

УДК 332.1

К.С. Мухтарова*, А.Т. Мылтыкбаева

Казахский национальный университет имени аль-Фараби, Республика Казахстан, г. Алматы

*E-mail: kmukhtarova@inbox.ru

**Анализ данных науки, образования и инновационной деятельности
с использованием методов анализа паттернов по регионам
Республики Казахстан за 2003-2012 гг.**

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

Ключевые слова: инновация, анализ паттернов, региональное инновационное развитие, корреляционный анализ.

K.S. Mukhtarova, A.T. Myltykbayeva

**The analysis of data of science, education and innovative activity with use of methods
of the analysis of patterns on regions of the Republic of Kazakhstan for 2003-2012**

The problem of the analysis of data of science, education and innovative activity is a complex and complex challenge, and to conduct the researches connected with these spheres of human activity, it is possible in the various ways, with use of different techniques and different methodical approaches.

Keywords: innovation, analysis of patterns, regional innovative development, correlation analysis.

Қ.С. Мухтарова, А.Т. Мылтыкбаева

**2003-2012 жылдарға арналған Қазақстан Республикасы аймақтары бойынша
паттерндерді талдау әдістерін қолдану аясында ғылым, білім мәліметтерін және
инновациялық қызметті талдау**

Ғылым, білім мәліметтерін және инновациялық қызметті талдау мәселесі күрделі, кешенді міндет болып табылады, адам қызметінің бұл саласына қатысты зерттеулерді жүргізу әртүрлі тәсілдер мен түрлі әдістемелерді қолдану арқылы жүзеге асырылады. Бұл мақалада паттерндерді талдау әдістері, олардың нәтижелерінің қойылған мақсаттарға және Қазақстан Республикасының аймақтарындағы ғылым мен білім дамуындағы қолданылуы сипатталады.

Түйін сөздер: инновация, паттерндерді талдау, аймақтық экономикалық даму, корреляциялық талдау.

Анализ паттернов – это новая область анализа данных, связанная с поиском взаимосвязей исследуемых объектов, построение 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 объек-

тов, которые выделяют объекты этого паттерна среди всей остальной совокупности. 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 [1].

С учетом того, что для динамического анализа паттернов необходимо иметь данные по всем показателям за выбранный промежуток времени, решено использовать данные лишь по 10 годам: 2003–2012 гг. Далее все показатели z_i^x (индекс x относится к объектам-регионам) были преобразованы аналогично [2] и получены нормированные показатели по всем блокам согласно формуле:

$$\tilde{z}_i^x = \frac{z_i^x - \min(z_i^x)}{\max(z_i^x) - \min(z_i^x)}. \quad (1)$$

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

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 [1].

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 взаи-

мосвязи 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 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 [1].

Как оказалось, все блоки можно считать некоррелированными (табл. 1).

Можно сказать, что на основании выбранной базовой системы показателей получена комплексная оценка развития и результативности науки, образования и инновационной деятельности в каждом регионе Казахстана, причем в динамике за 10 лет. Каждому из исследуемых объектов-регионов поставлен в соответствие вектор признаков описаний, составленный из 6 агрегированных показателей (Блок 1, Блок 2, Блок 3, Блок 4, Блок 5 и Блок 6). Каждый из этих показателей представляет собой число от 0 до 1, все показатели являются некоррелированными.

Таблица 1 – Корреляция между блоками показателей для 2003 г.

2003	Блок 1	Блок 2	Блок 3	Блок 4	Блок 5	Блок 6
Блок 1. Социально-экономические условия	1					
Блок 2. Образовательный потенциал	0,14	1,00				
Блок 3. Потенциал научно-технической деятельности	0,68	0,11	1			
Блок 4. Результативность исследований и разработок	0,01	0,32	0,33	1		
Блок 5. Потенциал инновационной деятельности	-0,24	0,09	-0,06	0,49	1,00	
Блок 6. Результативность инновационной деятельности	0,04	-0,09	0,13	-0,10	-0,17	1

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. 1 и рис. 2. Стоит отметить, что здесь и на всех последующих рисунках справа могут быть отмечены не все, а лишь часть регионов, представленных на рисунке.

Кажущееся хаотичным нагромождением изломанных линий при ближайшем рассмотрении проявляет некоторые закономерности и видно, что многие регионы имеют если неравные, 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.

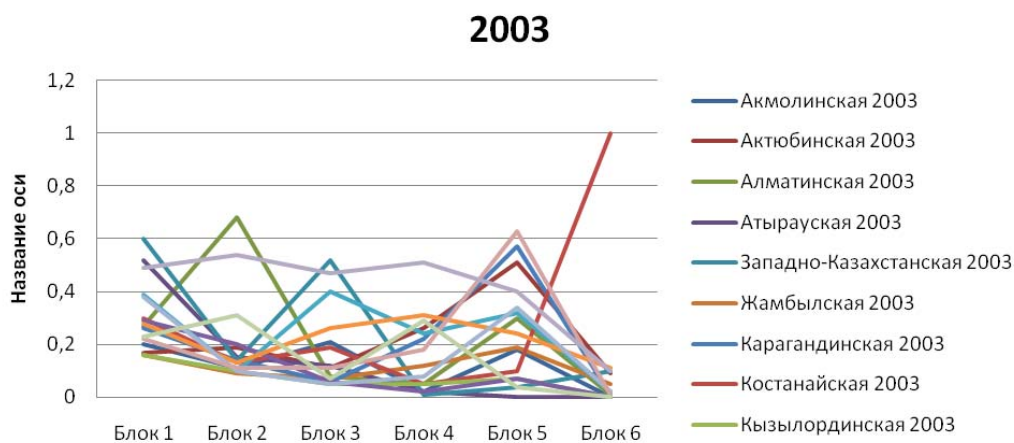


Рисунок 1 – Кусочно-линейные функции, описывающие объекты, для 2003 г.

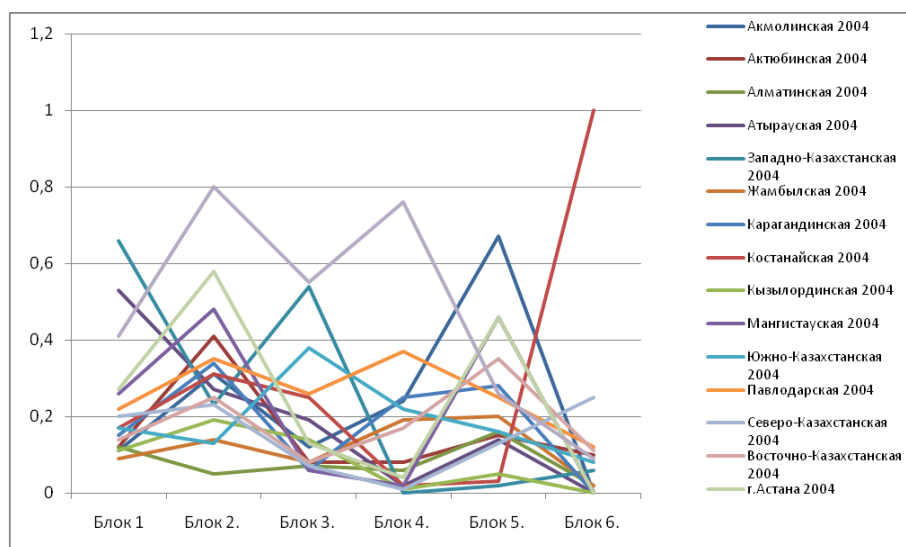


Рисунок 2 – Кусочно-линейные функции, описывающие объекты, для 2004 г.

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 $16 \times 10 = 160$ объектов для кластеризации.

Мы использовали два метода кластеризации: k-средних и иерархический метод кластеризации, в каждом из которых использовалась евклидова метрика, и проводили многошаговую процедуру кластерного анализа, комбинируя два вышеупомянутых метода. В результате было получено 33 паттерна данных, включающих более двух объектов внутри паттерна.

Динамический анализ паттернов позволяет отследить, какому паттерну из вышеописанных

следовал каждый из регионов в исследуемые 10 лет, 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. 2.

Таблица 2 – Динамический анализ паттернов

Регионы	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Акмолинская	33	6	16	16	16	16	18	16	16	16
Актюбинская	6	18	18	2	2	33	2	20	20	26
Алматинская	18	11	11	11	11	33	11	11	19	19
Атырауская	24	8	8	30	8	8	4	8	8	8
Западно-Казахстанская	22	22	1	1	1	1	1	1	33	21
Жамбылская	28	33	15	15	15	7	7	15	29	29
Карагандинская	6	2	23	23	5	4	27	4	8	30
Костанайская	7	7	7	7	7	7	33	7	7	21
Кызылординская	33	33	33	32	32	5	31	24	24	31
Мангистауская	28	18	18	4	4	30	3	3	3	33
Южно-Казахстанская	10	33	10	10	10	33	10	10	10	33
Павлодарская	33	2	9	33	26	33	14	14	14	14
Северо-Казахстанская	28	17	24	5	24	33	17	17	24	17
Восточно-Казахстанская	6	28	5	27	5	6	6	5	33	33
г.Астана	2	18	33	9	9	9	9	25	25	33
г.Алматы	33	12	13	13	13	33	12	12	12	12

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 иссле-

довательской, 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 Казахстана по таким показателям как уровень социально-экономических условий, потенциал и результативность науки, образования и инновационной деятельности в динамике за 10 лет с 2003 по 2012 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]:

1) Выявления неявно выраженного взаимно-

го 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;

– 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.

Литература

1. 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.

2. Рейтинг инновационного развития субъектов Российской Федерации: аналитический доклад / под ред. Л.М. Гохберга. М.: Национальный исследовательский университет «Высшая школа экономики», 2012.

References

1. Analiz dannyh nauki, obrazovanija i innovacionnoj dejatel'nosti s ispol'zovaniem metodov analiza patternov : preprint WP7/2012/07 [Tekst] / F.T. Aleskerov, L.M. Gohberg, L.G. Egorova, A.L. Mjachin, G.S. Sagieva ; Nac. issled. un-t «Vysshaja shkola jekonomiki». – M. : Izd. dom Vysshej shkoly jekonomiki, 2012.

2. Rejting innovacionnogo razvitija sub#ektov Rossijskoj Federacii: analiticheskij doklad / pod red. L.M. Gohberga. M.: Nacional'nyj issledovatel'skij universitet «Vysshaja shkola jekonomiki», 2012.