

Ю.Ю. Петрунин

ИССЛЕДОВАНИЕ ЭЛЕКТОРАЛЬНОГО ПОВЕДЕНИЯ: САМООРГАНИЗУЮЩИЕСЯ КАРТЫ КОХОНЕНА VERSUS СТАТИСТИЧЕСКОГО АНАЛИЗА

В статье проводится кластерный анализ электорального поведения регионов по шкале «левые—правые». Используются как классические (статистические) методы, так и методы на основе самоорганизующихся карт Кохонена. Исследование проводится на основе результатов выборов 1990 г. в Греции и 1998 и 2004 гг. во Франции.

Ключевые слова. Электоральное поведение, кластерный анализ, самоорганизующиеся карты Кохонена.

The paper deals with cluster analysis of regional electoral behavior using the “right-left” scale. The study is based on classical statistic methods and self-organizing maps. Data source: official results of elections in Greece (1990, 1993, 1996) and France (1998, 2004).

Key words. Electoral Behavior, Cluster Analysis, Self-Organizing Maps

Введение

Еще каких-нибудь 15 лет назад в литературе часто встречалось выражение «красный пояс вокруг Москвы», которое отражало особенности электорального поведения ряда областей, придерживавшихся преимущественно коммунистической ориентации. В настоящее время можно выделить группу регионов, схожих между собой по отношению к расколу «партия власти — оппозиция». Моделирование электорального поведения как научное направление на стыке математики, политологии, экономики, государственного управления сформировалось еще в середине прошлого века¹. Обзор эволюции основных методов исследований дан в недавней работе А.С. Ахременко². В большинстве случаев речь идет о регрессионном и факторном анализе. Но если выделять группы объектов со сходным

Петрунин Юрий Юрьевич – доктор философских наук, профессор кафедры теории и технологий управления факультета государственного управления МГУ им. М.В. Ломоносова, e-mail: petrunin@sps.msu.ru.

¹ *Hotelling H.* Stability in Competition // *Economic Journal*. 1929. N 39; *Arrow K.* Social Choice and Individual Values. N.Y., 1951; *Downs A.* An Economic Theory of Democracy. N.Y., 1957; *Black D.* The Theory of Committees and Elections. Cambridge, 1958.

² См.: *Ахременко А.С.* Структуры электорального пространства. М., 2007.

электоральным поведением, то разумным представляется использовать также и кластерный анализ. В кластеры объединяются регионы с близкими характеристиками. При этом, однако, возникает ряд принципиальных вопросов, требующих теоретического осмысления и практической проверки. Какие переменные использовать для определения сходства различных регионов? Какие методы наиболее адекватны и точны?

В данной работе рассматривается установление схожести/различия регионов на основе лишь одной агрегированной переменной, по предложенному алгоритму «расщепляемой» на три. При этом используются как классические (статистические) методы вычисления близости/удаленности электорального поведения регионов, так и более гибкие и мощные методы на основе нейронных сетей. Во избежание излишней ангажированности фактический материал взят по результатам выборов 1990-х гг. в Греции и 1998, 2004 гг. во Франции.

Методология исследования

Для описания такого сложного феномена, как электоральное поведение, исследователи обычно стремятся ввести несколько переменных. Использование одной переменной, представляется, существенно обедняет как глубину полученной модели, так и ее предсказательную силу. Однако это не всегда так.

Возьмем в качестве примера такое хорошо известное измерение, как «правые — левые». При активно протекающей политической жизни вряд ли можно судить о политической «окраске» региона только по результатам одного политического события, например региональных или федеральных выборов. Очевидно, более точно можно охарактеризовать регион по результатам нескольких выборов. Но в таком случае использование некоторой усредненной меры предпочтений населения для определенного отрезка времени вызывает ряд вопросов. Такие же вопросы возникают тогда, когда в более крупное территориальное образование включены несколько более мелких с существенно разнящимся электоральным поведением. Речь идет о том, что запрашивающееся в данном примере среднее арифметическое может скрывать картину происходящего. Действительно, два региона могут обладать одним и тем же индексом политических пристрастий (то есть разницей проголосовавших за правые партии и за левые партии), но значительно отличающимся разбросом избирателей, проголосовавших в разные годы (или в разных более мелких территориальных образованиях, входящих в них) за правых и за левых. Объединять два региона в одну группу (кластер) в таком случае было бы ошибочным.

Меньший разброс индекса политических пристрастий показывает большее единодушие избирателей, больший — неопределенность пристрастий избирателей.

Другим показателем, который необходимо принимать во внимание при объединении регионов в схожие группы (кластеры), является наличие или отсутствие выбросов по результатам голосований в разные годы (или в разных более мелких регионах, входящих в более крупный). Выбросы существенно искажают среднее арифметическое для всего региона. С другой стороны, они свидетельствуют либо о маргинальных регионах, либо о возможном направлении электоральной динамики (или ее инерционности в некоторых частях). В связи с изложенными соображениями возникает идея учитывать при объединении регионов не только среднее, типичное (среднеарифметическое) значение, но и стандартное (среднеквадратичное) отклонение и асимметрию по той или иной переменной. Разумеется, речь идет об агрегированных переменных.

Исходные данные

В качестве «экспериментального поля» использовались данные о результатах региональных выборов в 1990-х — начале 2000-х гг. во Франции³ и парламентских в Греции⁴. При этом для Греции были взяты все регионы (генеральная совокупность), для Франции 11 (43 департамента) из 22 (96 департаментов)⁵

Выбранные регионы равномерно распределены по всей территории Французской Республики. Проверка по социально-экономическим показателям также показала, что выбранные регионы достаточно типичны.

Поскольку каждый регион состоит из нескольких департаментов и выборы проводились на выделенном отрезке дважды, для переменной «индекс политических пристрастий» были вычислены показатели среднего арифметического, стандартного отклонения и асимметрии (табл. 1).

Аналогичные вычисления были проделаны по результатам парламентских выборов в Греции в 1990, 1993 и 1996 гг. (табл. 2).

³ См.: *Петрунин Ю.Ю., Зернова Ю.А.* Статистические и нейросетевые методы исследования политической ситуации во Франции на примере региональных выборов 1998 и 2004 гг. // Государственное управление Электронный вестник. http://e-journal.spa.msu.ru/14_2008.htm

⁴ См.: *Петрунин Ю.Ю.* Классификация при ограниченной информации: микроанализ данных // Вестн. Моск. ун-та. Сер. 21. Управление (государство и общество). 2006. № 2. [http://www.spa.msu.ru/publ_\(vestnik_2_2006\).html](http://www.spa.msu.ru/publ_(vestnik_2_2006).html)

⁵ Речь идет только о континентальных регионах.

Таблица 1

Средний индекс политических пристрастий (Average), стандартное отклонение (St dev) и асимметрия (Skew) для разных регионов Франции по результатам региональных выборов 1998 и 2004 гг.

Регион	St dev	Average	Skew
Нижняя Нормандия (Basse-Normandie)	7,572188	10,99333	-0,70915
Рона-Альпы (Rhône-Alpes)	6,772582	10,658	-1,29223
Иль-де-Франс (Ile De France)	8,57934	8,0975	-0,21053
Эльзас (Alsace)	6,095843	31,155	0,083423
Аквитания (Aquitaine)	10,53566	-7,546	0,35864
Бретань (Bretagne)	12,48004	-6,65125	-0,24352
Земли Луары (Pays de la loire)	16,0035	10,008	0,111085
Лимузен (Limousin)	18,1257	-8,47667	0,074256
Лангедок (Languedoc)	16,37504	8,242	0,776873
Франш-Конте (Franche-Comte)	3,65657	10,02667	0,469937
Лотарингия (Lorraine)	9,868368	11,1325	-0,0285

Таблица 2

Средний индекс политических пристрастий (Average), стандартное отклонение (St dev) и асимметрия (Skew) для разных регионов Греции по результатам парламентских выборов 1990, 1993 и 1996 гг.

Регион	St dev	Average	Skew
Вост. Македония-Фракия	0,1973	9,03099	-1,87945
Центр. Македония	-2,8196	7,14719	-0,35709
Зап. Македония	0,2083	13,40266	0,23562
Эпир	-8,424	6,62789	0,21923
Фессалия	-10,1213	4,2431	0,32307
Центр. Греция	-5,338	9,44038	-0,17794
Ионические о-ва	-16,6583	7,62677	0,51357
Зап. Греция	-13,4978	7,51658	0,13502
Аттика	-12,218	12,38909	-0,1282
Пелопоннес	5,2727	9,69867	0,30916
Сев. Эгейский бассейн	-17,7278	8,41093	0,37118
Юж. Эгейский бассейн	-11,095	10,1768	0,6648
Крит	-26,4658	8,58109	0,34264

Кластерный анализ для Франции

Стандартизовав данные, проведем кластерный анализ⁶. Дендрограмма на рис. 1 подсказывает выделить 4 кластера.

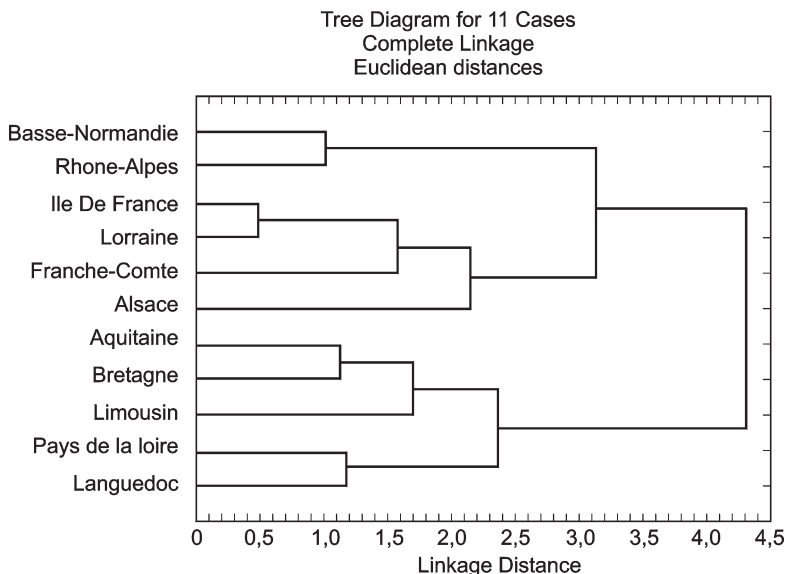


Рис. 1. Иерархическое объединение регионов Франции по результатам региональных выборов в 1998 и 2004 гг.

Состав кластеров и интерпретация

В 1 кластер вошли Иль-де-Франс (Ile De France), Франш-Конте (Franche-Comte), Лотарингия (Lorraine). Низкий разброс голосов, среднее арифметическое, средняя асимметрия. «Болото», слегка сдвинутое влево.

Во 2 кластере оказались Аквитания (Aquitaine), Бретань (Bretagne), Земли Луары (Pays de la loire), Лимузен (Limousin), Лангедок (Languedoc). Высокий разброс, низкое среднее, средняя асимметрия. Левый регион с широким разбросом голосов.

3 кластер объединил Нижнюю Нормандию (Basse-Normandie) и Рону-Альпы (Rhone-Alpes). Похож на 1 кластер. Отличие — низкая асимметрия, говорящая о неадекватности среднего. Вероятно, в некоторых департаментах региона преобладают ультралевые настроения, которые и сдвигают влево регион в целом.

⁶ Использовалась программа Statistica 6.0.

4 кластер — Эльзас (Alsace). Регион отличается низким разбросом индекса, высоким средним значением, средней асимметрией. Иными словами, кластер характеризуется правыми настройками электората и значительным единодушием.

Естественно, понятие «сдвиг вправо» или «сдвиг влево» имеет отношение к среднему по стране. Соответственно, говоря о «левом сдвиге», мы имеем в виду не то, что в нем голосует за левых больше, чем за правых, а то, что в нем голосует больше за левых, чем в среднем по стране.

График средних для 4 кластеров показан на рис. 2.

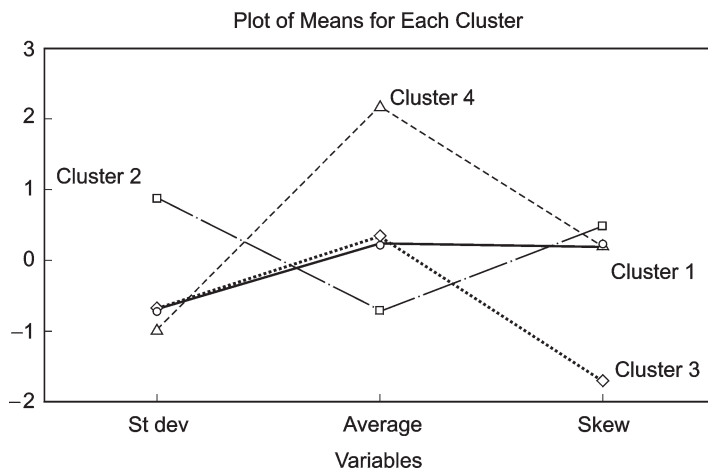


Рис. 2. Средние значения каждого кластера по 3 переменным.
Данные стандартизированы

Ограниченность статистических методов. Нейронные сети без учителя

Недостатком статистических методов является, в частности, то, что они предполагают наличие некоторых априорных представлений об изучаемой реальности. Например, при проведении кластерного анализа мы считаем, что переменные, описывающие объекты, взаимонезависимы. Но это совсем не так. При увеличении показателя асимметрии среднее арифметическое изменяется и, самое главное, становится неинформативным, искажающим типическое значение переменной.

Выходом из создавшегося положения может служить использование моделей нейронных сетей, которые позволяют изучать данные, не укладывающиеся в прокрустово ложе статистики. Перспективным

представляется использование разновидностей нейронных сетей без учителя — так называемых «самоорганизующихся карт Кохонена» (по имени их создателя), которые могут решать задачи кластерного анализа⁷.

На основе приведенных выше данных были получены группы схожих регионов (кластеры), показанные на рис. 3⁸.

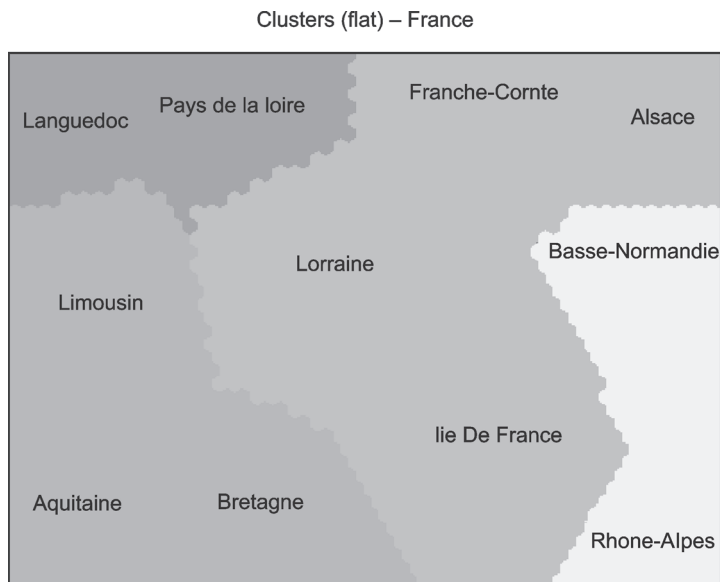


Рис. 3. Выделение 4 кластеров для Франции по результатам региональных выборов 1998 и 2004 гг.

График на рис.4 показывает, что кластеры, выделенные нейронной сетью, не совпадают с кластерами, выделенными традиционными статистическими методами.

1 кластер: Иль-де-Франс (Ile De France), Франш-Конте (Franche-Comte), Лотарингия (Lorraine), Эльзас (Alsace). Правые: разброс не очень большой, асимметрия мала.

2 кластер состоит из Аквитании (Aquitaine), Бретани (Bretagne) и Лимузена (Limousin). Левые: Разброс большой, асимметрия мала.

3 кластер — Нижняя Нормандия (Basse-Normandie), Рона-Альпы (Rhone-Alpes) — совпадает с 3 кластером предшествующего анализа. Правые в абсолютном понимании, но левее, чем в среднем по стране. Разброс небольшой, но сильная асимметрия говорит о ряде непред-

⁷ Kohonen N. Self-Organizing Maps. N.Y., 1995; Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт. М., 2001.

⁸ Использовалась программа Viscosity SOMine 3.0.

сказуемых результатов и, кроме того, что регион на самом деле правее среднего арифметического.

4 кластер включает в себя Лангедок (Languedoc) и Земли Луары (Pays de la Loire). Чуть правее центра с очень большим разбросом и возрастающей асимметрией. Весьма непредсказуемый регион.

Внимательный анализ полученных кластеров показывает, что, во-первых, самообучающиеся нейронные сети адекватней отражают электоральные особенности регионов и, во-вторых, легче интерпретируются, поскольку напрямую говорят о левых или правых регионах, в отличие от статметодов, где выражение «левее» или «правее» применяется всегда относительно средних значений электоральной переменной⁹.

Кластерный анализ для Греции

Перейдем к Греции. Будем придерживаться того же алгоритма. Дендрограмма на рис. 4 также подсказывает выделение 4 кластеров.

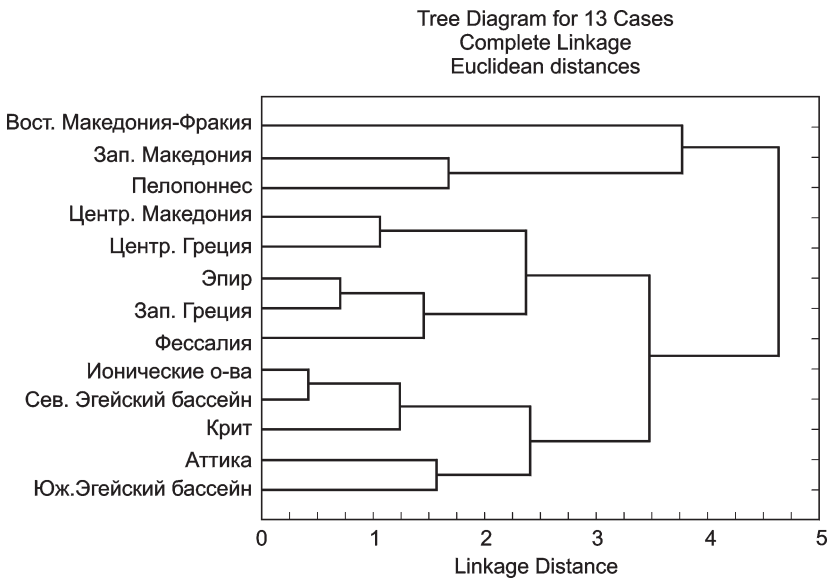


Рис. 4. Иерархическое объединение регионов Греции по результатам парламентских выборов 1990-х гг.

⁹ Так как данные были стандартизированы. В противном случае переменные не были бы сопоставимы.

Состав кластеров и интерпретация

1 кластер: Зап. Македония, Центр. Греция, Аттика, Пелопоннес. Правый, относительно страны в целом, регион с отсутствием единости и допустимой асимметрией.

2 кластер: Ионические о-ва, Зап. Греция, Сев. Эгейский бассейн, Юж. Эгейский бассейн, Крит. Левый регион с небольшим разбросом и положительной асимметрией. Следовательно, на самом деле еще более левый.

3 кластер: Центр. Македония, Эпир, Фессалия. Центристский регион с редким единым и не искажающей асимметрией.

4 кластер: Вост. Македония—Фракия. Правый регион с нормальным разбросом и очень большой отрицательной асимметрией, говорящей о том, что на самом деле регион еще более правый.

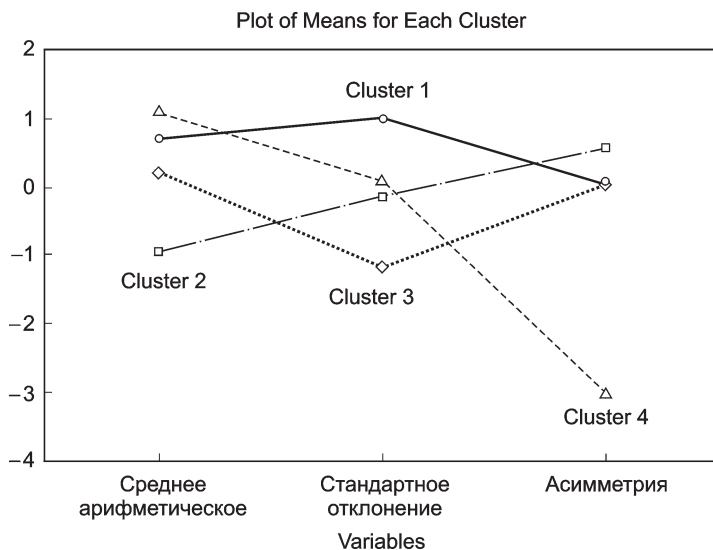


Рис. 5. График средних для 4 кластеров по 3 переменным.
Данные стандартизированы

Нейросети без учителя для Греции

На рис. 6 показан результат разделения всех регионов Греции по схожести электорального поведения на 4 кластера методом нейронных сетей без учителя. Состав кластеров также несколько отличается от кластеров, полученных методом *k*-средних.

1 кластер: Центр. Македония, Эпир, Фессалия, Зап. Греция, Центр. Греция. Левые, небольшой разброс, малая асимметрия (точность, предсказуемость).

2 кластер: Зап. Македония, Аттика, Пелопоннес, Юж. Эгейский бассейн. Менее левые, ближе к центру. Большой разброс. Неоднородность. Растет положительная асимметрия, но в пределах нормы.

3 кластер: Ионические о-ва, Сев. Эгейский бассейн, Крит. Очень левые, единодушные. Растет асимметрия.

4 кластер: Вост. Македония—Фракия. Непредсказуемые.



Рис. 6. Графическое объединение электорального поведения разных регионов Греции с помощью самоорганизующихся карт Кохонена

Сопоставив образованные кластеры с исходными данными (табл. 2), мы также видим большую правдоподобность сближения/отдаления различных регионов по итогам выборов.

Общие выводы

Исследования проводились и для другого количества кластеров (от 3 до 5), но общие результаты совпадают. Самообучающиеся карты Кохонена показали более точные, адекватные, лучше интерпретируемые результаты, чем методы кластерного анализа методом k -средних. Представляется перспективным дальнейшее использование моделей нейронных сетей при исследовании электорального поведения, а также дополнение в качестве еще одной переменной, получаемой из первоначальной агрегированной, эксцесса распределения голосов для каждого региона.

Список литературы

1. *Ахременко А.С.* Количественный анализ результатов выборов: современные методы и проблемы. М., 2008.
2. *Ахременко А.С.* Структуры электорального пространства. М., 2007.
3. *Дебок Г., Кохонен Т.* Анализ финансовых данных с помощью самоорганизующихся карт. М., 2001.
4. *Лапкин В.В.* Возможности количественного описания электоральной динамики // Полис. 2000. № 2.
5. *Липсет С., Роккан С.* Структуры размежеваний, партийные системы и предпочтения избирателей. Предварительные замечания // Политическая наука. 2004. № 4.
6. *Мелешкина Е.Ю.* «Воронка причинности» в электоральных исследованиях // Полис. 2002. № 5.
7. *Мелешкина Е.Ю.* Концепция социально-политических размежеваний: проблема универсальности // Политическая наука. 2004. № 4.
8. *Мелешкина Е.Ю.* Факторы структурирования. Структура и динамика российского электорального пространства // Полис. 2000. № 2.
9. *Новинская М.И.* Реплика о структуре. Структура и динамика российского электорального пространства // Полис. 2000. № 2.
10. *Пантин В.И.* Теоретическое описание: проблемы и подходы. Структура и динамика российского электорального пространства // Полис. 2000. № 2.
11. *Петрунин Ю.Ю.* Классификация при ограниченной информации: микроанализ данных // Вестн. Моск. ун-та. Серия 21 «Управление (государство и общество)». № 2. 2006.
12. *Петрунин Ю.Ю., Зернова Ю.А.* Статистические и нейросетевые методы исследования политической ситуации во Франции на примере региональных выборов 1998 и 2004 годов // Государственное управление. Электронный вестник. 2008. № 14 — http://e-journal.spa.msu.ru/14_2008Petrunin_Zernova.html
13. *Соловьев А.И.* Концепт и коммуникативный метод. Структура и динамика российского электорального пространства // Полис. 2000. № 2.
14. *Яргомыская Н.Б.* Избирательная система. Структура и динамика российского электорального пространства // Полис. 2000. № 2.
15. *Arrow K.* Social Choice and Individual Values. N.Y., 1951.
16. *Black D.* The Theory of Committees and Elections. Cambridge University Press, 1958.
17. *Downs A.* An Economic Theory of Democracy. N.Y., 1957.
18. *Hotelling H.* Stability in Competition // Economic Journal. 1929. N 39.
19. *Kohonen N.* Self-Organizing Maps. N.Y., 1995.
20. *Norris P.* Electoral Engineering // Voting Rules and Political Behavior. Cambridge University Press, 2004.