

Кластеризация временных рядов для исследования динамики научно-технического прогресса

Р.Д. Зайцев

Московский физико-технический институт (государственный университет)

Попытки изучения динамики мирового научно-технического развития, а также влияния научного прогресса на мировую экономику предпринимаются на протяжении многих лет, и интерес к таким исследованиям непрерывно возрастает [1]. Как отмечено в статье [2], общепринятые подходы содержат слишком много предварительных экстраполяционных гипотез, которые усложняют исследование глобальных процессов, таких как научно-техническое развитие.

Для содержательного изучения динамики научно-технического развития различных стран необходима как можно более точная и интерпретируемая кластеризация. Такой подход позволит упростить последующие исследования, такие как прогнозирование динамики развития каждой страны в зависимости от группы, в которую она входит, или поиск новых неявных закономерностей, являющихся характерными для стран внутри каждого кластера.

В данном исследовании предложен подход к решению проблемы кластеризации стран по типу их научно-технического развития на основе непараметрических многомерных мер различия между временными рядами, основанных на алгоритме динамического выравнивания DTW [3], [4]. В частности, предложено обобщение на многомерный случай адаптивной меры CORT [5], основанной на уточнении расстояния DTW с помощью коэффициента временной корреляции. В работе [6] показано, что для выборок относительно небольшого объема (десятки и сотни записей) данный подход является более предпочтительным по причине высокой точности кластеризации. При его применении к выборкам большого объема на первый план выходит проблема относительно высокой вычислительной сложности алгоритма $\sim O(n^2)$.

Для кластеризации 56 стран по типу динамики их научно-технического развития с 1992 по 2013 годы были выбраны следующие показатели: количество патентных заявок по месту (стране) обращения; количество патентных заявок по гражданству заявителя; количество исследователей и инженеров в научной области (на 10^6 населения); высокотехнологичный экспорт в \$USA по текущему курсу; количество научных публикаций (источники: база данных Всемирного Банка, data.worldbank.org; база данных Всемирной организации интеллектуальной собственности, ipstats.wipo.int).

С помощью описанной методики удалось произвести разбиение исследуемой выборки на четыре устойчивых интерпретируемых кластера.

Первый кластер, куда вошли 26 стран, характеризуется устойчивым ростом по четырём или пяти показателям. По одному показателю допускается стабильный тренд. В первый кластер вошли не только развитые страны (Китай, США, Франция, Германия), но также страны с активно развивающимся научно-техническим комплексом (Таиланд, Бразилия).

Второй кластер, состоящий из 16 стран, характеризуется особым поведением патентной динамики – снижение патентов, поданных в данной стране с одновременным увеличением патентной активности её граждан (Рисунок 1). Это свидетельствует о том, что граждане данных стран предпочитают подавать патентные заявки за рубежом (Эстония, Латвия, Литва, Польша, Болгария, Чехия).

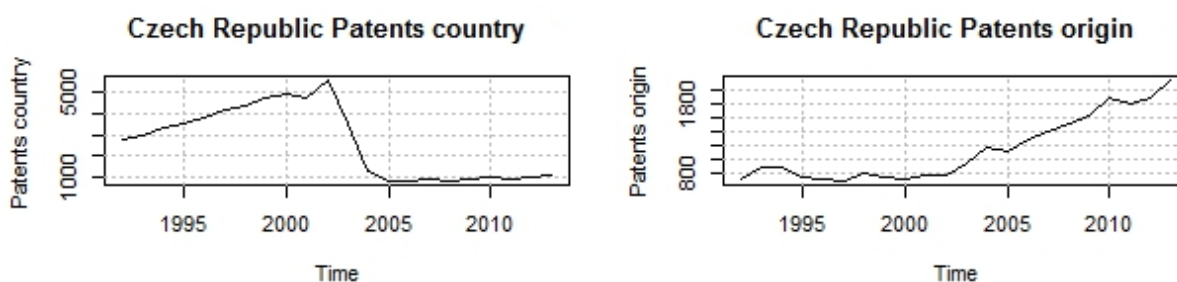


Рисунок 1. Патентные заявки в Чехии

Третий кластер характеризует смена устойчивого роста на первом этапе падением или стагнацией. Он состоит из 4 стран: Япония, Великобритания, Швеция, Финляндия. Такая динамика развития говорит о стагнации на конечном этапе развития (Рисунок 2). Возможно меры, предпринимаемые для развития науки, теряют свою эффективность с течением времени.

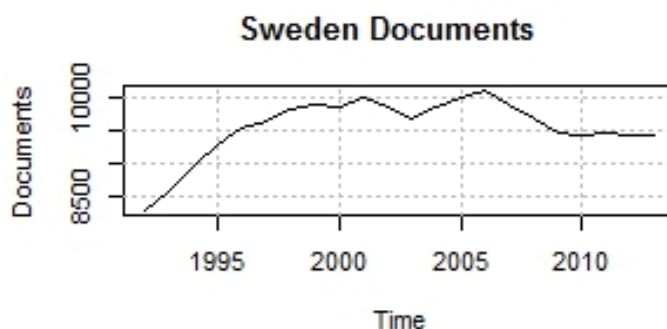


Рисунок 2. Публикации в Швеции

Четвёртый кластер наиболее неоднородный, в него вошли 10 стран, в том числе Россия и Украина. В данном случае нельзя говорить о росте по большинству показателей – три или четыре показателя демонстрируют стагнацию или падение.

Если установить отношения доминирования, основанное на наибольшей предпочтительности устойчивого роста по всем показателям, то четвёртый кластер следует считать наихудшим. Динамика показателей России очень схожа с динамикой Украины, например, количество исследователей и учёных падало почти линейно на протяжении последних двадцати лет (Рисунок 3). Устойчивый рост наблюдается только в области высокотехнологического экспорта. Это может свидетельствовать о недостатке эффективности мер развития научно-технического комплекса в нашей стране.

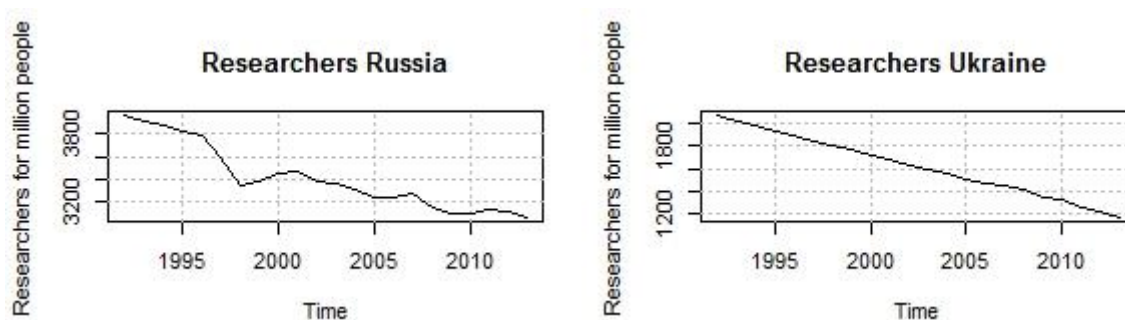


Рисунок 3. Исследователи в России и на Украине.

Литература

1. *Salter A. J., Martin B. R.* The economic benefits of publicly funded basic research: a critical review - *Research policy*. - 2001. – Т. 30. – №. 3. – С. 509-532.
2. *Геловани В.А, Бритков В.Б.* Методология глобального моделирования в анализе социально-экономических проблем // Системные исследования. Методологические проблемы: Ежегодник 2013-2014. Вып. 37. М.: ЛЕНАНД. - 2014. - С. 49-63.
3. *Berndt D. J., Clifford J.* Using Dynamic Time Warping to Find Patterns in Time Series - *KDD workshop*. – 1994. – Т. 10. – №. 16. – С. 359-370.
4. *Holt G., Reinders M.*, Multi-dimensional dynamic time warping for gesture recognition, in *Annual Conference on the Advanced School for Computing and Imaging*. - 2007.
5. *Chouakria D.A., Nagabhushan P.N.* Adaptive Dissimilarity Index for Measuring Time Series Proximity. - *Advances in Data Analysis and Classification*. – 2007 - 1(1), 5–21.
6. *Зайцев Р.Д.* Применение непараметрических мер различия между временными рядами для кластеризации показателей научно-технического развития» // Вопросы

науки: Современные технологии и технический прогресс: сборник статей по материалам IV международной научно-практической конференции. – 2015. С. 13-19.