

ПРИМЕНЕНИЕ МЕТОДА КЛАСТЕРНОГО АНАЛИЗА ПРИ ГРУППИРОВКЕ СТРАН ДЛЯ ПРОГНОЗИРОВАНИЯ СПРОСА НА НЕФТЕПРОДУКТЫ¹

COUNTRIES GROUPING FOR PETROLEUM PRODUCTS DEMAND FORECASTING USING CLUSTER ANALYSIS

Д.А. ГРУШЕВЕНКО – ведущий эксперт Центра изучения мировых энергетических рынков Института энергетики Высшей школы экономики, м.н.с., ИНЭИ РАН

D.A. GRUSHEVENKO – Leading expert Center for Energy Market Studies, Energy Institute Higher school of Economics, Junior Researcher, ERIRAS

Е.В. ГРУШЕВЕНКО – н.с., Центр изучения мировых энергетических рынков, ИНЭИ РАН, ст. преподаватель кафедры системных исследований в энергетике РГУ нефти и газа им. И.М. Губкина

E.V. GRUSHEVENKO – Senior Researcher, Center for Energy Market Studies, ERIRAS, senior lecturer, System research of energy markets, of the RSU n.a. Gubkinl.M.

- ▶ Прогнозирование спроса на нефть и нефтепродукты является сложной и комплексной задачей. Зачастую, для облегчения ее выполнения различные организации группируют страны и осуществляют прогнозирование по укрупненным регионам, что не всегда корректно в силу наличия существенных особенностей в закономерностях развития спроса в отдельных странах. Для проверки адекватности проводимых агрегаций, применяемых при прогнозировании спроса, авторы исследования предлагают использовать инструменты кластерного анализа.
- ▶ Petroleum demand forecasting is a very complicated and complex task. To facilitate this task, different organizations group countries, relying to their geographical location. This aggregation method could be incorrect in case of difference in the oil demand trends in different countries. To challenge the adequacy of geographical countries aggregation authors use the cluster analyses methods.

КЛЮЧЕВЫЕ СЛОВА: модельный комплекс, прогноз, нефтепродукты, спрос, мировая энергетика

KEYWORDS: modeling complex forecast of the oil market, petroleum products, demand, energy market

Введение

Зачастую при прогнозировании будущих объемов спроса на нефть и нефтепродукты, расчеты осуществляются не по каждой стране в отдельности, а по группам стран, объединенных по территориальному признаку. Подобная агрегация заметно снижает трудоемкость процесса, но зачастую может вносить существенные неточности в расчеты, так как факторы, влияющие на спрос, отнюдь не одинаковы для стран, расположенных в пределах одного географического региона.

Одна из наиболее распространенных моделей прогнозирования, разработанная Международным энергетическим агентством (МЭА), ETSAP-TIAM, осуществляет прогнозирование по следующим географическим регионам: Африка, Австралия и Новая Зеландия, Канада, Центральная и Южная Америка, Китай, Восточ-

ная Европа, Бывший СССР, Индия, Япония, Мексика, Ближний Восток, прочие страны развивающейся Азии, Южная Корея, США, Западная Европа [?]. Агрегация по этим группам означает, что для всех стран, вошедших в определенную группу, прогнозирование осуществляется по единой формуле, без частных изменений коэффициентов для каждой страны. При этом возникают сомнения, что тенденции в развитии спроса на нефть и отдельные нефтепродукты будут едиными для нефтедобывающих России, Азербайджана и Казахстана, и для нетто-импортеров нефти: Украины и Армении, вошедших в единую группу «Страны Бывшего СССР».

¹ Подготовлено при грантовой поддержке РФФ

² Pernille Seljom Global analyses with ETSAP-TIAM // CenSES Annual Conference, 30 November 2012.

Аналогично, группировки проводятся и в американской модели Generate World oil Balance (GWOB), ключевой модели системы WEPS+, используемой при формировании ежегодного обзора International Energy Outlook [3] Агентства энергетической информации США. Американская система прогнозирования детализирует спрос и предложение на нефть, и нефтепродукты по 44 странам. Региональные группировки в модели сохраняются для «Прочих стран ОЭСР» (Исландии, Люксембурга, Новой Зеландии); Стран ОПЕК (Алжир, Иран, Нигерия, Саудовская Аравия, Венесуэла); Восточной Европы (Казахстана и Украины); Африки (Камеруна, Египта, Кении, Марокко, Судана, Туниса, Уганды); Латинской Америки (Боливии, Колумбии, Коста-Рики, Доминиканы, Эквадора, Панамы, Парагвая, Перу, Уругвая); Прочих стран мира (Бангладеш, Израиля, Мьянмы, Пакистана, Сирии, Вьетнама) [4]. Как видно из этого списка, WEPS не всегда логично агрегирует страны даже по географическому признаку, объединяя прогнозы для стран внутри группы едиными системами уравнений. Также вызывает сомнения выбор единого тренда в развитии спроса на нефть и нефтепродукты, к примеру, для Казахстана и Украины – это страны, которые по-разному обеспечены нефтяным сырьем и имеют разные стимулы к энергосбережению. Не менее странно объединение в одну группу развитого Израиля и Бангладеш, где более 85% населения, по оценкам ООН, находятся за чертой бедности.

В целях устранения этих недочетов, существующих в современных системах прогнозирования, для минимизации ошибки при определении будущих трендов в развитии спроса при агрегации стран по территориальному принципу, предлагается использовать методы кластерного анализа для проверки адекватности группировки стран по географическому признаку.

Методология исследования

Кластерный анализ – «совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных»⁵. Методики кластерного анализа используются, когда каждый объект исследования обладает набором схожих, но несоизмеримых характеристик, проводится посредством формализации всех этих характеристик.

Методики кластерного анализа, в отличие от задач классификации, позволяют группировать объекты:

- с применением статистических и математических инструментов классификации. Так, если задача классификации, как правило, решается

через качественный анализ параметров исследуемых объектов классификации, то методы кластерного анализа, как правило, построены на статистическом анализе набора параметров;

- с одновременным учетом различных типов показателей, описывающих объект. Так, если для качественного анализа обычно применяются однотипные показатели (например, балансы производства и потребления энергоносителей), то методы кластеризации позволяют дополнительно учитывать и другие разнородные показатели.

Задача данной работы – проверить методами кластерного анализа корректность объединения отдельных стран в региональные группы по географическому признаку и задания единых математических формул при прогнозировании спроса на нефть и нефтепродукты для стран внутри выделенных региональных групп. Поскольку задача региональной агрегации решается для последующего решения задачи прогнозирования спроса на нефть и отдельные нефтепродукты, видится целесообразным использовать при кластеризации три параметра, определяющие спрос на нефть и нефтепродукты:

1) душевой ВВП. Большинство исследований, направленных на прогнозирование спроса, сходятся в том, что именно экономический и демографический рост оказывают основное влияние на спрос на нефть и нефтепродукты [6, 7, 8, 9]. Для снижения числа параметров, по которым будет проводиться кластеризация, целесообразным видится применение синтетического показателя душевого ВВП, отображающего качественный экономическо-демографический рост для отдельных стран;

2) обеспеченность нефтью, выраженная через показатель разницы между собственной добычей нефти внутри страны и потреблением нефти на НПЗ – нетто-экспорт. Логично, что страны, вынужденные импортировать нефть и нефтепродукты, будут активнее стимулировать энергосбережение, снижая нефтеемкость экономики и сдерживая спрос на нефтепродукты;

3) соотношение потребления бензина и дизельного топлива, отображающее потребительские предпочтения страны в ключевом секторе потребления нефти – транспортном. Именно этот показатель будет характеризовать предпочтительное для населения моторное топливо.

⁶ О. Б. Брагинский. Прогнозирование российского рынка автомобильных видов топлива. Открытый семинар «Экономические проблемы энергетического комплекса» 24 апреля 2012 года.

⁷ Бобылев Ю. Н., Приходько С. В., Дробышевский С. М., Тагор С. В. Факторы формирования цен на нефть. Институт экономики переходного периода, М.: 2006.

⁸ Миловидов К. Н., Коржубаев А. Г., Эдер Л. В. Нефтегазообеспечение глобальной экономики: Учебное пособие. – М.: ЦентрЛит-НефтеГаз, 2006. – 400с.

⁹ Макаров А. А., Митрова Т. А., Григорьев Л. М. Прогноз развития энергетики мира и России до 2040 года. ИНЭИ РАН – АЦ при правительстве РФ, М.: 2014.

³ EIA International Energy Outlook, 2013

⁴ The World Energy Projection System Plus (WEPS+): Global Activity Module, October 2013

⁵ Бузыкина Т. А. Кластерный анализ на примере химического кластера Самарской области. Аудит и Финансовый анализ, № 4, 2010.

Для решения поставленной задачи исследования был выбран алгоритм проведения кластеризации, основанный на построении оптимальной классификации объектов по группам. При этом оптимальная классификация может быть определена как требование минимизации среднеквадратической ошибки разбиения:

$$e^2(X, L) = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2$$

где: c_j – «центр масс» кластера j (точка со средними значениями характеристик для данного кластера).

Этот алгоритм более известен, как метод k -средних. Суть алгоритма заключается в построении кластеров, которые располагаются на как можно больших расстояниях друг от друга. При этом количество гипотез относительно числа кластеров – k должно быть максимально большим, или же основываться на предыдущих исследованиях, иерархических методах (если позволяет массив данных), теоретических выкладках или интуиции. Сам алгоритм выглядит следующим образом:

- случайно выбрать k точек, являющихся начальными «центрами масс» кластеров;
- отнести каждый объект к кластеру с ближайшим «центром масс»;
- пересчитать «центры масс» кластеров согласно их текущему составу;
- если критерий остановки алгоритма не удовлетворен, вернуться к п. 2.

В качестве критерия прекращения работы алгоритма выбирают минимальное изменение среднеквадратической ошибки. Так же возможно останавливать работу алгоритма, если на шаге 2 не было объектов, переместившихся из кластера в кластер [10]. При использовании данного алгоритма необходимо предварительно сформулировать гипотезу относительно наиболее вероятного количества кластеров – k .

Результаты исследования

Перед проведением кластерного анализа было определено количество кластеров, учитывая различные сочетания показателей, использованных в качестве входных данных. Поскольку все показатели являются двусоставными, то количество возможных сочетаний – восемь.

Расчеты проводились для отдельных стран в динамике, по состоянию на 1980 г., 2010 г. и на прогнозный – 2040 г. (исходя из прогнозных значений, опубликованных в работе ИНЭИ РАН-АЦ при правительстве РФ¹¹).

¹⁰ И. А. Чубукова. Data Mining. Учебное пособие. – М.: Интернет-Университет Информационных технологий; БИНОМ. Лаборатория знаний, 2006. – 382с. – (Серия «Основы информационных технологий»).

¹¹ Макаров А. А., Григорьев Л. М., Митрова Т. А. Прогноз развития энергетики мира и России до 2040 года, М.: 2014.

По результатам расчетов получено восемь различных (по составу стран) кластеров (). Важно отметить, что на протяжении времени кластеры менялись не только по составу, но и по самой характеристике кластеров. Так, к примеру, к 2040 г. исчезает один кластер с суперэкспортерами – Россия и Саудовская Аравия – и число кластеров уменьшается до семи. Это объясняется значительным ростом добычи нефти в кластере крупных экспортеров и снижением добыче в кластере суперэкспортеров.

Несмотря на разницу кластеров по составу и качественным характеристикам, все же можно выделить ряд кластеров с определёнными признаками (стабильные группы стран), например:

- в первом кластере – импортёры с высоким ВВП на душу и преобладанием дизельного топлива;
- в третьем кластере – импортеры со средним ВВП на душу;
- в пятом кластере – крупные потребители бензина;
- в седьмом кластере – крупные экспортеры (за исключением России и Саудовской Аравии);

Для этих стабильных групп стран, безусловно, прогнозирование спроса на нефтепродукты целесообразно проводить по общему для группы (кластера) тренду. Однако важно отметить, что не всегда эти страны связаны едиными географическими признаками: в одном кластере могут присутствовать и некоторые европейские страны, и страны Латинской Америки, и страны бывшего СССР.

При этом существует достаточная большая группа стран, которая на протяжении исследуемого периода так и не формирует стабильный кластер, перемещаясь из одного кластера в другой. Это говорит о нестабильности трендов и закономерностей прогнозирования спроса, а значит, о некорректности агрегации этих стран по географическому признаку для проведения модельных расчетов и прогнозирования спроса в связи с невозможностью построить единое уравнение, которое адекватно описывало бы спрос в странах одного географического региона.

Выводы

Кластеризация стран по ключевым показателям, определяющим будущие и настоящие значения спроса на нефть и нефтепродуктов, показала нестабильность трендов в развитии спроса на нефть и нефтепродуктов для стран, которые могли бы объединяться в единый географический регион. Это свидетельствует о том, что прогнозирование спроса на нефтепродукты практически для всех стран мира необходимо проводить отдельно, а не по географическому признаку. Региональные группировки могут быть сохранены только для ряда стран Европы.

Результаты кластерного анализа за 1980г., 2010г. и 2040г.*

	1	2	3	4	5	6	7	8
1980г.	Канада, Австрия, Бельгия, Дания, Финляндия, Франция, Германия, Греция, Италия, Люксембург, Нидерланды, Испания, Швеция, Исландия, Норвегия, Швейцария, Ирак, Израиль, Ливия	США, Япония	Бразилия, Чили, Аргентина, Кипр, Венгрия, Ирландия, Польша, Португалия, Румыния, Великобритания, Малайзия, ЮАР	Мексика, Колумбия, Эквадор, Венесуэла, Австралия, Новая Зеландия	Нигерия	Болгария, Чехия, Мальта, Словакия, Албания, Турция, Корея, Пакистан, Китай, Индия, Индонезия, Непал, Ангола Шри Ланка, Иран, Алжир	Кувейт, Катар, ОАЭ	СССР Саудовская Аравия
2010г.	Франция, Германия, Италия, Испания, Япония, Корея	США, Китай	Бразилия, Чили, Колумбия, Эквадор, Аргентина, Австрия, Бельгия, Болгария, Кипр, Хорватия, Чехия, Дания, Эстония, Финляндия, Греция, Венгрия, Ирландия, Латвия, Литва, Нидерланды, Польша, Португалия, Румыния, Словакия, Словения, Швеция, Великобритания, Албания, Босния и Герцеговина, Македония, Сербия, Швейцария, Турция, Киргизия, Таджикистан, Туркмения, Узбекистан, Армения, Белоруссия, Грузия, Молдавия, Украина, Австралия, Новая Зеландия, Монголия, Индонезия, Малайзия, Пакистан, Непал, Шри Ланка, Израиль, ЮАР, Индия	Люксембург, Исландия, Мальта, Катар	Ангола Нигерия		Канада, Мексика, Венесуэла, Норвегия, Казахстан, Азербайджан, Иран, Ирак, ОАЭ, Алжир, Ливия, Кувейт	Россия, Саудовская Аравия
2040г.	Австрия, Бельгия, Финляндия, Германия, Ирландия, Люксембург, Нидерланды, Швеция, Великобритания, Исландия, Норвегия, Швейцария, Австралия, Япония, Корея, Израиль	США, Индия	Мексика, Чили, Аргентина, Болгария, Хорватия, Чехия, Дания, Эстония, Франция, Греция, Венгрия, Италия, Латвия, Литва, Мальта, Польша, Португалия, Словакия, Словения, Испания, Турция, Туркмения, Белоруссия, Новая Зеландия, Малайзия, Колумбия, Эквадор, Венесуэла, Кипр, Румыния, Албания, Босния и Герцеговина, Македония, Сербия, Киргизия, Таджикистан, Узбекистан, Армения, Азербайджан, Грузия, Молдавия, Украина, Монголия, Индонезия, Пакистан, Непал, Шри Ланка, Иран, Алжир, Ангола, ЮАР	Катар	Нигерия	Китай	Канада, Бразилия, Казахстан, Россия, Ирак, Кувейт, Ливия Саудовская Аравия, ОАЭ,	

*Состав кластеров по годам не совпадает
Источник: составлено авторами