

И. К. Будникова, Е. В. Приймак, О. И. Илларионова

КЛАСТЕРНЫЙ АНАЛИЗ В СЕГМЕНТИРОВАНИИ РЫНКА ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИИ И МОЩНОСТИ

Ключевые слова: кластерный анализ, иерархический и итерационный методы, дендрограмма, сегмент рынка.

В работе рассматриваются возможности применения кластерного анализа для сегментирования оптового рынка электрической энергии и мощности. Для обоснования правильности выбора кластеров анализируются математические характеристики кластеров.

Keywords: cluster analysis, hierarchical and iterative methods, the dendrogram, segment of the market.

In this work the possibility of application of cluster analysis to segment the wholesale market of electric energy and power. To justify the choice of cluster analyses the mathematical characteristics of clusters.

Оптовый рынок электрической энергии и мощности (ОРЭМ) — сфера обращения электрической энергии и мощности в рамках Единой энергетической системы России, с участием крупных производителей и покупателей электрической энергии и мощности.

В настоящее время на территории Российской Федерации действует двухуровневый оптовый и розничный рынок электроэнергии и мощности. На оптовом рынке электроэнергии функционируют несколько секторов, различающихся условиями заключения сделок и сроками поставки. В их число входят:

Регулируемый договор (РД) – цены на поставку электрической энергии и мощности устанавливаются и рассчитываются государством.

Свободный договор (СД) – участники рынка самостоятельно определяют контрагентов, цены и объемы поставки.

Рынок на «сутки вперед» (РСВ) представляет собой проводимый коммерческим оператором конкурентный отбор ценовых заявок поставщиков и покупателей за сутки до реальной поставки.

Балансирующий рынок (БР) – торговля отклонениями, позволяет приводить в соответствие объемы производства и потребления электроэнергии в реальном времени.

Помимо рынка электроэнергии, существует рынок мощностей. Целью формирования рынка мощностей является создание благоприятных условий для привлечения инвестиций в электроэнергетику, обеспечивающих создание новых генерирующих мощностей в объеме, достаточном для удовлетворения спроса на электроэнергию.

На рынке мощностей так же существуют несколько секторов: сектор регулируемых договоров (РД), сектор свободных договоров (СД) и конкурентный отбор мощности (КОМ). Понятия регулируемых и свободных договоров аналогичны предыдущим. По итогам конкурентного отбора мощности генерирующим компаниям дается право получить с оптового рынка оплату за мощность в следующем периоде, что гарантирует оплату мощности поставщику на период строительства новых мощностей.

С 2008 года ОАО «Генерирующая компания» является полноправным субъектом оптового рынка электроэнергии и мощности. Основными видами деятельности компании являются выработка электрической и тепловой энергии на электростанциях, включенных в единую энергетическую систему; выдача энергии электрическим и тепловым сетям с поддержанием нормального качества отпускаемой энергии. Реализация электрической и тепловой энергии производится круглогодично. Однако максимальный объем производства и потребления приходится на 1-й и 4-й кварталы года – осенне-зимний максимум [1].

В данной работе проведены исследования структуры продаж на оптовом рынке электроэнергии и мощности с целью сегментирования рынка для последующего прогнозирования его конъюнктуры.

Исследования выполнены методом кластерного анализа в пакете Statistica [2, 3, 4].

Кластерный анализ представляет собой класс методов, используемых для классификации объектов или событий в относительно однородные группы, которые называют кластерами.

Для изучения полученного разбиения объектов на однородные группы применяют математические характеристики кластеров, такие как:

- центр кластера – это среднее геометрическое место точек в пространстве переменных;
- дисперсия кластера – это мера рассеяния точек в пространстве относительно центра кластера;
- среднеквадратичное отклонение объектов относительно центра кластера;
- радиус кластера – максимальное расстояние точек от центра кластера.

Методы кластеризации делятся на иерархические и неиерархические (итерационные).

Иерархические методы кластеризации различаются правилами построения кластеров. В качестве правил выступают критерии, которые используются при решении вопроса о "схожести" объектов при их объединении в группу

Иерархические алгоритмы связаны с построением дендрограмм, которые описывают близость отдельных точек и кластеров друг к другу. Они представляют в графическом виде последовательность объединения кластеров. Преимуществом

иерархических методов кластеризации является их наглядность.

Разбиение на кластеры зависит от выбора меры сходства или расстояния между объектом и кластером и метода кластеризации.

Для вычисления расстояния между объектами используются различные меры сходства (близости), называемые также метриками. Евклидово расстояние – наиболее общий тип расстояния. Оно попросту является геометрическим расстоянием в многомерном пространстве.

Выбрав меру сходства, нужно выбрать метод (правило) объединения. На первом шаге, когда каждый объект представляет собой отдельный кластер, расстояния между этими объектами определяются выбранной мерой. Когда связываются вместе несколько объектов, следует определить расстояния между кластерами. Другими словами, необходимо установить правило объединения для двух кластеров. Существуют различные правила, называемые методами объединения или связи для двух кластеров. Правило объединения – метод одиночной связи (метод ближайшего соседа), это метод, в котором расстояние между двумя кластерами определяется расстоянием между двумя наиболее близкими объектами (ближайшими соседями) в различных кластерах.

Наиболее распространенным среди неиерархических методов является алгоритм *K*-средних. В отличие от иерархических методов, для использования этого метода необходимо иметь гипотезу о наиболее вероятном количестве кластеров. Алгоритм *K*-средних строит *K* кластеров, расположенных на возможно больших расстояниях друг от друга. Основной тип задач, которые решает алгоритм *K*-средних – наличие предположений (гипотез) относительно числа кластеров, при этом они должны быть различны настолько, насколько это возможно. Выбор числа *K* может базироваться на результатах предшествующих исследований, теоретических соображениях или интуиции.

Метод основан на минимизации суммы квадратов расстояний между каждым элементом исходных данных и центром его кластера.

В процессе реализации алгоритма меняются значения центров кластеров и принадлежность элементов текущему кластеру. Каждое изменение ведет к уменьшению функции цели.

После получения результатов кластерного анализа методом *K*-средних следует проверить правильность кластеризации (т.е. оценить, насколько кластеры отличаются друг от друга). Для этого сравниваются значения дисперсий. Чем меньше значение внутригрупповой дисперсии и больше значение межгрупповой дисперсии, тем лучше признак характеризует принадлежность объектов к кластеру.

В работе проведем кластерный анализ на основании исходных данных по объемам поставки электроэнергии и мощности на ОРЭМ ОАО «Генерирующей компанией». В таблице 1 приведены объемы поставки в процентах по разным договорам [1].

Таблица 1 – Структура продаж на ОРЭМ

	Объемы поставки электроэнергии и мощности, %						
	РДэ	РСВэ	БРэ	СДэ	РДм	КОМм	СДм
1	88,29	8,87	0,6	2,24	76,23	8,36	15,41
2	101,4	-4,95	0,73	2,82	76,08	10,64	13,28
3	97,43	2,19	0,38	0	76,41	22,21	1,38
4	95,77	4,48	-0,25	0	75,14	5,32	19,54
5	90,78	5,88	1,08	2,26	74,15	4,63	21,22
6	78,8	17,33	-0,08	3,95	74,12	-0,66	26,54
7	66,9	13,47	2,57	17,06	59	-0,3	41,3
8	68,1	13,57	2,1	16,23	58,16	-0,58	42,42
9	76,41	1,73	1,82	20,06	60,13	0,1	39,77
10	66,38	16,27	2,83	14,62	59,09	-0,3	41,21
11	72,43	10,48	0,05	17,04	60,51	-0,1	39,59
12	73,21	0,35	1,64	24,8	60,2	-0,03	39,83
13	54,14	8,94	2,94	33,98	53,57	-0,71	47,14
14	53,1	38,55	1,62	6,3	54,89	-0,07	45,28
15	57,85	28,22	1,37	12,56	53,2	0	46,8
16	68,14	24,9	2,78	4,18	51,2	0,04	48,76
17	60,02	32,84	0,3	6,84	51,41	0	48,59
18	59,29	29,7	0,43	10,58	51,41	0	48,59
19	30,11	64,54	-4,78	10,13	46,54	0	53,46
20	29,86	66,88	-5,76	9,02	38,93	0,59	60,48
21	33,03	59,73	-6,36	13,6	37,15	0	62,85
22	40,87	53,3	-3,77	9,6	37,68	0	62,32
23	41,04	47,69	1,29	9,98	35,59	0	64,41
24	40,69	46,99	0,6	11,72	35,58	0	64,42

Крайний левый столбец таблицы представляет последовательную нумерацию месяцев с января 2009 года по декабрь 2010 года.

В процессе отработки методики кластерного анализа в пакете Statistica для метода иерархической классификации принято: правило объединения – метод одиночной связи и мера близости – Евклидово расстояние.

Результат иерархической классификации представлен на рис.1. Анализируя дендрограмму можно сказать, что получили три кластера:

- первый кластер включает элементы СДм, РСВэ;
- второй кластер включает элементы РДм, РДэ;
- третий кластер включает элементы СДэ, КОМм, БРэ.

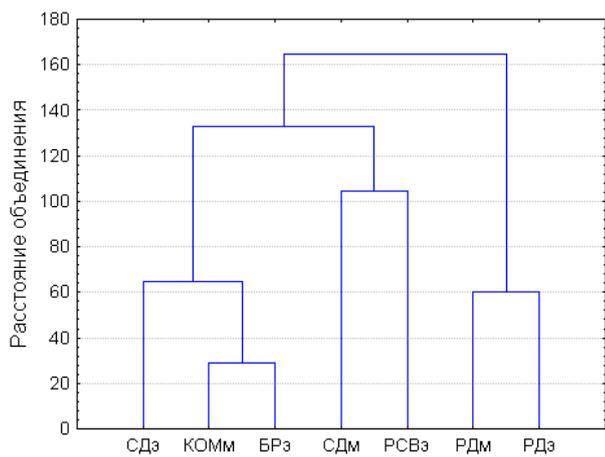


Рис. 1 – Дендрограмма кластеризации

Матрица расстояний (таблица 2) показывает значения евклидова расстояния между парами элементов. Совокупный анализ дендрограммы и матрицы расстояний позволяет обоснованно утверждать, что при иерархическом методе получили три кластера, которые исходя из поставленной задачи исследования, можно классифицировать сегментами рынка.

Таблица 2 – Матрица расстояний

Вид договора	Евклидово расстояние						
	РДэ	РСВэ	БРэ	СДэ	РДм	КОМм	СДм
РДэ	0	281	329	289	60	318	214
РСВэ	281	0	166	133	227	162	104
БРэ	329	166	0	65	283	29	221
СДэ	289	133	65	0	240	71	168
РДм	60	227	283	240	0	272	165
КОМм	318	162	29	71	272	0	219
СДм	214	104	221	168	165	219	0

На втором этапе исследования выполнен кластерный анализ методом *K*-средних. Целью метода *K*-средних является классификация объектов в заданное количество кластеров. Алгоритм перемещает объекты в разные кластеры с целью минимизации изменчивости внутри кластеров и максимизации изменчивости между кластерами. В этом методе задается число кластеров, полученное на основании предыдущего исследования – метода иерархической кластеризации.

В результате выполнения данного метода получаем следующую информацию: средние для каждого кластера (усреднения производятся внутри кластера); евклидовы расстояния и квадраты евклидовых расстояний между кластерами; описательные статистики для каждого кластера. Дисперсионный анализ выводит значения межгрупповых и внутригрупповых дисперсий, а также *F*-критерий Фишера и уровень значимости. Анализ перечисленных пара-

метров подтверждает, что кластеризация выполнена качественно.

Таблица 3 показывает количество кластеров и их классификацию по принадлежности. Расстояние до центра кластера обосновывает принадлежность элемента к конкретному кластеру.

Таблица 3 – Состав кластеров

Вид договора	Элемент кластера	Номер кластера	Расстояние до центра кластера
РДэ	1	2	6,15
РСВэ	2	1	10,64
БРэ	3	3	4,86
СДэ	4	3	9,03
РДм	5	2	6,15
КОМм	6	3	5,92
СДм	7	1	10,64

Таким образом, при выполнении кластерного анализа методом *K*-средних, получили точно такие же кластеры, как при иерархической классификации.

В первый кластер объединяются элементы, которые представляют сектор конкурентного отбора ценовых заявок поставщиков и покупателей за сутки до реальной поставки электрической энергии (РСВэ) и сектор свободных договоров на цены и объемы поставки мощности (СДм).

Во второй кластер объединяются элементы, которые представляют сектор регулируемых договоров на поставку электрической энергии (РДэ) и сектор регулируемых договоров на поставку мощности (РДм).

В третий кластер объединяются элементы, которые представляют: сектор торговли отклонениями, то есть балансирующий рынок по объему производства и потребления электроэнергии в реальном времени (БРэ), сектор свободных договоров по выбору контрагентов, определению цены и объемов поставки электрической энергии (СДэ), а также сектор конкурентного отбора мощности (КОМм).

В данном случае график средних (рис.2) для трех кластеров показывает динамику изменения средних значений кластеров по времени – за два года. При хорошей кластеризации должны быть получены сильно отличающиеся средние для всех измерений или хотя бы большей их части, что наблюдаем на графике.

Программная реализация алгоритмов кластерного анализа, представленная в пакете Statistica, является удобным инструментом для многомерного статистического анализа в сегментации рынка.

В ходе исследования выделено три сегмента рынка, которые позволяют анализировать динамику состояния рынка.

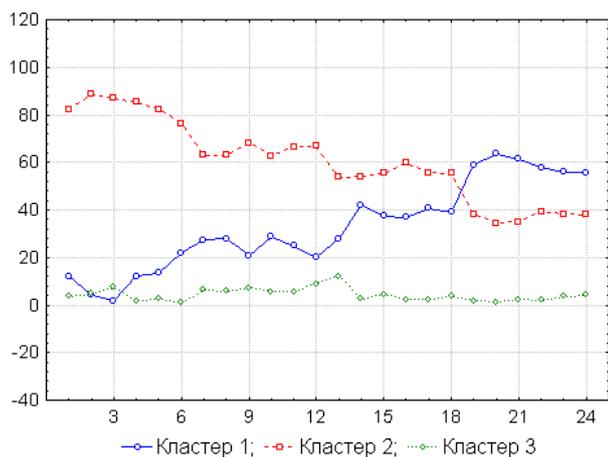


Рис. 2 – График средних

Исходя из полученных результатов, можно сделать вывод о том, что методику кластерного ана-

лиза можно рекомендовать для обработки статистических данных по структуре продаж на ОРЭМ с целью формирования прогнозов для управления процессом поставки электроэнергии и мощности, а также расчета возможных рисков.

Литература

1. ОАО «Генерирующая компания»: <http://www.tatgencom.ru/operations/realization>.
2. Э.А. Вуколов, *Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL*. ФОРУМ: ИНФРА-М, Москва, 2010. 464 с.
3. И.К. Будникова, Е.В. Приймак, *Вестник Казанского технологического университета*, **15**, 15, 247–249 (2012).5/5
4. И.К. Будникова, Е.В. Приймак, *Вестник Казанского технологического университета*, **15**, 20, 199–202 (2012).

© **И. К. Будникова** - канд. техн. наук, доц. каф. инженерная кибернетика КГЭУ, ikbudnikova@yandex.ru; **Е. В. Приймак** – канд. хим. наук, доц. каф. аналитической химии, сертификации и менеджмента качества КНИТУ, Lenaprima@yandex.ru; **О. И. Илларионова** – студ. КГЭУ.