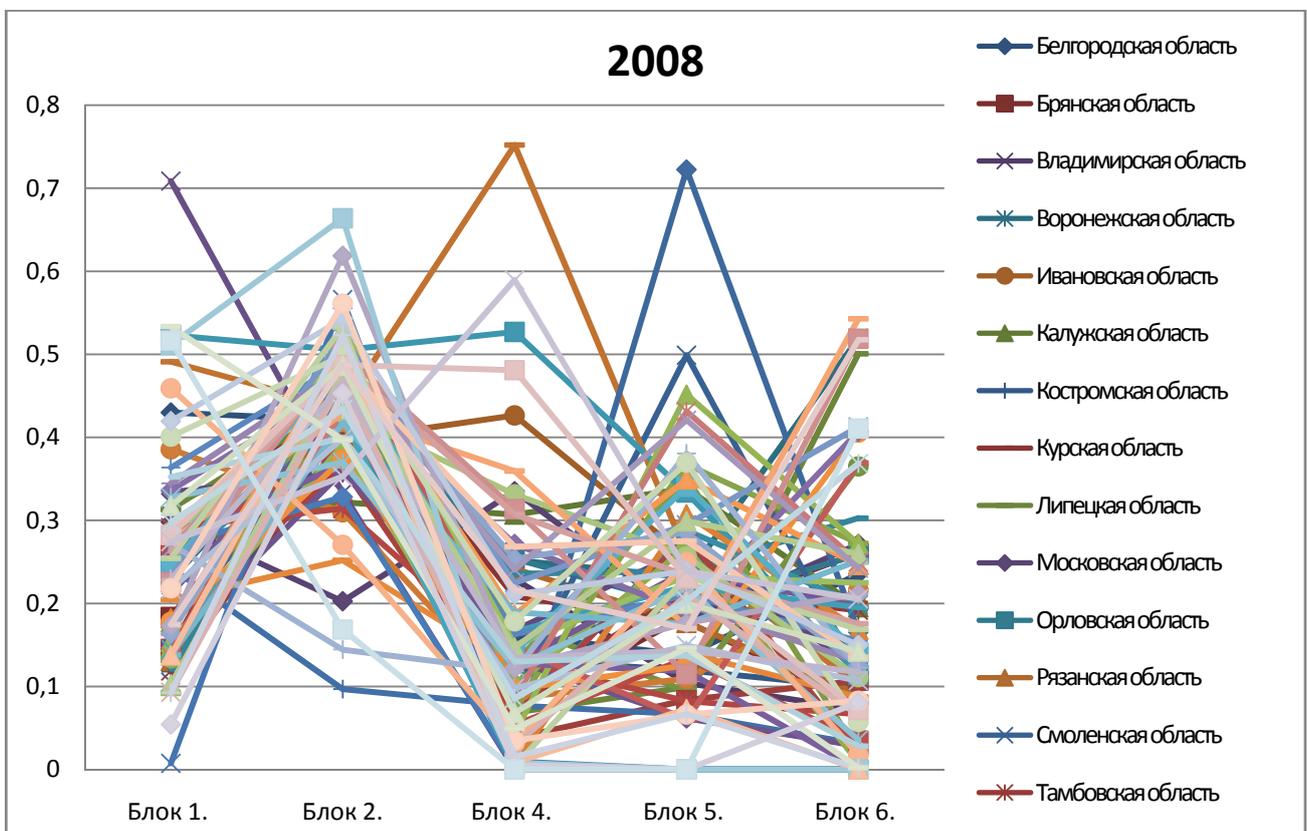


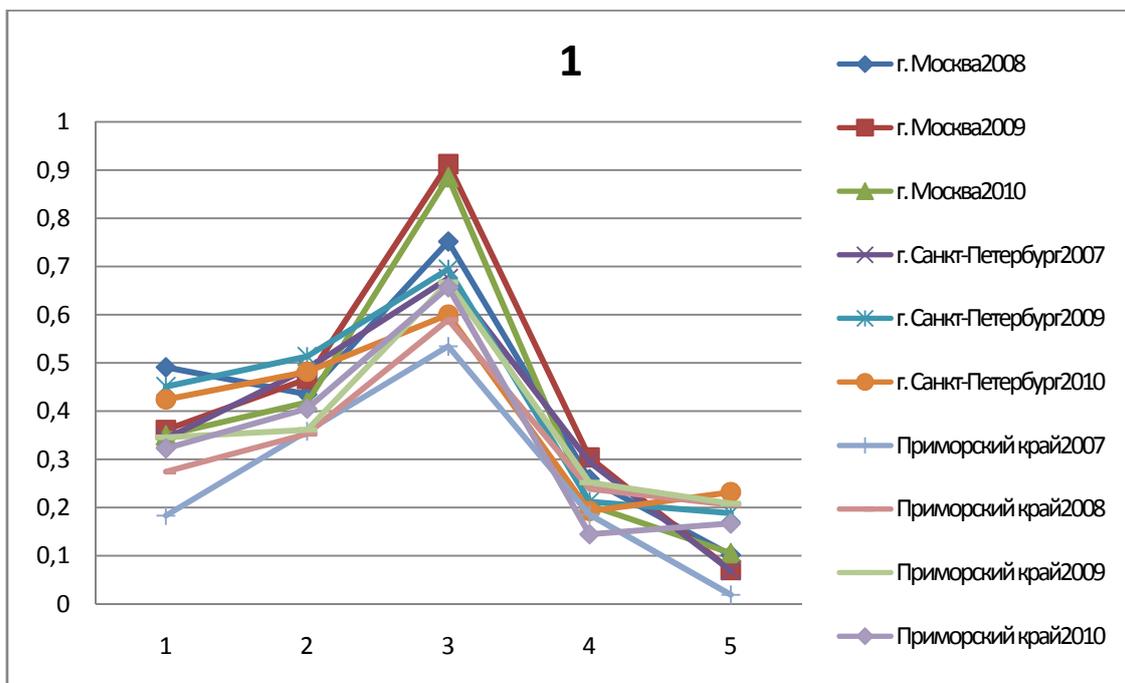
Рис. 4. Кусочно-линейные функции, описывающие объекты, для 2008 г.

Кажущееся хаотичным нагромождением изломанных линий при ближайшем рассмотрении проявляет некоторые закономерности и видно, что многие регионы имеют если не равные, то очень схожие соотношения параметров. Логично считать, что регионы, имеющие приблизительно одинаковую форму такой кусочно-линейной функции, имеют и схожую структуру показателей в признаковых описаниях, а значит, и схожи в выбранной модели развития науки, образования и инновационной деятельности.

3.2. Результаты анализа

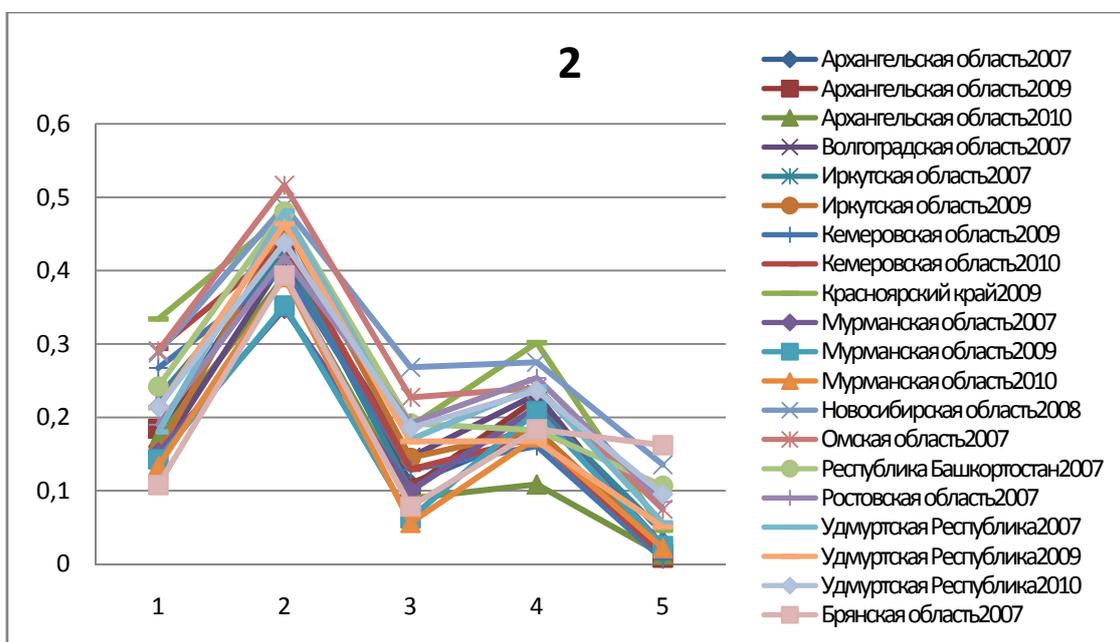
Учитывая, что целью работы является динамический анализ поведения регионов, то полученные кусочно-линейные кривые для каждого года были объединены в общую выборку. Итого получилось $83 \times 4 = 332$ объекта для кластеризации.

Мы использовали два метода кластеризации: *k*-средних и иерархический метод кластеризации [13], в каждом из которых использовалась евклидова метрика, и проводили многошаговую процедуру кластерного анализа, комбинируя два вышеупомянутых метода. В результате было получено 24 паттерна данных, включающих более двух объектов внутри паттерна, и 24 уникальных паттерна (они представлены на одном, последнем рисунке). Сами данные по всем регионам представлены в Приложении 2.

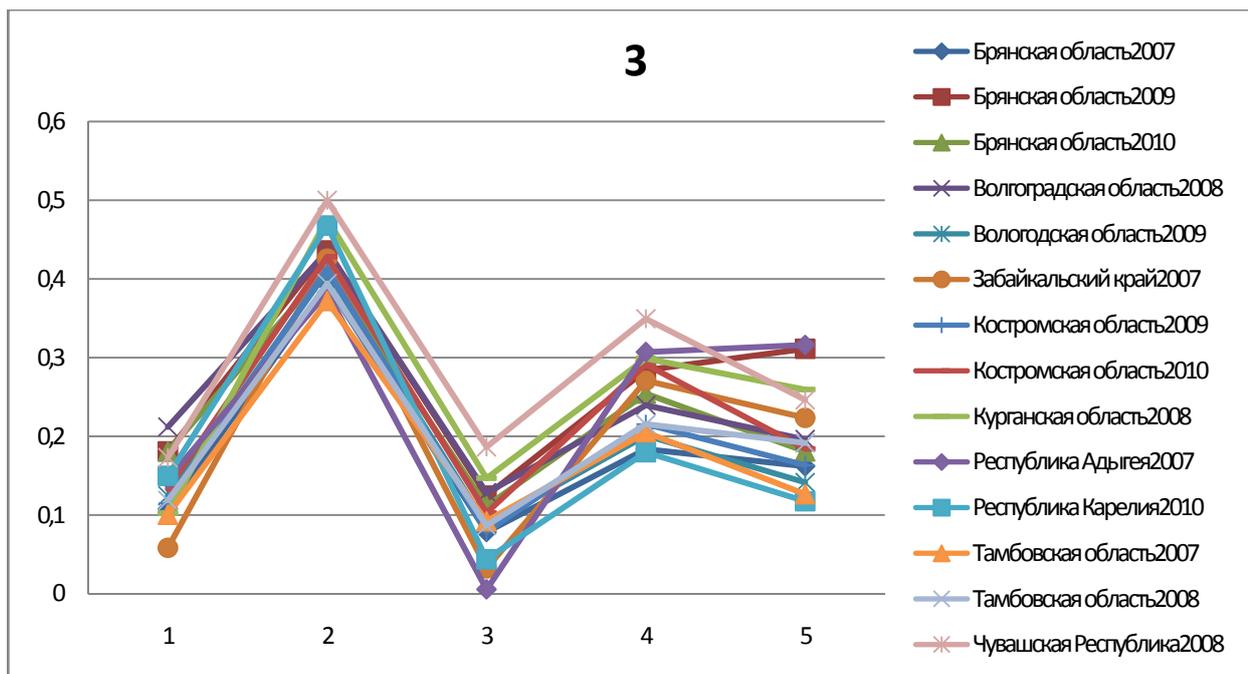


В 1-й паттерн попали 11 ломаных, представляющих город Москву и Приморский край за все 4 года с 2007 по 2010 г., а также город Санкт-Петербург в 2007, 2009, 2010 гг. Такой

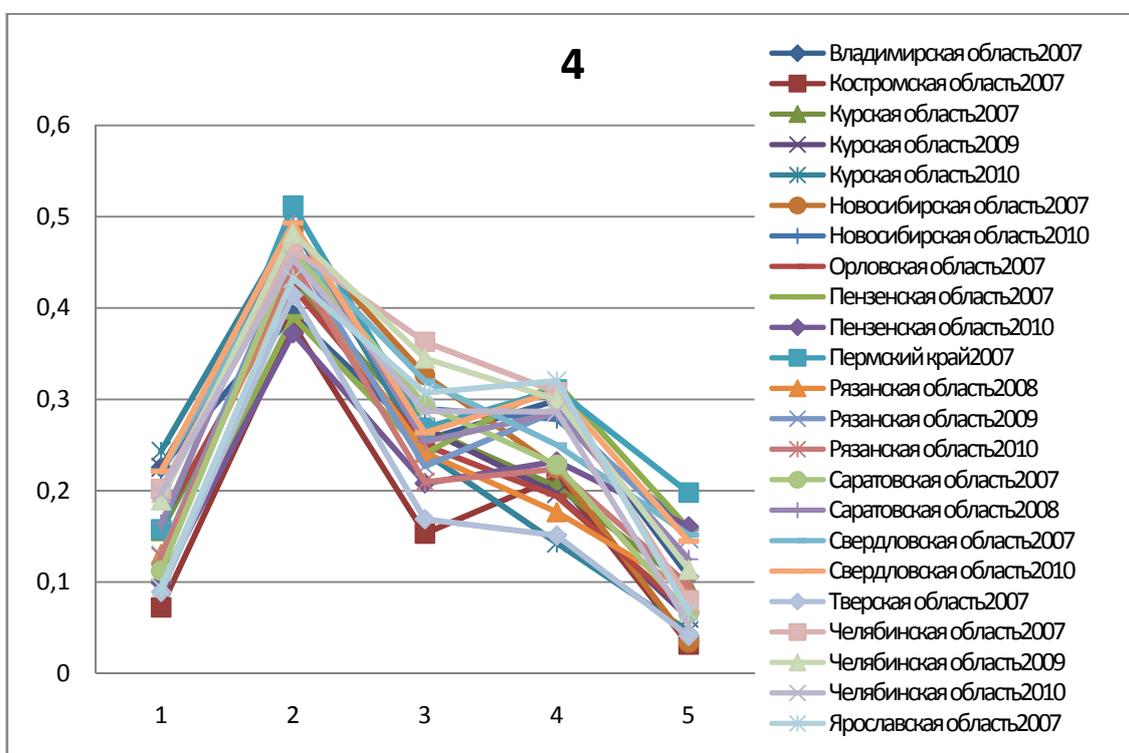
паттерн характеризуется умеренными (0,3–0,5) значениями первых двух показателей (Блок 1 «Социально-экономические условия» и Блок 2 «Образовательный потенциал»), чрезвычайно высоким (0,5–0,9) показателем Блока 4 «Результативность исследований и разработок», а также очень низкими (менее 0,2) показателями Блоков 5 и 6, отвечающими за потенциал и результативность инновационной деятельности. Напомним, что все показатели в базовой системе представляют собой нормированные показатели – они либо приведены к численности экономически активного населения, либо представляют собой процентное отношение к общему показателю. Мы намеренно отказались от абсолютных показателей, имея целью исследовать внутреннюю структуру взаимной зависимости науки, образования и инновационной деятельности.



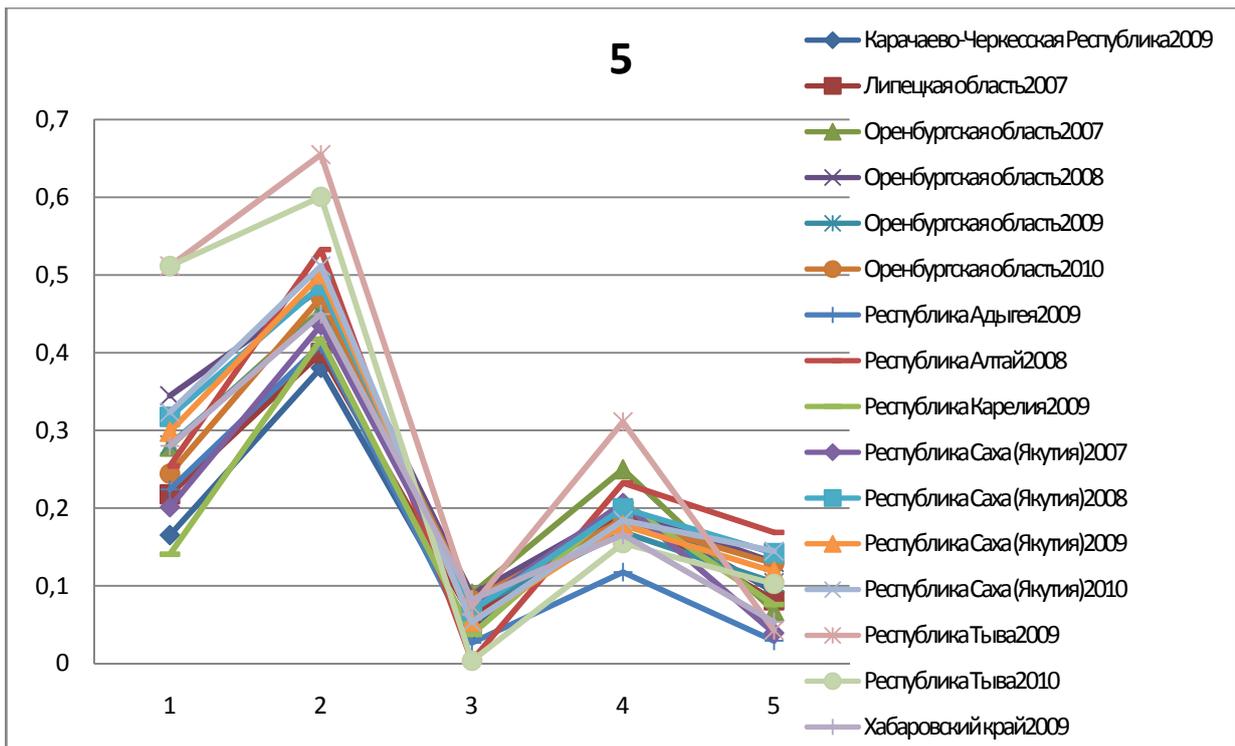
Во 2-м паттерне находятся 19 объектов. Основные характеристики этого паттерна: низкие (0,1–0,3) показатели блоков 1, 4, 5, 6 (социально-экономические условия, потенциал и результативность науки и инновационной деятельности), и средние (0,4–0,5) показатели образовательного потенциала.



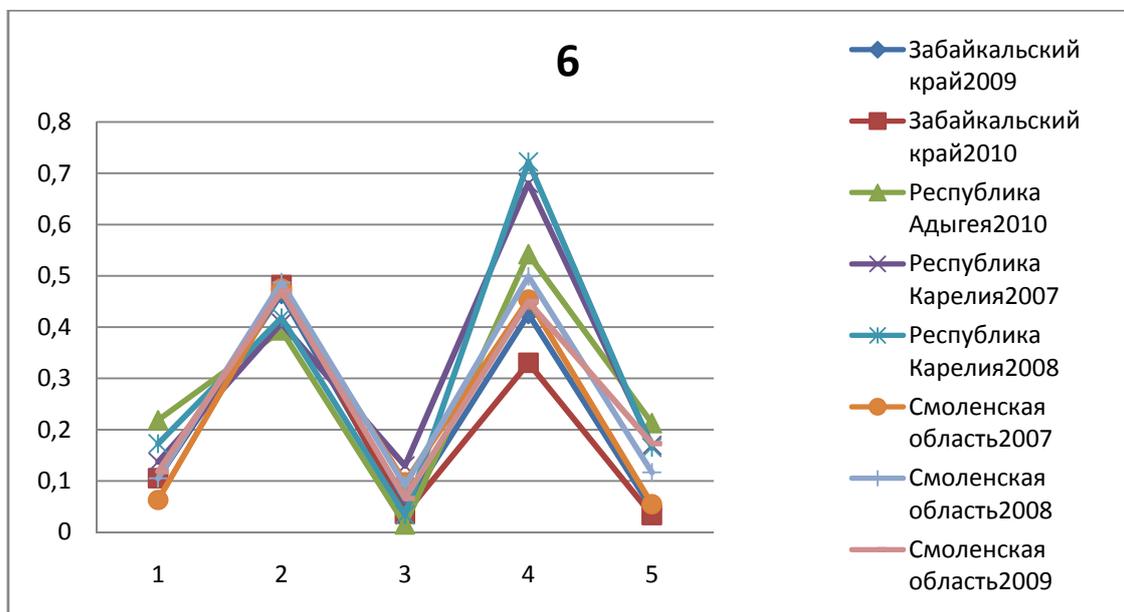
В 3-й паттерн входят 14 объектов. Этот паттерн мало отличается от второго паттерна – лишь показатель «Результативность исследований и разработок» в этом паттерне ниже, чем во втором, а показатель «Результативность инновационной деятельности» наоборот, выше.



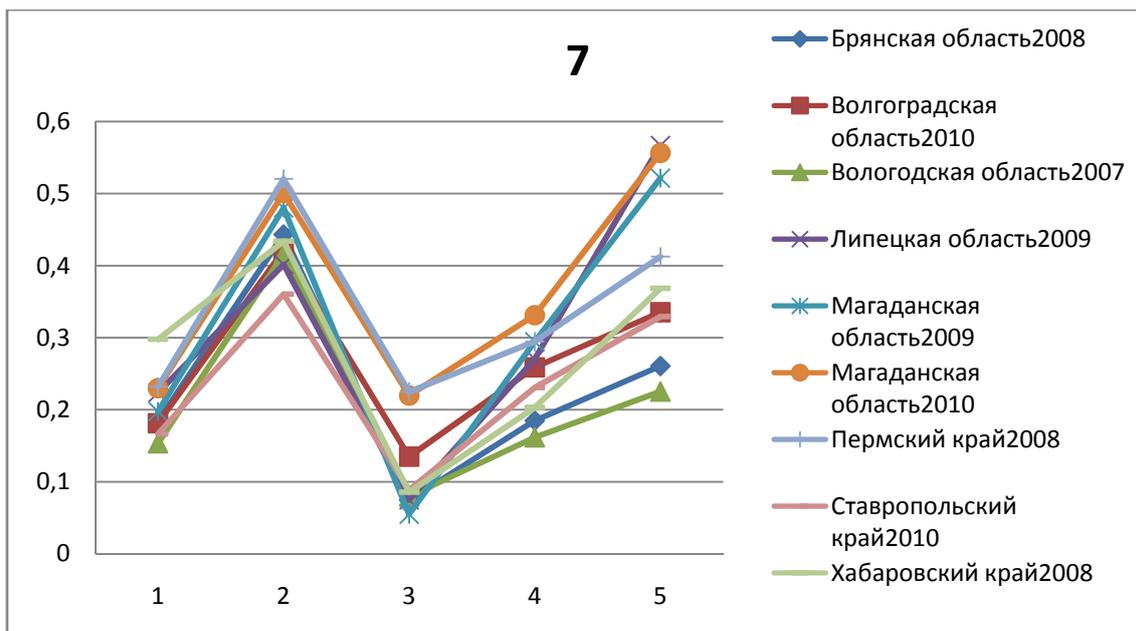
4-й паттерн во многом похож на второй, отличия заключаются в меньших значениях показателя «Социально-экономические условия» и более высоких значениях показателя «Потенциала инновационной деятельности». В этот паттерн входят 23 объекта.



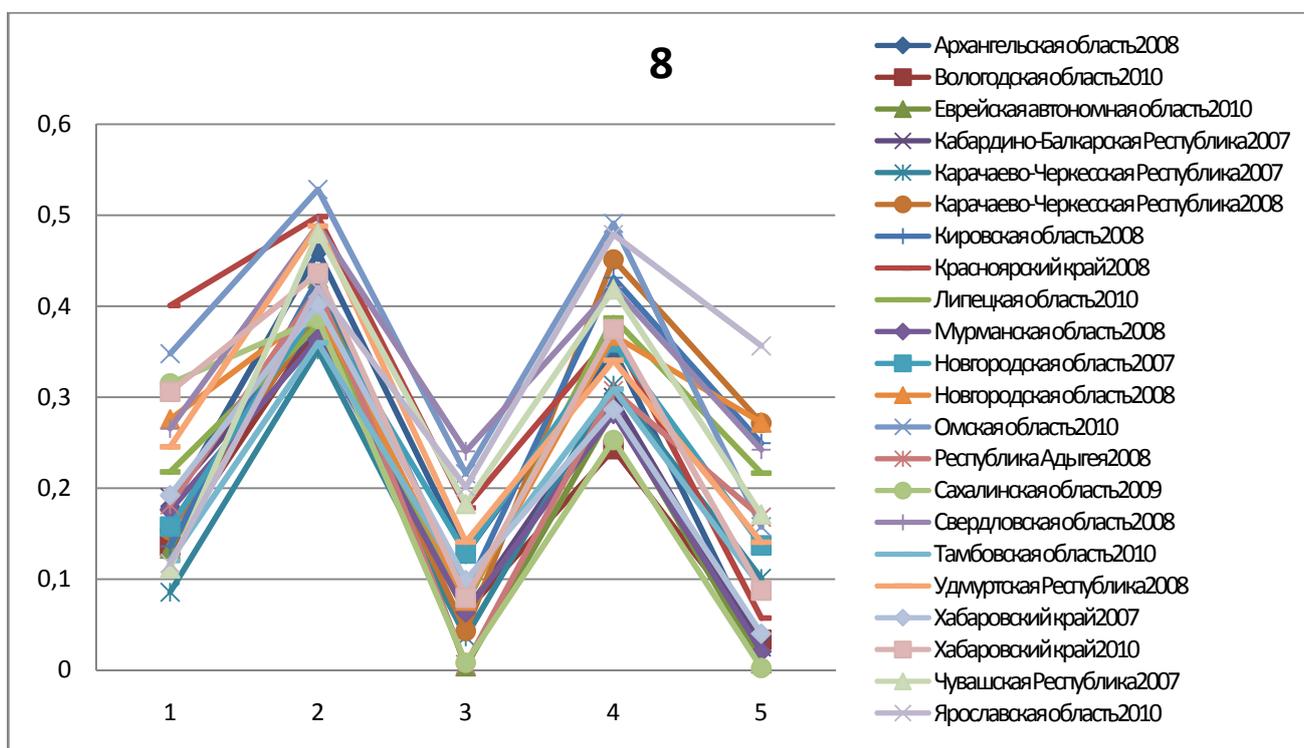
5-й паттерн представляет 18 объектов, имеющих очень высокий уровень образовательного потенциала, приемлемый уровень социально-экономического развития, слабый уровень потенциала инновационной деятельности и крайне низкие значения показателей результативности как науки, так и инновационной деятельности.



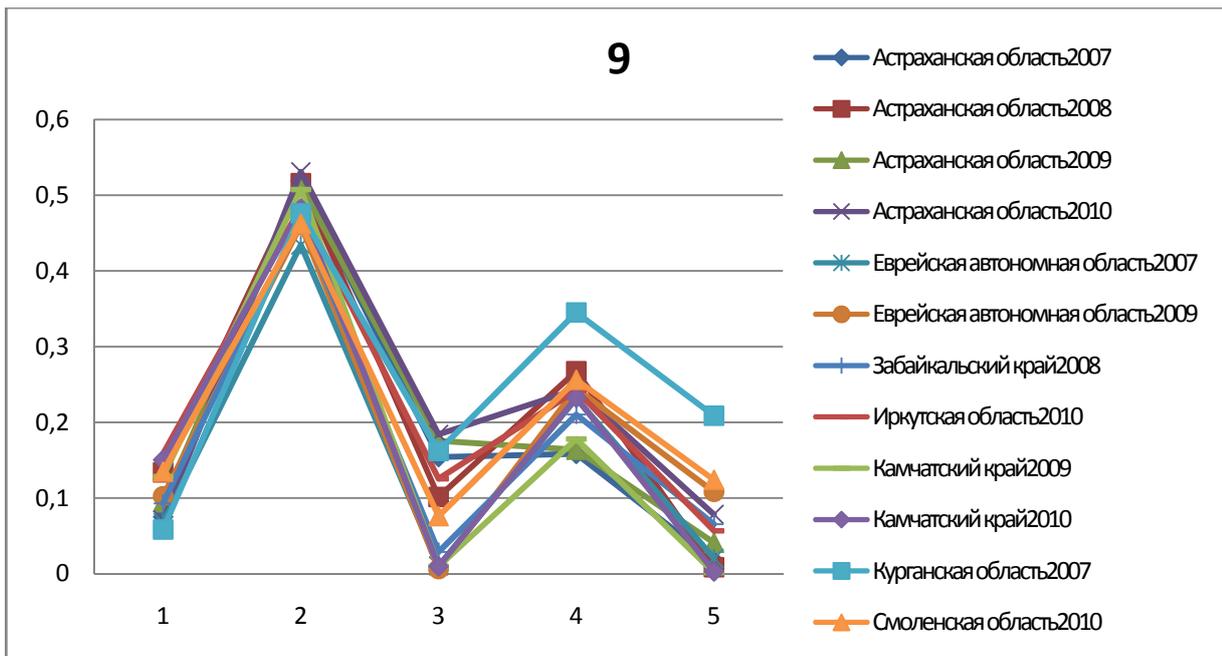
6-й паттерн – это 8 объектов, имеющих высокий уровень потенциала инновационной деятельности и образования, но крайне низкие значения как общего социально-экономического уровня, так и результативности науки и инновационной деятельности.



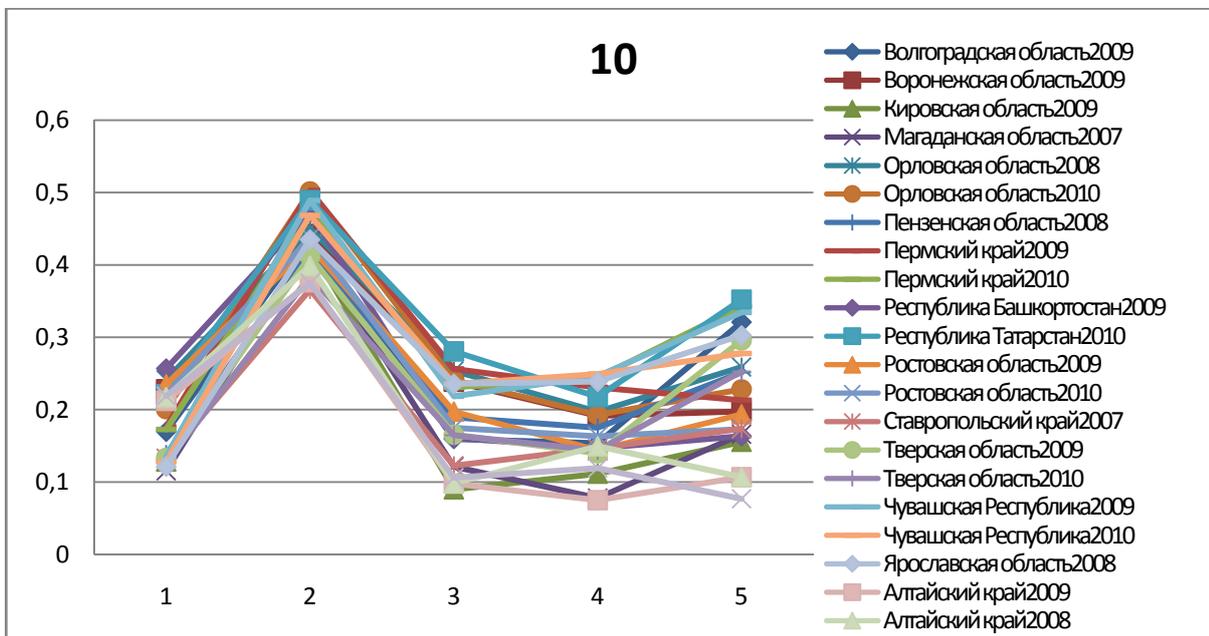
В 7-м паттерне находятся 9 объектов. Каждый из них имеет удовлетворительные показатели образовательного потенциала и развития инновационной деятельности, но весьма невысокие значения по уровню социально-экономического развития и результативности научной деятельности.



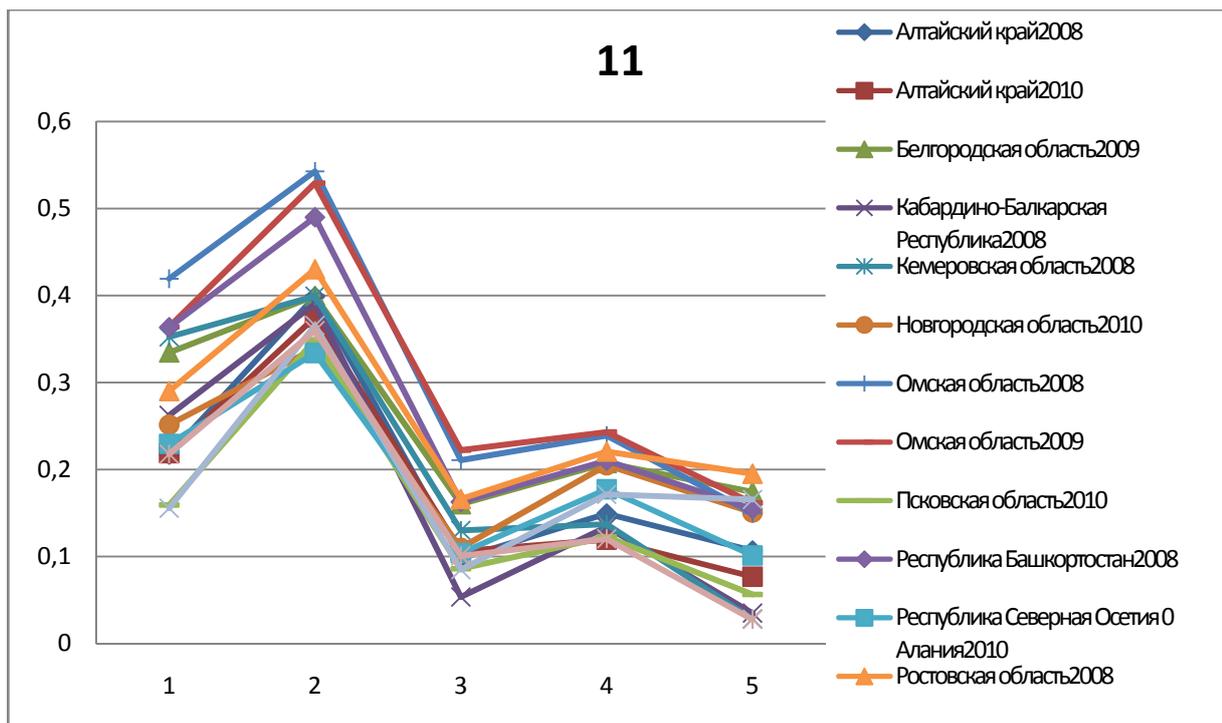
8-й паттерн представляют 22 объекта, отличительными чертами которых являются неплохие данные по уровню образовательного и инновационного потенциала, невысокий уровень социально-экономических условий и малая результативность науки и инновационной деятельности.



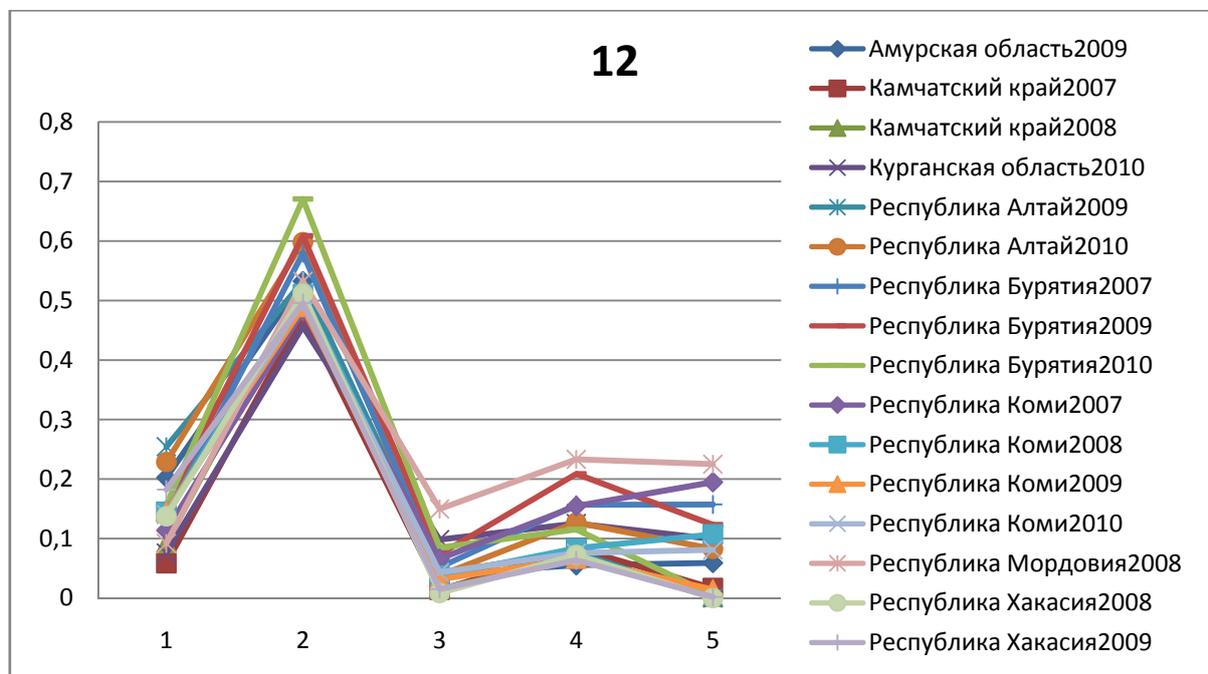
В 9-м паттерне находятся 12 объектов, которые при низких значениях показателей социально-экономических условий, результативности науки и инновационной деятельности, тем не менее имеют весьма неплохой потенциал как в образовании, так и в инновационной деятельности.



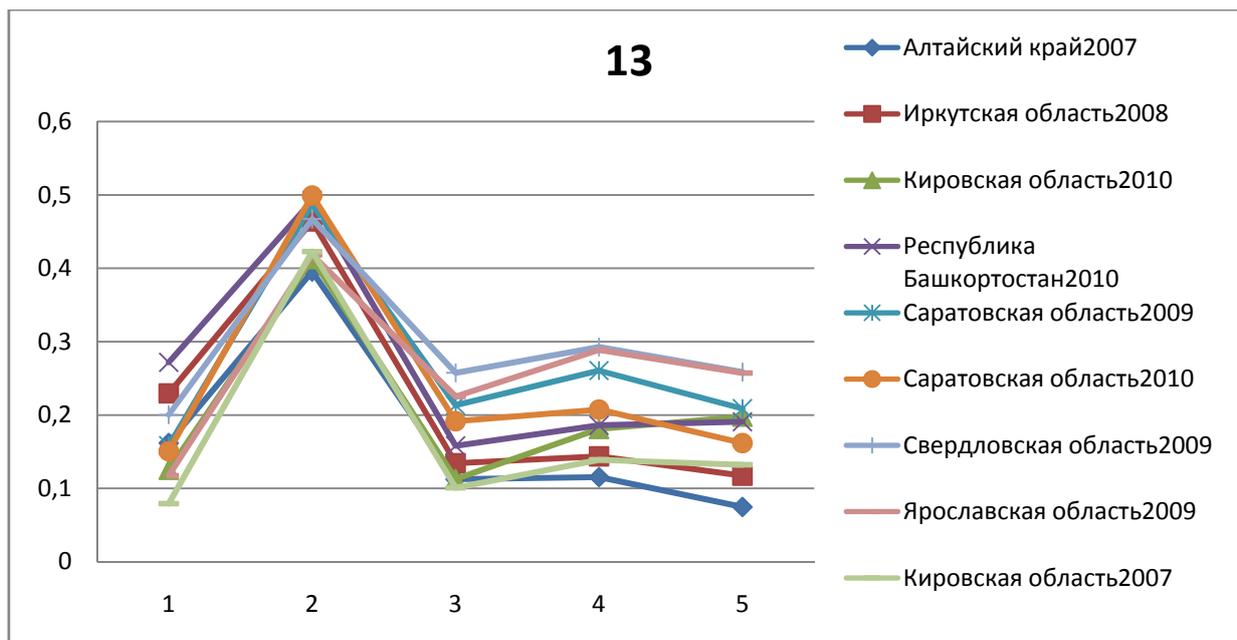
10-й паттерн – это 20 объектов, имеющих небольшие (0,1–0,3) значения всех показателей, кроме образовательного потенциала, который показывает средний (0,4–0,5) уровень.



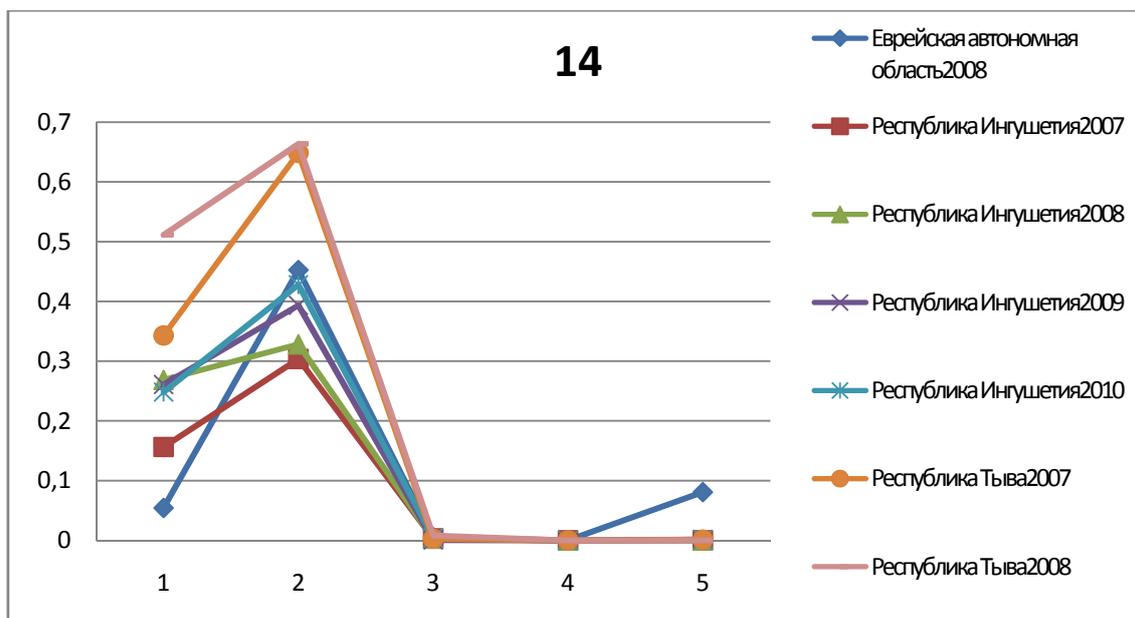
11-й паттерн может быть охарактеризован следующим образом: 14 объектов, находящихся в нем, имеют низкие значения показателей развития науки и инновационной деятельности, чуть более высокий уровень социально-экономических условий и средний уровень образовательного потенциала объектов.



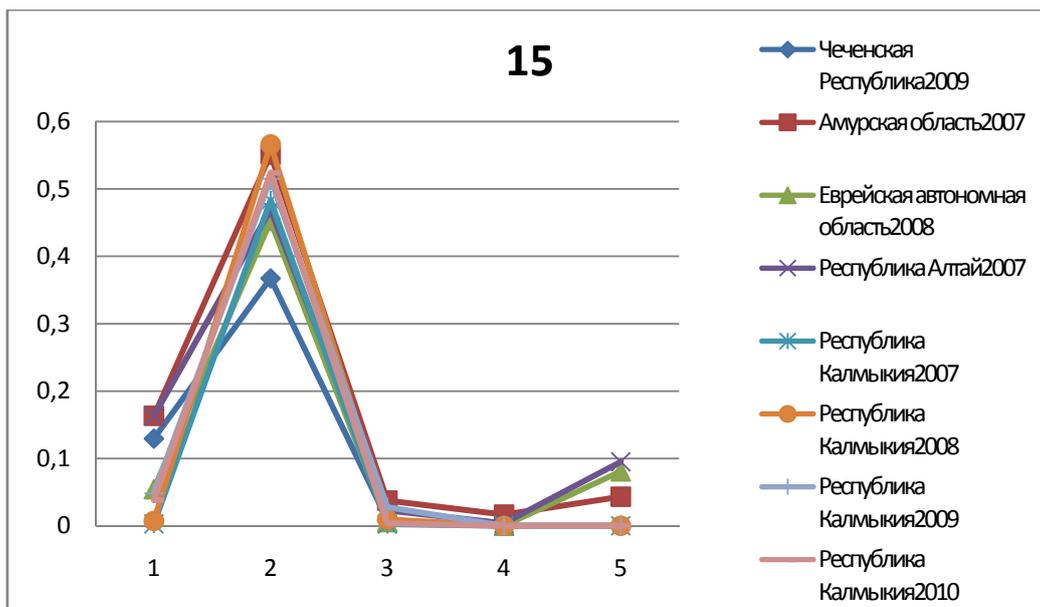
16 объектов, составляющих 12-й паттерн, имеют крайне низкие (0–0,2) значения всех показателей, кроме образовательного потенциала – у всех объектов данный показатель имеет значения 0,4–0,7.



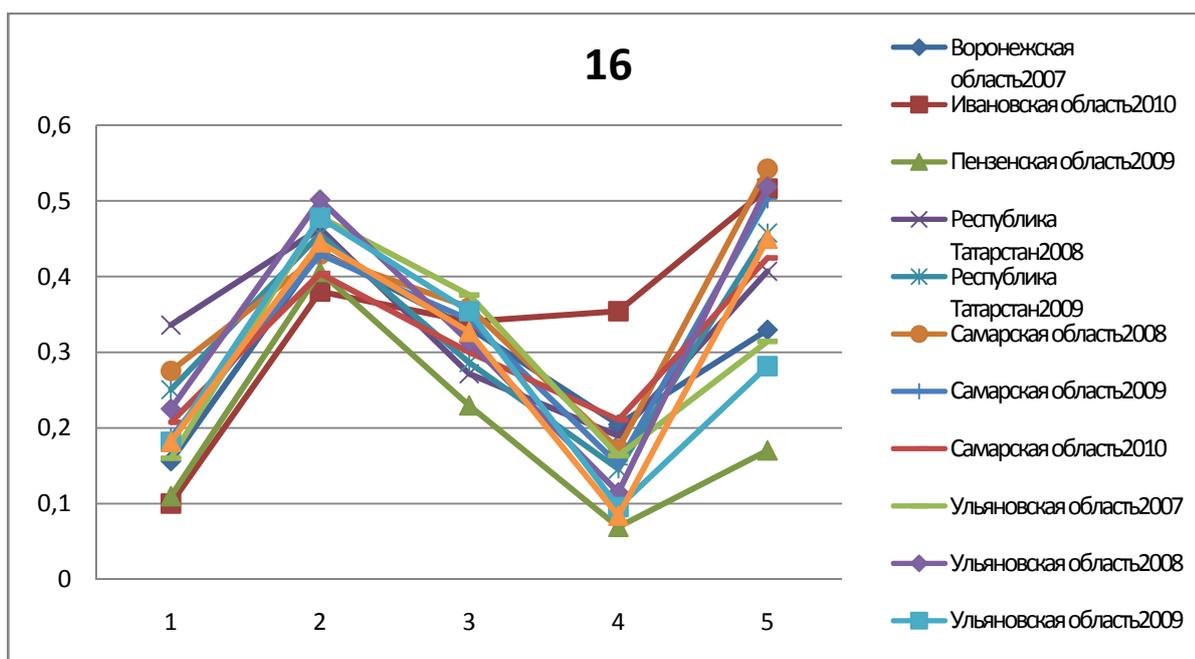
В 13-м паттерне находятся 9 регионов, имеющих похожую структуру показателей, как в 12 паттерне, лишь разрыв между образовательным потенциалом и остальными параметрами стал меньше (0,4–0,5 для Блока 2 и 0,1–0,3 для остальных показателей).



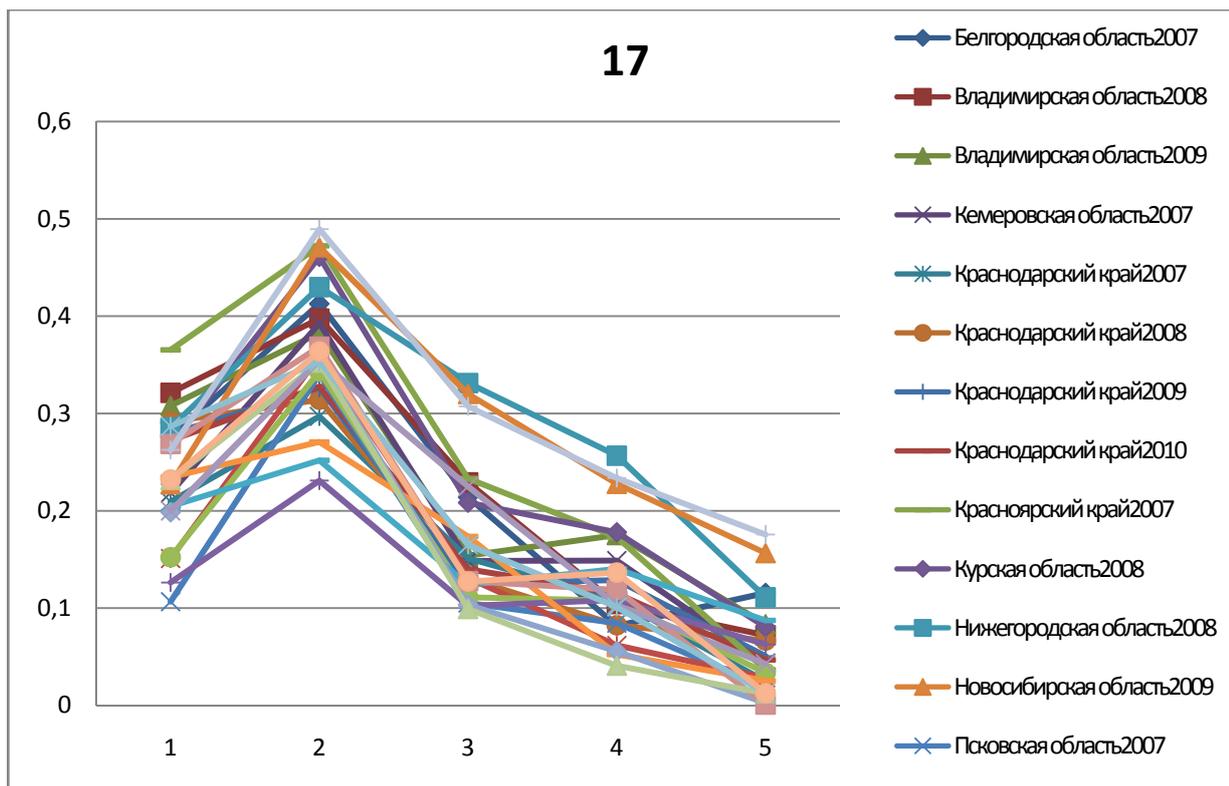
14-й паттерн состоит из 8 объектов, имеющих нулевые значения трех последних показателей, средний уровень социально-экономических условий и очень высокий потенциал образования.



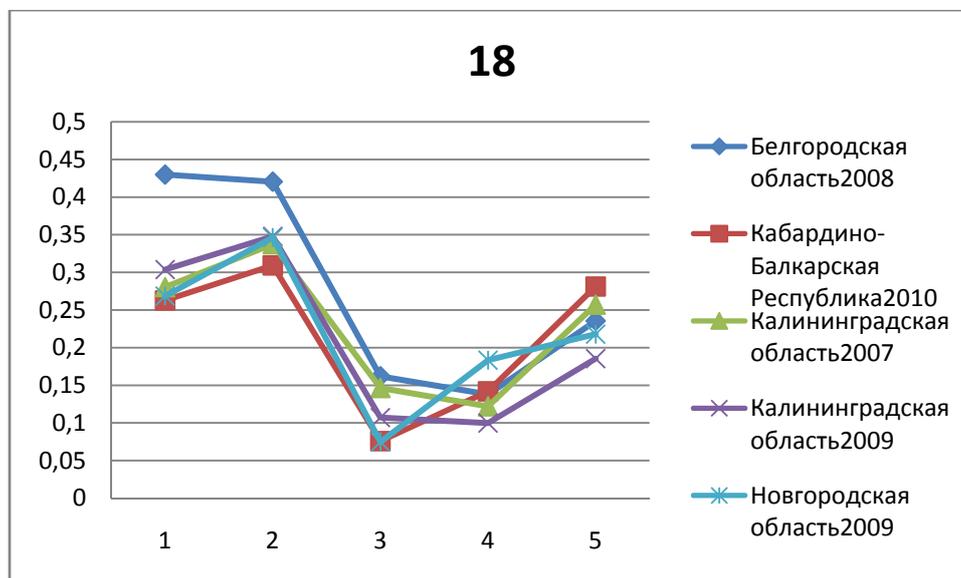
15-й паттерн весьма похож на предыдущий, отличие заключается в более низких значениях первого показателя. В этот паттерн входят 8 объектов.



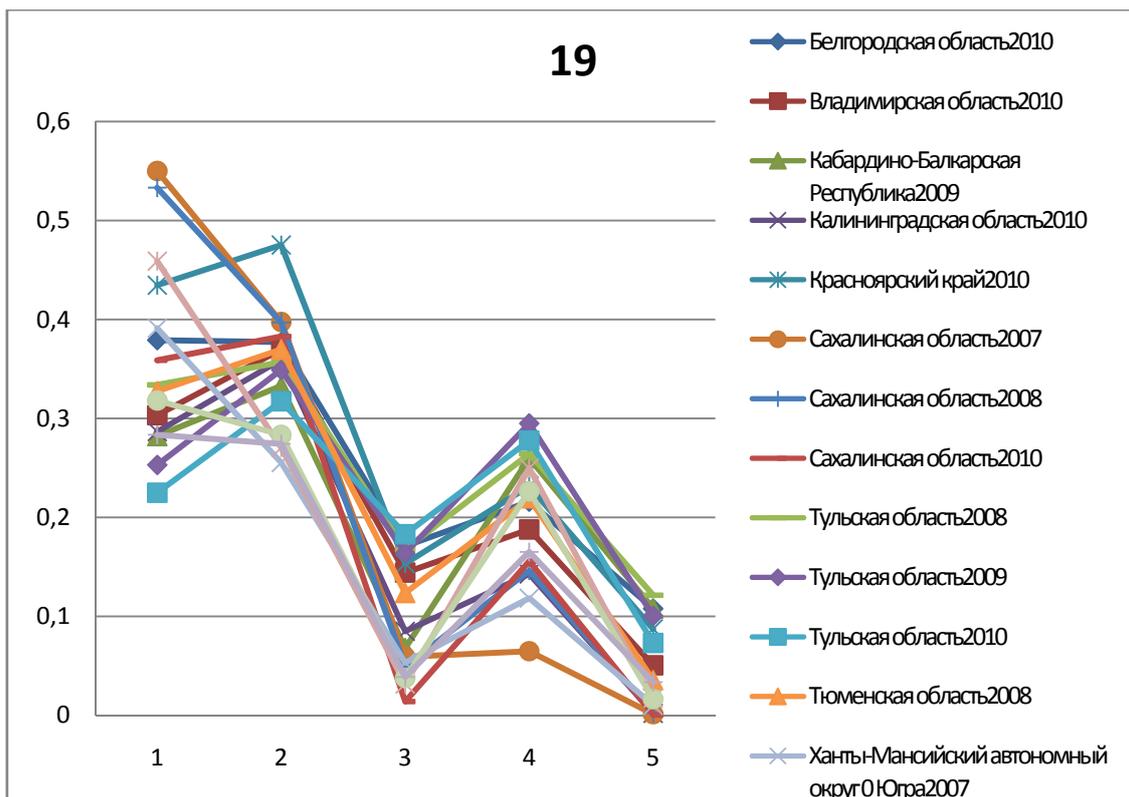
В 16-м паттерне находятся 12 объектов, имеющих средние показатели образовательного потенциала и результативности инновационной деятельности, чуть более низкие значения результативности науки, но низкие (0,1–0,3) значения показателей социально-экономических условий и потенциала инновационной деятельности.



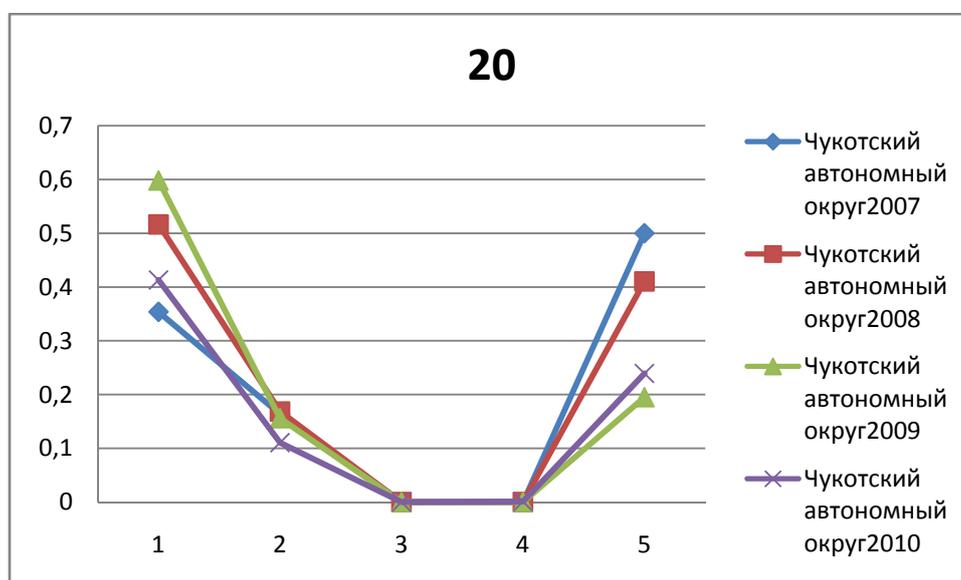
В 17-м паттерне находятся 25 объектов. Основные характеристики: умеренные показатели экономических условий, образования и науки в регионе, но низкие (с тенденцией к нулевому уровню отдачи от инноваций) показатели инновационной деятельности.



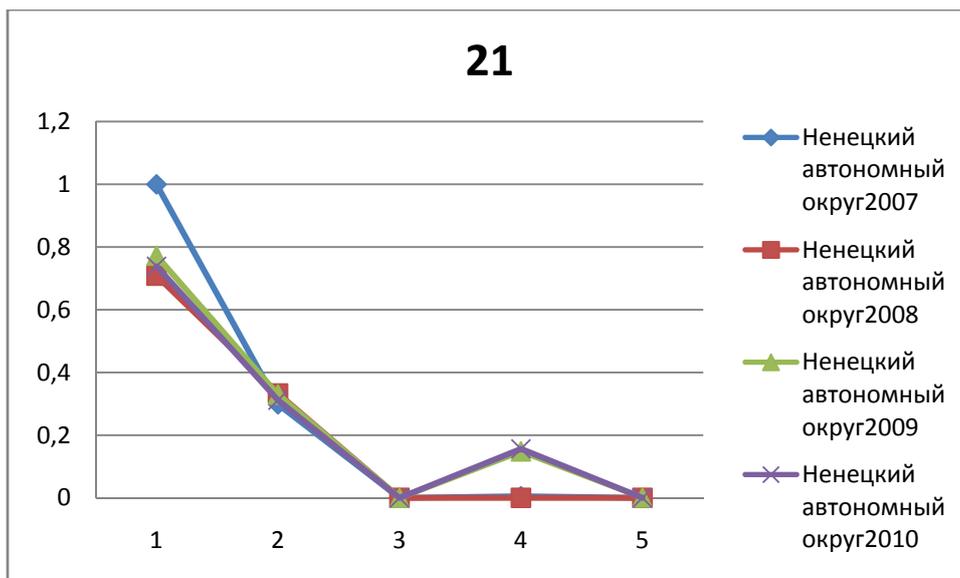
18-й паттерн – это 5 объектов, имеющих низкий уровень по Блокам 4, 5, 6, и умеренные значения по первым двум блокам.



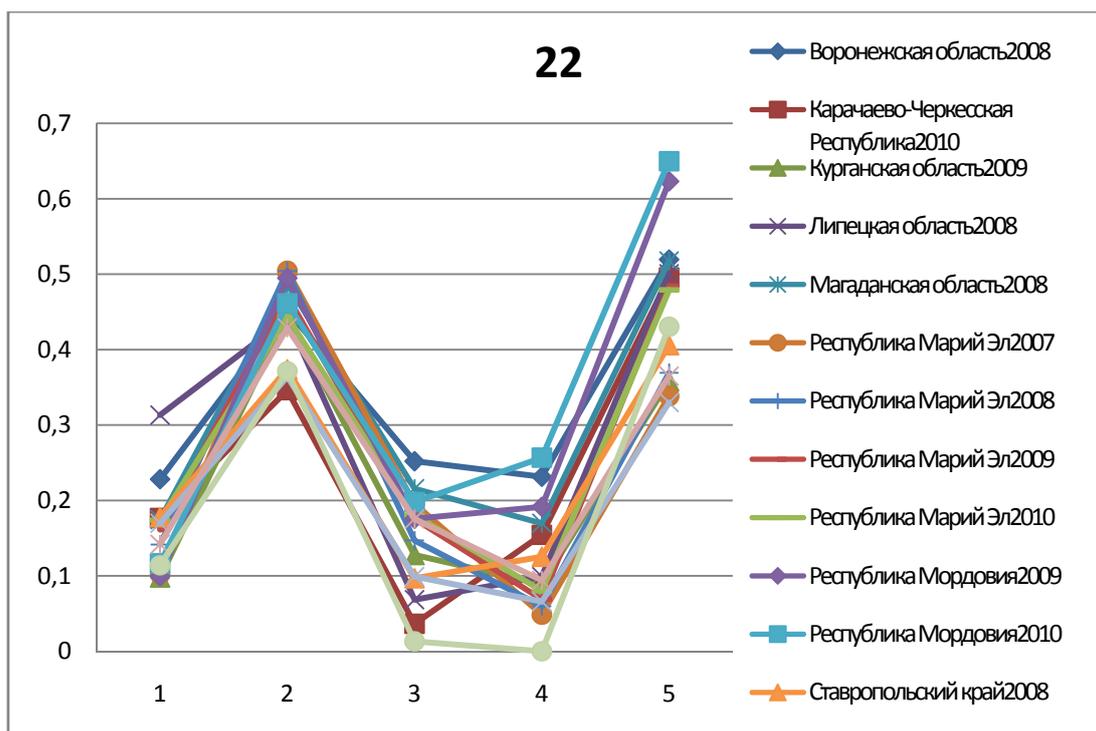
19-й паттерн можно описать как множество регионов, сходных в наличии неплохих социально-экономических условий, образовательного потенциала, но слабо использующих возможности инновационной и научной деятельности. Всего в этом паттерне 16 объектов.



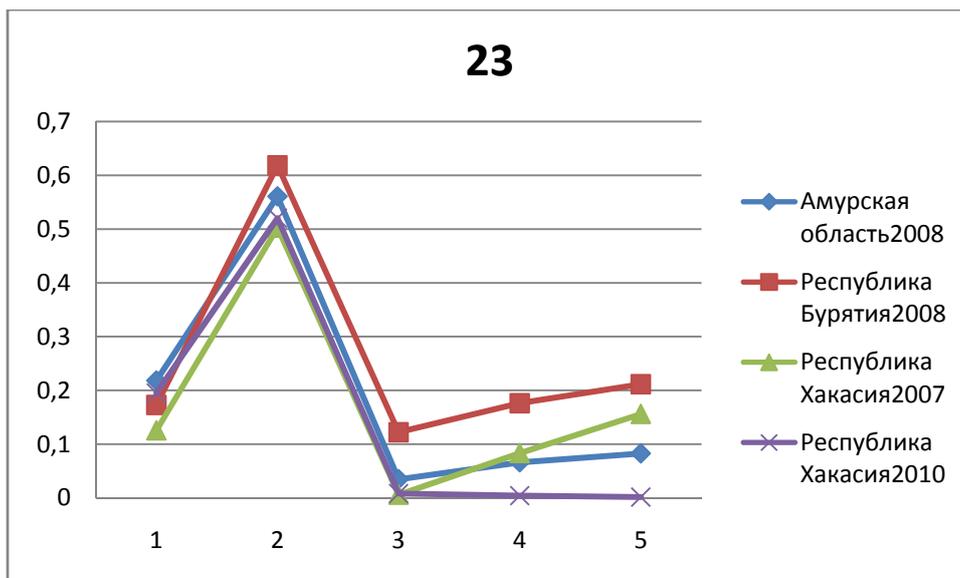
В 20-м паттерне находятся 4 объекта, представляющих Чукотский автономный округ в 4 выбранных для исследования года. Средние значения социально-экономических условий и результативности инновационной деятельности сопровождаются очень низкими показателями по всем остальным параметрам.



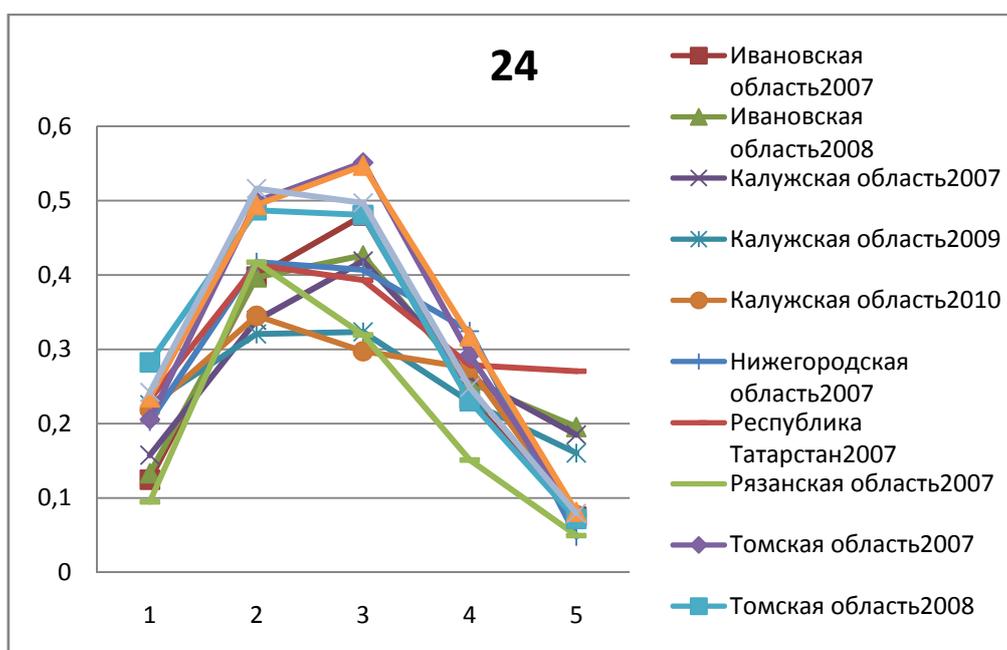
Аналогичная ситуация с 21-м паттерном. Его образуют данные по Ненецкому автономному округу. Его уникальность в очень высоких значениях социально-экономических условий, но крайне низких остальных показателях.



22-й паттерн – это 15 объектов, выделяющихся очень высокой результативностью инновационной деятельности, при умеренном образовательном потенциале, и слабом уровне оставшихся показателей.



23-й паттерн образуют объекты с уровнем образовательного потенциала выше среднего, но очень низкими значениями остальных показателей.



В 24-м паттерне 12 объектов, которые можно охарактеризовать как обладающие умеренными данными для образовательного потенциала и научной деятельности, но более низкими показателями по инновационной деятельности и социально-экономическим условиям.



25-й паттерн – это собрание уникальных паттернов, ни один из которых не попал в предыдущие 24 паттерна. Всего таких объектов 24.

3.3. Динамический анализ паттернов

Динамический анализ паттернов позволяет отследить, какому паттерну из вышеописанных следовал каждый из регионов в исследуемые 4 года, на основе траекторий развития объекта и динамических групп.

Траектория развития объекта – это определенное чередование паттернов, которое полностью описывает изменение направления развития объекта на рассматриваемом горизонте анализа. Динамическая группа – это совокупность объектов, характеризующаяся идентичными траекториями развития. Все элементы динамической группы одинаково реагируют на изменения внешней среды и характеризуются схожими стратегиями развития. В данном исследовании мы выделили только группу абсолютно устойчивых регионов.

Все данные приведены в табл. 3.

Таблица 3. Динамический анализ паттернов

Регион	2007	2008	2009	2010
Алтайский край	13	11	10	11
Амурская область	15	23	12	5
Архангельская область	2	8	2	2
Астраханская область	9	9	9	9
Белгородская область	17	18	11	19

Брянская область	3	7	3	3
Владимирская область	4	17	17	19
Волгоградская область	2	3	10	7
Вологодская область	7	5	3	8
Воронежская область	16	22	10	25
г. Москва	1	1	1	1
г. Санкт-Петербург	1	25	1	1
Еврейская автономная область	9	14	9	8
Забайкальский край	3	9	6	6
Ивановская область	24	24	25	16
Иркутская область	2	13	2	9
Кабардино-Балкарская Республика	8	11	19	18
Калининградская область	18	25	18	19
Калужская область	24	25	24	24
Камчатский край	12	12	9	9
Карачаево-Черкесская Республика	8	8	5	22
Кемеровская область	17	11	2	2
Кировская область	13	8	10	13
Костромская область	4	25	3	3
Краснодарский край	17	17	17	17
Красноярский край	17	8	2	19
Курганская область	9	3	22	12
Курская область	4	17	4	4
Ленинградская область	25	25	25	25
Липецкая область	5	22	7	8
Магаданская область	10	22	7	7
Московская область	25	25	25	25
Мурманская область	2	8	2	2
Ненецкий автономный округ	21	21	21	21
Нижегородская область	24	17	25	25
Новгородская область	8	8	18	11
Новосибирская область	4	2	17	4
Омская область	2	11	11	8
Оренбургская область	5	5	5	5
Орловская область	4	10	25	10
Пензенская область	4	10	16	4
Пермский край	4	7	10	10
Приморский край	1	1	1	1
Псковская область	17	17	17	11
Республика Адыгея	3	8	5	6
Республика Алтай	15	5	12	12
Республика Башкортостан	2	11	10	13
Республика Бурятия	12	23	12	12
Республика Дагестан	17	17	17	25
Республика Ингушетия	14	14	14	14
Республика Калмыкия	15	15	15	15

Республика Карелия	6	6	5	3
Республика Коми	12	12	12	12
Республика Марий Эл	22	22	22	22
Республика Мордовия	25	12	22	22
Республика Саха (Якутия)	5	5	5	5
Республика Северная Осетия – Алания	17	17	17	11
Республика Татарстан	24	16	16	10
Республика Тыва	14	15	5	5
Республика Хакасия	23	12	12	23
Ростовская область	2	11	10	10
Рязанская область	24	4	4	4
Самарская область	25	16	16	16
Саратовская область	4	4	13	13
Сахалинская область	19	19	8	19
Свердловская область	4	8	13	4
Смоленская область	6	6	6	9
Ставропольский край	10	22	22	7
Тамбовская область	3	3	11	8
Тверская область	4	22	10	10
Томская область	24	24	24	24
Тульская область	17	19	19	19
Тюменская область	17	19	17	11
Удмуртская Республика	2	8	2	2
Ульяновская область	16	16	16	16
Хабаровский край	8	7	5	8
Ханты-Мансийский автономный округ – Югра	19	19	19	19
Челябинская область	4	17	4	4
Чеченская Республика	14	15	14	22
Чувашская Республика	8	3	10	10
Чукотский автономный округ	20	20	20	20
Ямало-Ненецкий автономный округ	25	25	25	25
Ярославская область	4	10	13	8

14 регионов из 83 оказались абсолютно устойчивыми, т.е. ни разу за 4 года не изменили выбранному паттерну. Траектории таких регионов представлены на рис. 5.

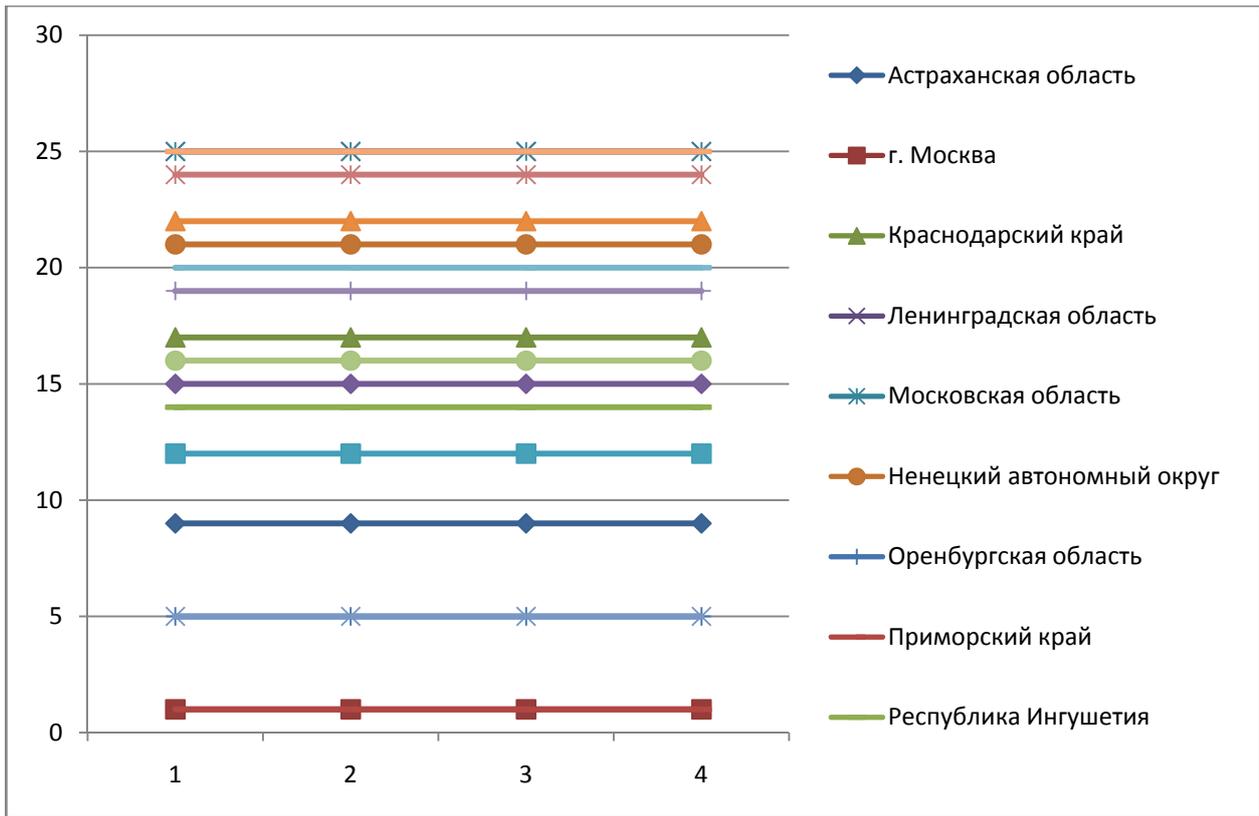
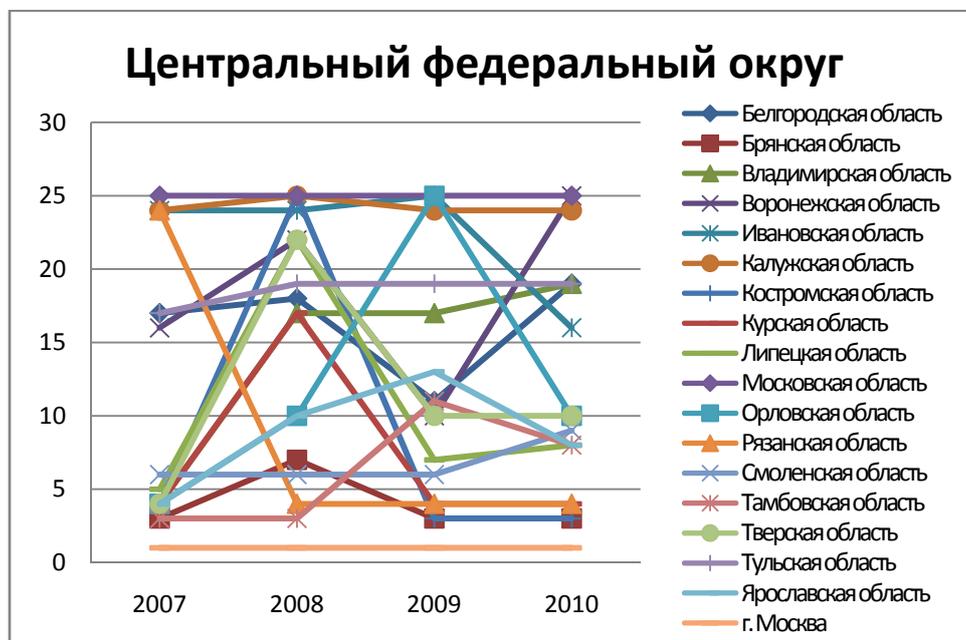


Рис. 5. Абсолютно устойчивые регионы

Остальные регионы сменили паттерны хотя бы раз. Для удобства мы изображаем графики траекторий не по всем регионам сразу, а по федеральным округам.



Как видно, в Центральном федеральном округе регионы имеют самые разные паттерны и, к тому же, часто меняют их. Впрочем, и большинство остальных регионов Российской Федерации также непостоянны. Все остальные данные приведены в Приложениях 3 и 4.

Выводы

Анализ паттернов данных является сравнительно молодым методом анализа данных, доказавшим свою эффективность и способность решать задачи, недоступные даже своему «прародителю» кластерному анализу. Способность строить не только статические классификации исходных объектов для анализа их структуры, но и исследовать процессы развития объектов, изучать законы изменения характеристик этих объектов во времени и задавать новые классификации на основе динамики развития составляет одно из основных преимуществ метода анализа паттернов данных.

Этот метод уже успешно зарекомендовал себя при решении таких разноплановых задач, как анализ банковской сферы, разработка процедуры банкротства кредитных организаций, принятие решений в менеджменте и управлении персоналом, исследовании уровня поддержки политических партий на основании результатов выборов и т.д.

Применение анализа паттернов к данным науки, образования и инновационной деятельности представляет огромный интерес как с исследовательской, так и с прикладной точки зрения.

Анализ паттернов и динамический анализ паттернов данных решает такие задачи, как:

- проведение классификации данных науки, образования и инновационной деятельности на основании различных характеристик;
- исследование траекторий развития найденных паттернов данных с течением времени и рассмотрении их в динамическом развитии;
- построение классификации объектов с учетом их специфики и тенденций измерения.

В результате исследования были изучены характеристики регионов России по таким показателям как уровень социально-экономических условий, потенциал и результативность науки, образования и инновационной деятельности в динамике за 4 года с 2007 по 2010 г. Получена классификация регионов по схожести внутренней структуры указанных показателей, также построены траектории развития регионов с течением времени и найдены группы регионов, придерживающихся выбранной стратегии.

Несомненный интерес представляет использование полученной информации для дальнейшего решения следующих задач:

- 1) Выявления неявно выраженного взаимного влияния индикаторов.
- 2) Определения характерных векторов направленности индикаторов и их качественная оценка для различных аналитических измерений:
 - разных видов экономической деятельности;

- высоко-, средне-, и низкотехнологичных отраслей;
 - федеральных округов;
 - регионов-доноров и дотационных регионов;
 - форм собственности предприятий.
- 3) Выявления трендов изменения индикаторов (показателей, рассчитываемых в ходе агрегации и преобразования статистических данных).
 - 4) Идентификации взаимного влияния траекторий развития индикаторов (их агрегатов).
 - 5) Выявления нетипичной динамики индикаторов.

Литература

1. Aleskerov F., Alper C.E. Inflation, Money, and Output Growths: Some Observations // Bogazici University Research Paper, No. SBE 96-06, 1996.
2. Aleskerov F., Alper C.E. A clustering approach to some monetary facts: a long-run analysis of cross-country data // The Japanese Economic Review. 2000. Vol. 51. No. 4. P. 555–567.
3. Aleskerov F., Ersel H., Gundes C., Minibas A., Yolalan R. Environmental Grouping of Bank Branches and their Performances // Yapi Kredi Discussion Paper Series. No. 97–03, 1997, Istanbul, Turkey.
4. Aleskerov F., Ersel H., Gundes C., Yolalan R. A Multicriterial Method for Personnel Allocation among Bank Branches // Yapi Kredi Discussion Paper Series, No. 98–01, 1998, Istanbul, Turkey.
5. Aleskerov F., Ersel H., Yolalan R. Multicriterial Ranking Approach for Evaluating Bank Branch Performance // International Journal of Information Technology and Decision Making. 2004. Vol. 3. No. 2. P. 321–335.
6. Aleskerov F.T., Martynova Y.I., Solodkov V.M. Assessment and analysis of the efficiency of banks and banking systems // Proceedings of the 2d International Conference “Mathematical Modeling of Social and Economic Dynamic” (MMSED - 2007), Moscow, 2007, 13–15 (ISBN 978-5-209-02632-7).
7. Aleskerov F., Nurmi H. A Method for Finding Patterns of Party Support and Electoral Change: An Analysis of British General and Finnish Municipal Elections // Mathematical and Computer Modelling, 2008, 1225–1253.
8. Few S. Multivariate Analysis Using Parallel Coordinates, 2006 (http://www.perceptualedge.com/articles/b-eye/parallel_coordinates.pdf).
9. Алексашин П.Г., Алескеров Ф.Т., Белоусова В.Ю., Попова Е.С., Солодков В.М. Динамический анализ бизнес-моделей российских банков в период 2006–2009 гг.: препринт WP7/2012/03. М.: Изд. дом НИУ ВШЭ, 2012.
10. Индикаторы науки: 2008. Статистический сборник. М.: ГУ ВШЭ, 2008 (<http://issek.hse.ru/news/49369919.html>).
11. Индикаторы образования: 2008. Статистический сборник. М.: ГУ ВШЭ, 2008 (<http://issek.hse.ru/news/49370471.html>).
12. Индикаторы инновационной деятельности: 2008. Статистический сборник. М.: ГУ ВШЭ, 2008 (<http://issek.hse.ru/news/49369377.html>).

13. Миркин Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор: препринт WP7/2011/03. М.: Изд. дом Национального исследовательского университета «Высшая школа экономики», 2011.
14. Рейтинг инновационного развития субъектов Российской Федерации: аналитический доклад / под ред. Л.М. Гохберга. М.: Национальный исследовательский университет «Высшая школа экономики», 2012.
15. Статистический сборник Федеральной службы государственной статистики «Российский статистический ежегодник» (http://www.gks.ru/bgd/regl/b11_13/Main.htm).
16. Статистический сборник Федеральной службы государственной статистики «Обследование населения по проблемам занятости» (http://www.gks.ru/bgd/regl/b12_30/Main.htm).
17. Статистический сборник Федеральной службы государственной статистики «Регионы России. Социально-экономические показатели» (http://www.gks.ru/bgd/regl/b11_14p/Main.htm).

Приложение 1. Матрицы корреляции показателей

2007	Блок 1	Блок 2	Блок 3.	Блок 4.	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1					
Блок 2. Образовательный потенциал	-0,2	1				
Блок 3. Потенциал научно-технической деятельности	0,16	-0,1	1			
Блок 4. Результативность исследований и разработок	0,11	0,11	0,7	1		
Блок 5. Потенциал инновационной деятельности	-0,2	0,14	0,22	0,33	1	
Блок 6. Результативность инновационной деятельности	-0,1	0,05	0,05	0,13	0,21	1

2008	Блок 1	Блок 2	Блок 3	Блок 4	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1,00					
Блок 2. Образовательный потенциал	-0,21	1,00				
Блок 3. Потенциал научно-технической деятельности	0,25	-0,12	1,00			
Блок 4. Результативность исследований и разработок	0,18	0,04	0,70	1,00		
Блок 5. Потенциал инновационной деятельности	-0,11	0,06	0,16	0,22	1,00	
Блок 6. Результативность инновационной деятельности	-0,08	0,04	0,16	0,24	0,11	1,00

2009	Блок 1	Блок 2	Блок 3	Блок 4	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1,00					
Блок 2. Образовательный потенциал	-0,19	1,00				
Блок 3. Потенциал научно-технической деятельности	0,21	-0,02	1,00			
Блок 4. Результативность исследований и разработок	0,11	0,12	0,69	1,00		
Блок 5. Потенциал инновационной деятельности	-0,03	0,26	0,27	0,35	1,00	
Блок 6. Результативность инновационной деятельности	-0,17	0,08	0,14	0,24	0,16	1,00

2010	Блок 1	Блок 2	Блок 3	Блок 4	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1,00					
Блок 2. Образовательный потенциал	-0,13	1,00				
Блок 3. Потенциал научно-технической деятельности	0,16	0,00	1,00			
Блок 4. Результативность исследований и разработок	0,10	0,06	0,58	1,00		
Блок 5. Потенциал инновационной деятельности	-0,07	0,15	-0,01	0,19	1,00	
Блок 6. Результативность инновационной деятельности	-0,19	-0,04	0,18	0,26	0,20	1,00

8	Удельный вес персонала, занятого исследованиями и разработками, в общей численности экономически активного населения в экономике региона, % (агрегированный)	0,03	0,29	0,63	-0,12	-0,34	0,84	0,05	1,00											
9	Удельный вес лиц, имеющих ученую степень в численности исследователей, % (агрегированный)	-0,07	-0,05	-0,09	-0,01	0,07	-0,18	-0,08	-0,16	1,00										
10	Число патентных заявок на изобретения, поданных в Роспатент национальными заявителями в расчете на 10 000 человек экономически активного населения региона	-0,08	0,15	0,60	-0,09	-0,24	0,37	-0,02	0,53	-0,01	1,00									
11	Число патентных заявок на полезные модели, поданных в Роспатент национальными заявителями в расчете на 10 000 человек экономически активного населения региона	-0,04	0,17	0,62	-0,04	-0,31	0,60	0,05	0,71	-0,05	0,63	1,00								
12	Число выданных охраняемых документов в расчете на 10 000 человек экономически активного населения региона	-0,07	0,17	0,65	-0,07	-0,27	0,44	-0,04	0,62	-0,02	0,96	0,74	1,00							
13	Число патентных заявок на полезные модели, выданных в Роспатент национальными заявителями в расчете на 10 000 человек экономически активного населения региона	-0,03	0,19	0,64	-0,01	-0,31	0,63	0,06	0,75	-0,08	0,62	0,97	0,72	1,00						

14	Число созданных передовых производственных технологий в расчете на 10 000 человек экономически активного населения региона	-0,01	0,13	0,36	-0,03	-0,28	0,69	0,17	0,63	-0,16	0,19	0,36	0,21	0,41	1,00						
15	Удельный вес организаций, приобретавших новые технологии, в общем числе организаций, осуществлявших технологические инновации, проценты (ДП)	-0,20	-0,19	0,12	0,05	-0,08	0,13	0,03	0,03	0,14	-0,02	0,16	0,01	0,15	0,12	1,00					
16	Удельный организаций, передававших новые технологии, в общем числе организаций, осуществлявших технологические инновации, проценты (ДП)	-0,09	0,05	0,32	0,11	-0,16	0,37	0,00	0,41	-0,05	0,30	0,30	0,32	0,32	0,42	0,18	1,00				
17	Затраты на технологические инновации, в процентах от общего объема отгруженных товаров, выполненных работ, услуг (ДП)	-0,18	-0,19	0,24	0,10	-0,18	0,26	-0,08	0,19	-0,21	0,21	0,29	0,25	0,32	0,27	0,20	0,25	1,00			
18	Затраты на технологические инновации, в процентах от общего объема отгруженных товаров, выполненных работ, услуг (связь)	-0,05	-0,18	-0,03	0,11	0,02	-0,11	0,02	-0,11	-0,04	-0,12	-0,12	-0,17	-0,10	0,12	0,12	0,35	0,15	1,00		
19	Объем инновационных товаров, работ, услуг, проценты (ДП)	-0,15	-0,07	0,16	0,12	-0,06	0,18	-0,14	0,14	-0,14	0,13	0,28	0,16	0,33	0,25	0,08	-0,01	0,58	0,05	1,00	
20	Объем инновационных товаров, работ, услуг, проценты (Связь)	-0,07	-0,03	-0,11	0,02	-0,03	-0,05	0,18	-0,06	0,00	-0,09	-0,12	-0,14	-0,09	0,03	-0,05	0,02	-0,15	0,30	-0,01	1,00

Приложение 2. Значения показателей для регионов России

Регион	Блок 1. Социально-экономические условия	Блок 2. Образовательный потенциал	Блок 3. Результативность исследований и разработок	Блок 4. Потенциал инновационной деятельности	Блок 5. Результативность инновационной деятельности	Номер паттерна
г. Москва 2007	0,47	0,43	0,84	0,21	0,07	1
г. Москва 2008	0,49	0,43	0,75	0,26	0,10	1
г. Москва 2009	0,36	0,47	0,91	0,30	0,07	1
г. Москва 2010	0,35	0,42	0,89	0,21	0,11	1
г. Санкт-Петербург 2007	0,34	0,49	0,68	0,29	0,07	1
г. Санкт-Петербург 2009	0,45	0,51	0,69	0,21	0,19	1
г. Санкт-Петербург 2010	0,42	0,48	0,60	0,19	0,23	1
Приморский край 2007	0,18	0,36	0,53	0,19	0,02	1
Приморский край 2008	0,27	0,35	0,59	0,24	0,21	1
Приморский край 2009	0,35	0,36	0,67	0,25	0,21	1
Приморский край 2010	0,32	0,41	0,66	0,14	0,17	1
Архангельская область 2007	0,18	0,45	0,11	0,18	0,01	2
Архангельская область 2009	0,18	0,42	0,10	0,23	0,01	2
Архангельская область 2010	0,17	0,39	0,09	0,11	0,01	2
Волгоградская область 2007	0,15	0,41	0,15	0,23	0,02	2
Иркутская область 2007	0,23	0,43	0,10	0,21	0,04	2
Иркутская область 2009	0,15	0,46	0,14	0,18	0,02	2
Кемеровская область 2009	0,27	0,40	0,13	0,16	0,01	2
Кемеровская область 2010	0,29	0,43	0,13	0,17	0,01	2
Красноярский край 2009	0,33	0,48	0,19	0,30	0,05	2
Мурманская область 2007	0,15	0,35	0,10	0,21	0,03	2
Мурманская область 2009	0,14	0,35	0,06	0,21	0,02	2
Мурманская область 2010	0,13	0,39	0,06	0,17	0,02	2

Новосибирская область 2008	0,29	0,49	0,27	0,28	0,14	2
Омская область 2007	0,29	0,52	0,23	0,24	0,07	2
Республика Башкортостан 2007	0,24	0,48	0,19	0,18	0,11	2
Ростовская область 2007	0,19	0,42	0,19	0,25	0,08	2
Удмуртская Республика 2007	0,18	0,48	0,17	0,24	0,06	2
Удмуртская Республика 2009	0,21	0,46	0,17	0,17	0,05	2
Удмуртская Республика 2010	0,21	0,44	0,19	0,24	0,10	2
Брянская область 2007	0,11	0,39	0,08	0,18	0,16	3
Брянская область 2009	0,18	0,44	0,12	0,28	0,31	3
Брянская область 2010	0,18	0,41	0,11	0,25	0,18	3
Волгоградская область 2008	0,21	0,44	0,13	0,24	0,20	3
Вологодская область 2009	0,12	0,41	0,09	0,20	0,14	3
Забайкальский край 2007	0,06	0,43	0,03	0,27	0,22	3
Костромская область 2009	0,11	0,41	0,08	0,22	0,16	3
Костромская область 2010	0,13	0,43	0,10	0,29	0,18	3
Курганская область 2008	0,10	0,48	0,15	0,30	0,26	3
Республика Адыгея 2007	0,15	0,38	0,01	0,31	0,32	3
Республика Карелия 2010	0,15	0,47	0,04	0,18	0,12	3
Тамбовская область 2007	0,10	0,37	0,09	0,21	0,13	3
Тамбовская область 2008	0,12	0,39	0,09	0,22	0,19	3
Чувашская Республика 2008	0,17	0,50	0,19	0,35	0,25	3
Владимирская область 2007	0,23	0,39	0,26	0,30	0,11	4
Костромская область 2007	0,07	0,38	0,15	0,21	0,03	4
Курская область 2007	0,16	0,43	0,27	0,20	0,10	4
Курская область 2009	0,23	0,46	0,27	0,20	0,06	4
Курская область 2010	0,24	0,49	0,24	0,14	0,05	4
Новосибирская область 2007	0,20	0,49	0,33	0,22	0,03	4
Новосибирская область 2010	0,22	0,46	0,29	0,28	0,16	4
Орловская область 2007	0,16	0,42	0,25	0,19	0,07	4
Пензенская область 2007	0,10	0,39	0,24	0,32	0,16	4

Пензенская область 2010	0,10	0,37	0,21	0,23	0,16	4
Пермский край 2007	0,16	0,51	0,27	0,31	0,20	4
Рязанская область 2008	0,13	0,44	0,24	0,18	0,09	4
Рязанская область 2009	0,13	0,46	0,23	0,29	0,15	4
Рязанская область 2010	0,13	0,45	0,21	0,22	0,10	4
Саратовская область 2007	0,11	0,46	0,29	0,23	0,07	4
Саратовская область 2008	0,16	0,49	0,25	0,29	0,12	4
Свердловская область 2007	0,22	0,47	0,32	0,25	0,15	4
Свердловская область 2010	0,22	0,49	0,26	0,31	0,14	4
Тверская область 2007	0,09	0,41	0,17	0,15	0,04	4
Челябинская область 2007	0,20	0,46	0,36	0,31	0,08	4
Челябинская область 2009	0,19	0,48	0,34	0,30	0,11	4
Челябинская область 2010	0,20	0,45	0,29	0,29	0,05	4
Ярославская область 2007	0,09	0,43	0,31	0,32	0,07	4
Амурская область 2010	0,25	0,51	0,04	0,16	0,06	5
Вологодская область 2008	0,24	0,42	0,06	0,29	0,20	5
Карачаево-Черкесская Республика 2009	0,17	0,38	0,04	0,21	0,09	5
Липецкая область 2007	0,22	0,40	0,06	0,20	0,08	5
Оренбургская область 2007	0,28	0,46	0,09	0,25	0,07	5
Оренбургская область 2008	0,34	0,48	0,09	0,20	0,13	5
Оренбургская область 2009	0,28	0,45	0,07	0,17	0,10	5
Оренбургская область 2010	0,24	0,47	0,08	0,19	0,13	5
Республика Адыгея 2009	0,22	0,41	0,03	0,12	0,03	5
Республика Алтай 2008	0,25	0,53	0,00	0,23	0,17	5
Республика Карелия 2009	0,14	0,42	0,04	0,21	0,08	5
Республика Саха (Якутия) 2007	0,20	0,44	0,07	0,21	0,04	5
Республика Саха (Якутия) 2008	0,32	0,49	0,07	0,20	0,14	5
Республика Саха (Якутия) 2009	0,30	0,50	0,05	0,18	0,12	5
Республика Саха (Якутия) 2010	0,32	0,51	0,05	0,18	0,14	5
Республика Тыва 2009	0,51	0,65	0,07	0,31	0,04	5

Республика Тыва 2010	0,51	0,60	0,00	0,15	0,10	5
Хабаровский край 2009	0,28	0,45	0,08	0,17	0,05	5
Забайкальский край 2009	0,11	0,46	0,05	0,43	0,05	6
Забайкальский край 2010	0,11	0,48	0,04	0,33	0,03	6
Республика Адыгея 2010	0,22	0,39	0,02	0,54	0,21	6
Республика Карелия 2007	0,14	0,41	0,13	0,68	0,17	6
Республика Карелия 2008	0,17	0,42	0,03	0,72	0,17	6
Смоленская область 2007	0,06	0,48	0,10	0,45	0,05	6
Смоленская область 2008	0,11	0,49	0,09	0,50	0,12	6
Смоленская область 2009	0,12	0,47	0,07	0,45	0,17	6
Брянская область 2008	0,18	0,44	0,07	0,18	0,26	7
Волгоградская область 2010	0,18	0,42	0,13	0,26	0,34	7
Вологодская область 2007	0,15	0,42	0,08	0,16	0,23	7
Липецкая область 2009	0,22	0,40	0,07	0,27	0,57	7
Магаданская область 2009	0,20	0,48	0,05	0,29	0,52	7
Магаданская область 2010	0,23	0,50	0,22	0,33	0,56	7
Пермский край 2008	0,23	0,52	0,22	0,29	0,41	7
Ставропольский край 2010	0,17	0,36	0,09	0,23	0,33	7
Хабаровский край 2008	0,30	0,43	0,09	0,20	0,37	7
Архангельская область 2008	0,18	0,46	0,13	0,35	0,02	8
Вологодская область 2010	0,14	0,38	0,06	0,24	0,03	8
Еврейская автономная область 2010	0,13	0,43	0,00	0,29	0,01	8
Кабардино-Балкарская Республика 2007	0,19	0,38	0,06	0,30	0,03	8
Карачаево-Черкесская Республика 2007	0,09	0,35	0,04	0,31	0,10	8
Карачаево-Черкесская Республика 2008	0,15	0,39	0,04	0,45	0,27	8
Кировская область 2008	0,14	0,43	0,09	0,43	0,25	8
Красноярский край 2008	0,40	0,50	0,18	0,37	0,06	8
Липецкая область 2010	0,22	0,38	0,07	0,39	0,22	8
Мурманская область 2008	0,18	0,36	0,06	0,28	0,02	8
Новгородская область 2007	0,16	0,40	0,13	0,36	0,14	8

Новгородская область 2008	0,28	0,40	0,08	0,37	0,27	8
Омская область 2010	0,35	0,53	0,22	0,49	0,16	8
Республика Адыгея 2008	0,18	0,42	0,01	0,31	0,17	8
Сахалинская область 2009	0,32	0,39	0,01	0,25	0,00	8
Свердловская область 2008	0,27	0,49	0,24	0,42	0,24	8
Тамбовская область 2010	0,12	0,36	0,09	0,31	0,09	8
Удмуртская Республика 2008	0,25	0,49	0,14	0,34	0,14	8
Хабаровский край 2007	0,19	0,40	0,10	0,29	0,04	8
Хабаровский край 2010	0,31	0,44	0,08	0,37	0,09	8
Чувашская Республика 2007	0,11	0,48	0,18	0,42	0,17	8
Ярославская область 2010	0,12	0,42	0,20	0,48	0,36	8
Астраханская область 2007	0,08	0,52	0,15	0,16	0,02	9
Астраханская область 2008	0,13	0,52	0,10	0,27	0,01	9
Астраханская область 2009	0,09	0,51	0,18	0,16	0,04	9
Астраханская область 2010	0,07	0,53	0,18	0,24	0,08	9
Еврейская автономная область 2007	0,07	0,43	0,01	0,23	0,02	9
Еврейская автономная область 2009	0,10	0,46	0,01	0,24	0,11	9
Забайкальский край 2008	0,09	0,47	0,03	0,21	0,07	9
Иркутская область 2010	0,16	0,48	0,13	0,24	0,06	9
Камчатский край 2009	0,13	0,51	0,01	0,18	0,00	9
Камчатский край 2010	0,15	0,48	0,01	0,23	0,00	9
Курганская область 2007	0,06	0,48	0,16	0,35	0,21	9
Смоленская область 2010	0,13	0,46	0,08	0,26	0,12	9
Волгоградская область 2009	0,17	0,44	0,16	0,15	0,32	10
Воронежская область 2009	0,23	0,44	0,24	0,19	0,20	10
Кировская область 2009	0,13	0,42	0,09	0,11	0,16	10
Магаданская область 2007	0,12	0,43	0,12	0,08	0,17	10
Орловская область 2008	0,24	0,45	0,25	0,20	0,26	10
Орловская область 2010	0,20	0,50	0,24	0,19	0,23	10
Пензенская область 2008	0,14	0,42	0,19	0,18	0,25	10

Пермский край 2009	0,18	0,50	0,26	0,23	0,21	10
Пермский край 2010	0,17	0,48	0,23	0,24	0,34	10
Республика Башкортостан 2009	0,26	0,47	0,16	0,15	0,16	10
Республика Татарстан 2010	0,22	0,49	0,28	0,22	0,35	10
Ростовская область 2009	0,23	0,43	0,20	0,14	0,19	10
Ростовская область 2010	0,22	0,44	0,17	0,16	0,17	10
Ставропольский край 2007	0,13	0,37	0,12	0,15	0,17	10
Тверская область 2009	0,13	0,41	0,17	0,14	0,30	10
Тверская область 2010	0,13	0,38	0,16	0,14	0,25	10
Чувашская Республика 2009	0,14	0,49	0,22	0,25	0,33	10
Чувашская Республика 2010	0,13	0,47	0,24	0,25	0,28	10
Ярославская область 2008	0,12	0,44	0,24	0,24	0,30	10
Алтайский край 2009	0,21	0,38	0,10	0,07	0,11	10
Алтайский край 2008	0,22	0,40	0,10	0,15	0,11	11
Алтайский край 2010	0,22	0,37	0,11	0,12	0,08	11
Белгородская область 2009	0,33	0,40	0,16	0,21	0,17	11
Кабардино-Балкарская Республика 2008	0,26	0,39	0,05	0,13	0,03	11
Кемеровская область 2008	0,35	0,40	0,13	0,14	0,03	11
Новгородская область 2010	0,25	0,34	0,11	0,20	0,15	11
Омская область 2008	0,42	0,54	0,21	0,24	0,15	11
Омская область 2009	0,36	0,53	0,22	0,24	0,16	11
Псковская область 2010	0,16	0,35	0,09	0,12	0,06	11
Республика Башкортостан 2008	0,36	0,49	0,16	0,21	0,15	11
Республика Северная Осетия – Алания 2010	0,23	0,33	0,10	0,18	0,10	11
Ростовская область 2008	0,29	0,43	0,17	0,22	0,20	11
Тамбовская область 2009	0,16	0,37	0,08	0,17	0,17	11
Тюменская область 2010	0,22	0,36	0,10	0,12	0,03	11
Амурская область 2009	0,20	0,53	0,04	0,06	0,06	12
Камчатский край 2007	0,06	0,48	0,01	0,08	0,02	12

Камчатский край 2008	0,09	0,52	0,02	0,07	0,00	12
Курганская область 2010	0,08	0,46	0,10	0,13	0,10	12
Республика Алтай 2009	0,25	0,53	0,04	0,08	0,00	12
Республика Алтай 2010	0,23	0,60	0,04	0,13	0,08	12
Республика Бурятия 2007	0,14	0,58	0,05	0,16	0,16	12
Республика Бурятия 2009	0,15	0,61	0,07	0,21	0,12	12
Республика Бурятия 2010	0,16	0,67	0,08	0,12	0,00	12
Республика Коми 2007	0,11	0,48	0,07	0,15	0,19	12
Республика Коми 2008	0,15	0,51	0,04	0,08	0,11	12
Республика Коми 2009	0,14	0,49	0,03	0,07	0,01	12
Республика Коми 2010	0,14	0,50	0,04	0,08	0,08	12
Республика Мордовия 2008	0,09	0,53	0,15	0,23	0,22	12
Республика Хакасия 2008	0,14	0,51	0,01	0,07	0,00	12
Республика Хакасия 2009	0,18	0,50	0,02	0,06	0,00	12
Алтайский край 2007	0,16	0,40	0,11	0,12	0,07	13
Иркутская область 2008	0,23	0,46	0,13	0,14	0,12	13
Кировская область 2010	0,13	0,41	0,11	0,18	0,20	13
Республика Башкортостан 2010	0,27	0,49	0,16	0,19	0,19	13
Саратовская область 2009	0,16	0,49	0,21	0,26	0,21	13
Саратовская область 2010	0,15	0,50	0,19	0,21	0,16	13
Свердловская область 2009	0,20	0,47	0,26	0,29	0,26	13
Ярославская область 2009	0,12	0,42	0,23	0,29	0,26	13
Кировская область 2007	0,08	0,42	0,10	0,14	0,13	13
Республика Ингушетия 2007	0,16	0,30	0,00	0,00	0,00	14
Республика Ингушетия 2008	0,27	0,33	0,00	0,00	0,00	14
Республика Ингушетия 2009	0,26	0,39	0,00	0,00	0,00	14
Республика Ингушетия 2010	0,25	0,43	0,00	0,00	0,00	14
Республика Тыва 2007	0,34	0,65	0,00	0,00	0,00	14
Чеченская Республика 2009	0,13	0,37	0,01	0,00	0,00	14
Еврейская автономная область 2008	0,05	0,45	0,01	0,00	0,08	14

Чеченская Республика 2007	0,05	0,39	0,00	0,00	0,00	14
Амурская область 2007	0,16	0,55	0,04	0,02	0,04	15
Республика Алтай 2007	0,16	0,47	0,02	0,00	0,10	15
Республика Калмыкия 2007	0,00	0,48	0,00	0,00	0,00	15
Республика Калмыкия 2008	0,01	0,57	0,01	0,00	0,00	15
Республика Калмыкия 2009	0,05	0,52	0,03	0,00	0,00	15
Республика Калмыкия 2010	0,04	0,52	0,00	0,00	0,00	15
Республика Тыва 2008	0,51	0,66	0,01	0,00	0,00	15
Чеченская Республика 2008	0,14	0,43	0,00	0,00	0,00	15
Воронежская область 2007	0,16	0,43	0,33	0,20	0,33	16
Ивановская область 2010	0,10	0,38	0,34	0,35	0,52	16
Пензенская область 2009	0,11	0,41	0,23	0,07	0,17	16
Республика Татарстан 2008	0,34	0,46	0,27	0,19	0,41	16
Республика Татарстан 2009	0,25	0,46	0,29	0,15	0,46	16
Самарская область 2008	0,28	0,43	0,36	0,17	0,54	16
Самарская область 2009	0,19	0,43	0,34	0,15	0,50	16
Самарская область 2010	0,21	0,40	0,30	0,21	0,42	16
Ульяновская область 2007	0,16	0,48	0,38	0,16	0,31	16
Ульяновская область 2008	0,23	0,50	0,31	0,11	0,52	16
Ульяновская область 2009	0,18	0,48	0,35	0,10	0,28	16
Ульяновская область 2010	0,18	0,45	0,33	0,08	0,45	16
Белгородская область 2007	0,28	0,41	0,21	0,08	0,12	17
Владимирская область 2008	0,32	0,40	0,23	0,10	0,07	17
Владимирская область 2009	0,31	0,38	0,15	0,18	0,08	17
Кемеровская область 2007	0,22	0,39	0,15	0,15	0,04	17
Краснодарский край 2007	0,21	0,30	0,15	0,11	0,02	17
Краснодарский край 2008	0,29	0,31	0,13	0,08	0,07	17
Краснодарский край 2009	0,28	0,33	0,12	0,13	0,05	17
Краснодарский край 2010	0,27	0,33	0,14	0,12	0,05	17
Красноярский край 2007	0,37	0,47	0,23	0,17	0,04	17

Курская область 2008	0,27	0,46	0,21	0,18	0,08	17
Нижегородская область 2008	0,29	0,43	0,33	0,26	0,11	17
Новосибирская область 2009	0,23	0,47	0,32	0,23	0,16	17
Псковская область 2007	0,11	0,34	0,11	0,08	0,02	17
Псковская область 2008	0,15	0,36	0,13	0,06	0,03	17
Псковская область 2009	0,15	0,34	0,11	0,11	0,03	17
Республика Дагестан 2007	0,13	0,23	0,10	0,11	0,06	17
Республика Дагестан 2008	0,20	0,25	0,12	0,14	0,09	17
Республика Дагестан 2009	0,23	0,27	0,17	0,05	0,03	17
Республика Северная Осетия – Алания 2007	0,20	0,36	0,10	0,06	0,00	17
Республика Северная Осетия – Алания 2008	0,27	0,37	0,13	0,12	0,00	17
Республика Северная Осетия –Алания 2009	0,23	0,35	0,10	0,04	0,01	17
Тульская область 2007	0,20	0,35	0,23	0,10	0,04	17
Тюменская область 2007	0,29	0,35	0,16	0,10	0,01	17
Тюменская область 2009	0,23	0,36	0,13	0,14	0,01	17
Челябинская область 2008	0,26	0,49	0,31	0,23	0,18	17
Белгородская область 2008	0,43	0,42	0,16	0,14	0,24	18
Кабардино-Балкарская Республика 2010	0,26	0,31	0,08	0,14	0,28	18
Калининградская область 2007	0,28	0,34	0,15	0,12	0,26	18
Калининградская область 2009	0,30	0,35	0,11	0,10	0,19	18
Новгородская область 2009	0,27	0,35	0,07	0,18	0,22	18
Белгородская область 2010	0,38	0,38	0,17	0,22	0,11	19
Владимирская область 2010	0,30	0,37	0,14	0,19	0,05	19
Кабардино-Балкарская Республика 2009	0,28	0,33	0,07	0,26	0,11	19
Калининградская область 2010	0,29	0,36	0,09	0,14	0,00	19
Красноярский край 2010	0,43	0,48	0,15	0,23	0,09	19
Сахалинская область 2007	0,55	0,40	0,06	0,06	0,00	19
Сахалинская область 2008	0,53	0,40	0,05	0,15	0,00	19

Сахалинская область 2010	0,36	0,38	0,01	0,16	0,00	19
Тульская область 2008	0,33	0,36	0,17	0,26	0,12	19
Тульская область 2009	0,25	0,35	0,16	0,30	0,10	19
Тульская область 2010	0,22	0,32	0,18	0,28	0,07	19
Тюменская область 2008	0,33	0,37	0,12	0,22	0,04	19
Ханты-Мансийский автономный округ – Югра 2007	0,39	0,26	0,05	0,12	0,01	19
Ханты-Мансийский автономный округ – Югра 2008	0,46	0,27	0,03	0,25	0,02	19
Ханты-Мансийский автономный округ – Югра 2009	0,32	0,28	0,04	0,23	0,02	19
Ханты-Мансийский автономный округ – Югра 2010	0,28	0,27	0,04	0,17	0,03	19
Чукотский автономный округ 2007	0,35	0,16	0,00	0,00	0,50	20
Чукотский автономный округ 2008	0,52	0,17	0,00	0,00	0,41	20
Чукотский автономный округ 2009	0,60	0,16	0,00	0,00	0,20	20
Чукотский автономный округ 2010	0,41	0,11	0,00	0,00	0,24	20
Ненецкий автономный округ 2007	1,00	0,30	0,00	0,01	0,00	21
Ненецкий автономный округ 2008	0,71	0,33	0,00	0,00	0,00	21
Ненецкий автономный округ 2009	0,77	0,33	0,00	0,15	0,00	21
Ненецкий автономный округ 2010	0,74	0,31	0,00	0,16	0,00	21
Воронежская область 2008	0,23	0,45	0,25	0,23	0,52	22
Карачаево-Черкесская Республика 2010	0,18	0,35	0,04	0,15	0,50	22
Курганская область 2009	0,10	0,44	0,13	0,09	0,36	22
Липецкая область 2008	0,31	0,44	0,07	0,10	0,50	22
Магаданская область 2008	0,17	0,49	0,22	0,17	0,52	22
Республика Марий Эл 2007	0,10	0,50	0,20	0,05	0,34	22
Республика Марий Эл 2008	0,14	0,50	0,15	0,06	0,37	22
Республика Марий Эл 2009	0,16	0,47	0,18	0,07	0,48	22
Республика Марий Эл 2010	0,17	0,44	0,18	0,08	0,48	22
Республика Мордовия 2009	0,10	0,49	0,18	0,19	0,62	22

Республика Мордовия 2010	0,12	0,46	0,20	0,26	0,65	22
Ставропольский край 2008	0,18	0,37	0,10	0,13	0,41	22
Ставропольский край 2009	0,17	0,36	0,10	0,07	0,33	22
Тверская область 2008	0,14	0,43	0,18	0,09	0,36	22
Чеченская Республика 2010	0,11	0,37	0,01	0,00	0,43	22
Амурская область 2008	0,22	0,56	0,03	0,07	0,08	23
Республика Бурятия 2008	0,17	0,62	0,12	0,18	0,21	23
Республика Хакасия 2007	0,13	0,50	0,01	0,08	0,16	23
Республика Хакасия 2010	0,20	0,52	0,01	0,00	0,00	23
Ивановская область 2007	0,12	0,40	0,48	0,23	0,08	24
Ивановская область 2008	0,13	0,40	0,43	0,26	0,20	24
Калужская область 2007	0,16	0,34	0,42	0,26	0,18	24
Калужская область 2009	0,23	0,32	0,32	0,23	0,16	24
Калужская область 2010	0,22	0,35	0,30	0,28	0,08	24
Нижегородская область 2007	0,20	0,42	0,41	0,32	0,05	24
Республика Татарстан 2007	0,23	0,41	0,39	0,28	0,27	24
Рязанская область 2007	0,09	0,42	0,32	0,15	0,05	24
Томская область 2007	0,21	0,50	0,55	0,29	0,08	24
Томская область 2008	0,28	0,49	0,48	0,23	0,07	24
Томская область 2009	0,24	0,49	0,55	0,32	0,08	24
Томская область 2010	0,24	0,52	0,50	0,25	0,08	24
г. Санкт-Петербург 2008	0,52	0,51	0,53	0,33	0,11	25
Ямало-Ненецкий автономный округ 2008	0,27	0,14	0,12	0,38	0,12	25
Ивановская область 2009	0,12	0,39	0,41	0,48	0,18	25
Калужская область 2008	0,27	0,32	0,31	0,34	0,20	25
Ленинградская область 2007	0,16	0,11	0,08	0,15	0,01	25
Ленинградская область 2008	0,24	0,10	0,08	0,07	0,03	25
Ленинградская область 2009	0,25	0,11	0,11	0,08	0,04	25
Ленинградская область 2010	0,24	0,08	0,11	0,07	0,05	25
Московская область 2007	0,20	0,19	0,37	0,24	0,15	25

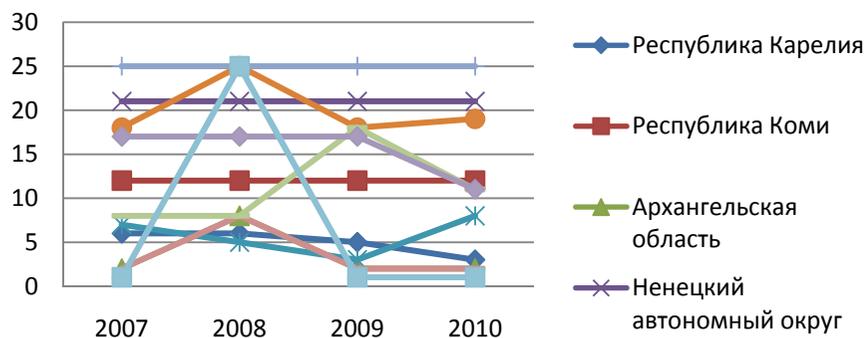
Московская область 2008	0,30	0,20	0,34	0,18	0,27	25
Московская область 2009	0,22	0,21	0,35	0,21	0,37	25
Московская область 2010	0,23	0,17	0,34	0,20	0,22	25
Республика Дагестан 2010	0,21	0,26	0,21	0,16	0,42	25
Республика Мордовия 2007	0,09	0,51	0,21	0,42	0,50	25
Самарская область 2007	0,20	0,42	0,44	0,23	0,55	25
Ямало-Ненецкий автономный округ 2007	0,24	0,14	0,13	0,17	0,04	25
Ямало-Ненецкий автономный округ 2009	0,18	0,17	0,07	0,08	0,01	25
Ямало-Ненецкий автономный округ 2010	0,18	0,15	0,07	0,11	0,03	25
Калининградская область 2008	0,39	0,31	0,09	0,11	0,22	25
Воронежская область 2010	0,22	0,41	0,27	0,25	0,25	25
Нижегородская область 2009	0,22	0,41	0,34	0,31	0,28	25
Нижегородская область 2010	0,22	0,39	0,26	0,33	0,33	25
Костромская область 2008	0,12	0,42	0,13	0,12	0,10	25
Орловская область 2009	0,20	0,45	0,21	0,16	0,12	25

Приложение 3. Динамическая группа абсолютно устойчивых регионов

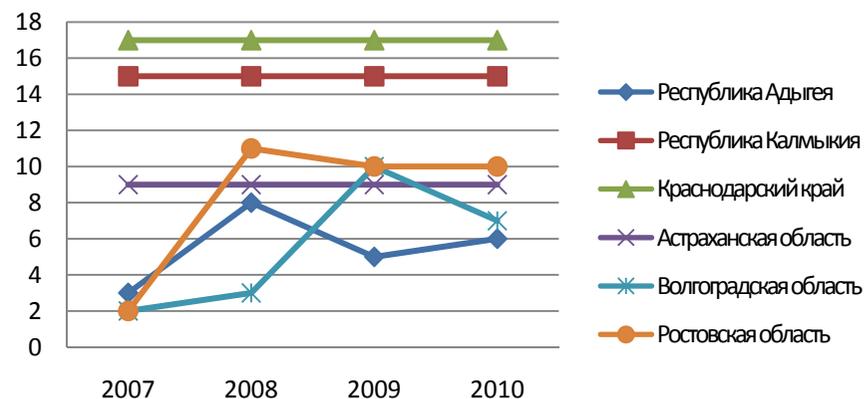
Астраханская область	9	9	9	9
г. Москва	1	1	1	1
Краснодарский край	17	17	17	17
Ленинградская область	25	25	25	25
Московская область	25	25	25	25
Ненецкий автономный округ	21	21	21	21
Оренбургская область	5	5	5	5
Приморский край	1	1	1	1
Республика Ингушетия	14	14	14	14
Республика Калмыкия	15	15	15	15
Республика Коми	12	12	12	12
Республика Марий Эл	22	22	22	22
Республика Саха (Якутия)	5	5	5	5
Томская область	24	24	24	24
Ульяновская область	16	16	16	16
Ханты-Мансийский автономный округ - Югра	19	19	19	19
Чукотский автономный округ	20	20	20	20
Ямало-Ненецкий автономный округ	25	25	25	25

Приложение 4. Траектории регионов в динамическом анализе паттернов

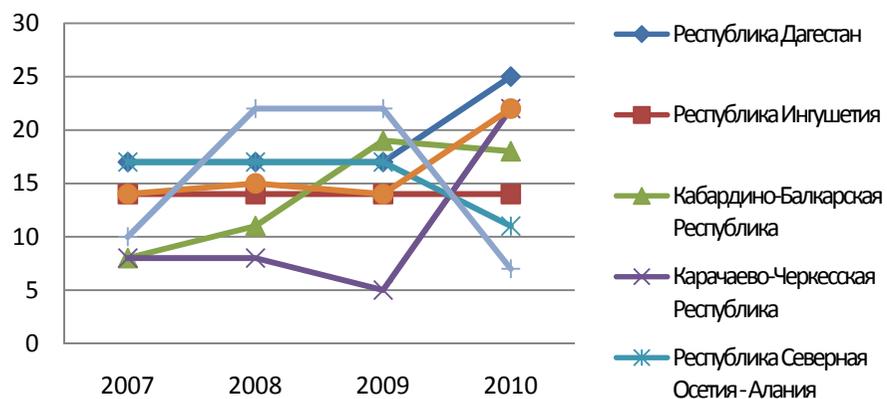
Северо-Западный федеральный округ



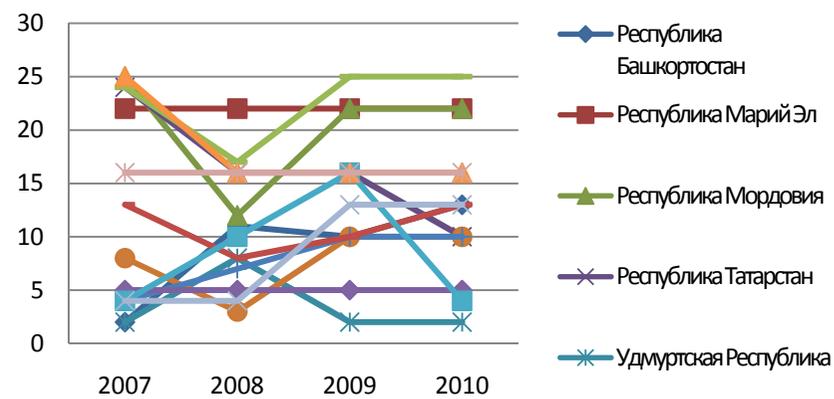
Южный федеральный округ



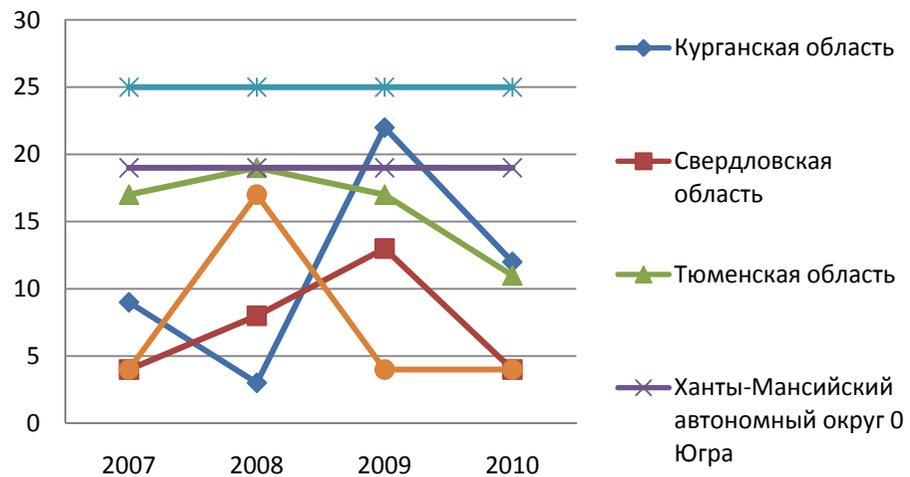
Северо-Кавказский федеральный округ



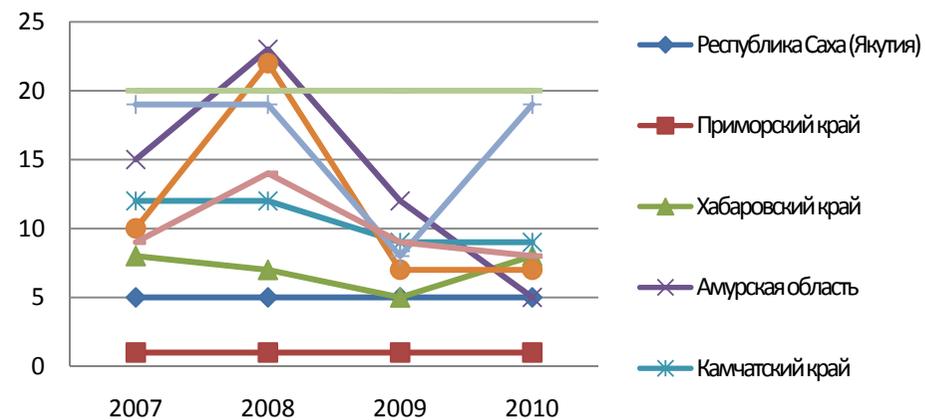
Приволжский федеральный округ



Уральский федеральный округ



Дальневосточный федеральный округ



Study of science, education and innovation data using the pattern analysis : Working paper WP7/2012/07 [Text] / F. Aleskerov, L. Gokhberg, L. Egorova, A. Myachin, G. Sagieva ; National Research University “Higher School of Economics”. – Moscow : Publishing House of the Higher School of Economics, 2012.

Analysis of the science, education and innovation data is very difficult and complex problem. In this paper we describe the method of pattern analysis and the results of its application to the problem of analyzing the development of science, education and the success of innovative activity in the regions of the Russian Federation.

In this study we examined characteristics of the regions of Russia such as the level of socio-economic conditions and the potential and efficiency of science, education and innovation from 2007 to 2010. Also we obtain a classification of regions by the similarity of the internal structure of these indicators, constructed trajectories of regional development over time, and found groups of regions carrying out similar strategies.

Keywords: pattern, cluster analysis, the regions of the Russian Federation, science, education, innovation.

Aleskerov F. – National Research University “Higher School of Economics”.

Gokhberg L. – National Research University “Higher School of Economics”.

Egorova L. – National Research University “Higher School of Economics”.

Myachin A. – National Research University “Higher School of Economics”.

Sagieva G. – National Research University “Higher School of Economics”.

Препринт WP7/2012/07

Серия WP7

Математические методы анализа решений
в экономике, бизнесе и политике

Алескеров Ф.Т., Гохберг Л.М., Егорова Л.Г.,
Мячин А.Л., Сагиева Г.С.

**Анализ данных науки, образования и инновационной
деятельности с использованием методов анализа паттернов**