

Обзор методов выявления точек роста экспортного потенциала с применением интеллектуального анализа данных*

Р.Е. Суворов, Д.А. Девяткин, Н.И. Усенко, Ю.С. Отмахова

Аннотация. В работе представлен обзор подходов и статистических баз данных для анализа возможностей расширения экспортного потенциала. Рассмотрены основные отечественные и зарубежные работы в области анализа международных торговых потоков. Проанализированные материалы разделены авторами на три основные группы согласно их целевому назначению. Предложена постановка задачи интеллектуального анализа данных для выявления точек роста экспортного потенциала и сделаны выводы о дальнейших направлениях исследований с применением интеллектуального анализа данных баз статистики международной торговли и семантического поиска по коллекциям научно-технических документов.

Ключевые слова: экспортный потенциал, интеллектуальный анализ данных, международная торговля, таможенная статистика, точки роста, научно-технические решения, технологии, семантический поиск.

Введение

В России отмечается серьезная неблагоприятная тенденция снижения стоимостных объемов экспорта и его относительной доходности. По данным таможенной статистики в 2015 г. экспорт составил 345,9 млрд. долл. США, что меньше по сравнению с 2014г. на 31,1%. В 2016г. тенденция снижения сохранилась, т.е. величина российского экспорта снизилась на 16,9% по сравнению с 2015г. и составила 285,5 млрд. долл. США. На фоне снижения общего объема экспорта в стоимостном выражении экспорт продовольственных товаров и сельскохозяйственного сырья отличается положительной динамикой удельного веса данной товарной группы в общем объеме. Так, в 2006 г. удельный вес данной товарной группы составил 1,8%, а в 2016 г. – 6%.

Данные изменения связаны с наличием таких разнонаправленных тенденций как снижение общего объема российского экспорта и роста экспорта продовольствия. Вместе с тем, стоит отметить, что по основным товарным позициям отмечается снижение стоимостных объемов экспорта товаров при одновременном росте их физических объемов [1]. Например, в настоящее время Россия является одним из ведущих экспортеров продовольственного зерна в мире. За прошлый сельскохозяйственный

год (июнь 2015 г. – июнь 2016 г.) величина российского экспорта оценивается в 24,5 млн. тонн зерна пшеницы, что позволило России впервые в современной истории занять лидирующие позиции в мире [2]. Однако высокие урожаи, большие запасы зерна и высокая конкуренция среди экспортеров на мировом рынке зерна пшеницы влияют на движение цен в сторону их понижения, обуславливают снижение относительного дохода от экспорта пшеницы.

Кроме того, анализ сложившейся товарной структуры российского экспорта продовольствия (по данным FAOSTAT [3], 2013г.) свидетельствует о слабо диверсифицированной товарной структуре экспортной номенклатуры, поскольку всего лишь на две позиции приходится около 40% в общем объеме экспорта продовольствия (экспорт пшеницы -27,5% и подсолнечное масло – 11,7%).

В неблагоприятных условиях антироссийских санкций, сопровождающихся снижением объема импорта, особую актуальность приобретает задача обеспечения роста экспорта за счет не только наращивания физических объемов поставок продукции традиционного экспортного ассортимента, но и посредством корректировки товарной структуры экспорта зарубежных поставок с учетом перспектив глобального спроса на мировом рынке. В свою очередь, точки роста экспортного потенциала нельзя рассматривать в отрыве от имеющихся

* Работа выполнена при финансовой поддержке РФФИ (грант № 16-29-12877)

производственных возможностей, в том числе перспективных технологий, которые в скором времени могут быть внедрены.

Оценка эффективности и реализуемости научно-технических решений и технологий является критически важным элементом при планировании государственных расходов на поддержку отечественных производителей. При этом компании, поставляющие продукцию не только на внутренний, но и на внешний рынки, несут больше пользы для государства по целому ряду причин, в том числе по объему налоговых отчислений, а также по вкладу в экономический рост и национальную безопасность. Таким образом, важно оценивать не только внутренние потребности, но также и перспективы роста экспорта.

В настоящее время нет однозначной трактовки экономического содержания термина «точки роста». В биологии под «точками роста» у растений понимают зоны делящихся клеток, поэтому и в экономике под данным понятием понимают начало нового развития и методы этого развития. Применительно к рынку товаров под точками роста понимают потенциал развития некоего рынка и источники экономического развития. Как указывается в статье [4] «точки роста» возникают на рынке (или сегменте рынка) как возможность удовлетворения первичного спроса на некий товар или услугу. При этом, с точки зрения авторов, сама возможность удовлетворить данный спрос возникает локально и имеет вполне конкретную территориальную, а, следовательно, и национальную привязку.

Необходимо отметить, что непосредственно рост может обеспечиваться как на основе совершенствования технологий и научно-технических решений, так и за счет повышения ресурсного потенциала и эффективности управления и т.д. Для понимания категории «точки роста» большое значение имеют ее свойства, а именно, высокая норма добавленной стоимости предприятий, входящих в производства «точки роста», наличие межотраслевых связей, мультипликативный эффект, устойчивый внешний спрос из-за значительных конкурентных преимуществ. В рамках настоящего исследования приводится обзор существующих подходов к выявлению и оценке точек роста применительно к экспорту.

1. Анализ подходов к определению точек роста

В разделе представлен обзор основных отечественных и зарубежных работ по теме данного исследования. Необходимо заметить небольшое количество работ, посвященных именно иденти-

фикации перспективных научно-технических решений и технологий с точки зрения экспортного потенциала, поэтому критерии поиска были расширены. Статья включалась в обзор, если она:

1. была посвящена одной из следующих тем:
 - анализу международных торговых потоков, направлений экспорта или факторов, влияющих на эффективность экспорта;
 - анализу перспективности научно-технических решений и технологий, выявлению точек роста;
 - исследованиям по смежным тематикам, с описанием методик и технологий, которые можно было бы использовать в качестве дополнительных факторов в рамках данной работы;
2. содержала анализ реальных статистических данных (экспериментальную часть);
3. содержала описание математических методов, с помощью которых выполнялся анализ.

Все найденные работы можно условно разделить на следующие группы.

1. Работы, в первую очередь анализирующие имеющиеся научно-технические решения и технологии, и потом оценивающие их с точки зрения эффективности.
2. Работы, в первую очередь анализирующие международные финансовые потоки:
 - на уровне стран в целом;
 - на уровне отдельных предприятий (firm-level);
 - по товарам.
3. Работы, посвященные источникам данных для анализа *экспорта и международных торговых потоков*.

К первой группе относятся все работы, посвященные применению методов технологического прогнозирования, анализа патентных баз для решения задач экономики.

Например, в статье [5] рассматривается система TechPioneer, реализующая тематический и структурный анализ описаний технологий (в виде патентов) с частичным привлечением экспертов в предметной области. Эксперты отвечают за начальное создание тезауруса релевантной лексики. Тезаурус необходим для построения начального списка ключевых слов для фильтрации патентов, а также для анализа структуры технологий. После задания тезауруса и отбора патентов, выполняется структурный анализ каждой технологии, представленной в выбранных патентах. Затем каждая комбинация морфологических характеристик оценивается по некоторой шкале с привлечением эксперта. После этого выполня-

ется оценка вклада каждого элемента технологии в ее успешность.

В работе [6] авторы делают обзор различных методологий по выявлению перспективных технологий, а также предлагают скомбинировать несколько подходов: библиометрический анализ для раннего обнаружения растущих областей и поиск патентов, а также модель Фишера-Прая для оценки момента времени, соответствующего внедрению.

В статье [7] показано, что изменение кумулятивного числа патентов в некоторой области может быть применено для определения технологических трендов и отражает распространение технологии. Этот подход был применен в этой работе для определения стадии развития технологий, связанных с интеллектуальной энергетикой. Кумулятивное число патентов аппроксимируется с помощью

логистической кривой $C = \frac{L}{a + e^{-bt}}$, где параметры

L , a и b находятся, например, с помощью метода наименьших квадратов. Затем определяется точка перегиба аппроксимирующей кривой, если она уже пройдена, то рост патентной массы замедляется и технология начинает стагнировать. Идея этого подхода состоит в том, что быстрый рост количества патентов, посвященных определенной технологии, говорит о значительных организационных и экономических усилиях, предпринимаемых научными и промышленными организациями для развития этой технологии. Снижение прироста патентной массы может говорить о падении интереса к технологии, что может быть вызвано появлением новой, замещающей технологии.

Вторая группа статей – наиболее многочисленная. Кратко все подходы, описанные в работах этой группы, можно разделить на (а) гравитационные модели; (б) вычислимые модели общего равновесия; (в) ранжирование вариантов с помощью эвристик; (г) анализ структуры графа и прогнозирование его изменчивости с помощью марковских моделей, оценок центральности и анализа потоков; (д) применение общих статистических приемов (регрессий, гистограмм) для ручного анализа ситуации.

Остановимся на наиболее интересных работах чуть подробнее.

В классической работе [8] описывается применение анализа по методике shift-share для выявления наиболее перспективных внешних рынков. Особенностью этой методики является прогнозирование динамики изменения не абсолютных показателей, а долей рынка.

В [9] авторы заключают, что распространенные гравитационные модели плохо подходят для

предсказания факта торговли между некоторыми двумя странами, однако эти модели позволяют достаточно точно оценить и спрогнозировать объем торговли, если известно, что между некоторыми двумя странами вообще есть торговые отношения. В [10] упоминается возможность использования Марковских моделей для прогнозирования изменчивости сети внешнеторговых финансовых потоков.

В [11] исследуется вопрос выявления таких направлений экспорта, которые могут наиболее эффективно развиваться при внутренней финансовой поддержке. Экспериментальная работа проводится на данных экспорта продуктов питания на уровне фирм и товаров. В качестве основы для математического аппарата используются гравитационные модели.

В работе отечественных авторов [12] модель общего равновесия применяется для количественной оценки решений в области государственного регулирования экспорта на примере стран Россия – Беларусь. В работе [13] коллеги также сравнивают различные отрасли промышленности Беларуси с точки зрения экспортного потенциала. Для этого авторы строят линейную модель, связывающую основные факторы (научеёмкость, фондоемкость, прямые затраты промежуточного продукта). Одной из особенностей подхода является совместный учет текущих затрат и затрат овеществленного труда, а также экологической нагрузки.

В [14] анализируется динамика изменения структуры импорта и экспорта мяса крупного рогатого скота с применением международной базы статистики FAOStat. В качестве основных факторов рассматривается скорость и направление изменения объемов производства, потребления, а также объемов существующих потоков экспорта. Проводится PEST-анализ (политико-правовой, экономический, социокультурный и технологический) внешней среды мирового рынка мяса крупного рогатого скота.

В [15] рассматривается задача выявления перспективных направлений экспорта в сфере услуг и товаров вместе с точки зрения возможностей для роста. Подход основан на последовательной фильтрации потенциальных рынков с помощью ряда эвристик, включающих оценку объема рынка в целом, уровня спроса, открытости рынка и т. п. Работа, по сути, является продолжением [16–18] тех же авторов.

В [19] исследуется связь миграционных потоков с внешней торговлей. Авторы заключают, что торговые потоки для дифференцированных товаров положительно и значимо коррелируют с

миграционными потоками. Этот фактор можно учитывать при анализе и оценке перспективности точек роста.

В [20] рассматривается вопрос взаимосвязи экспорта и «сложности» товаров, вводится метрика «сложности» или «технологичности» товаров. Авторы упоминают такие факты как зависимость скорости роста цен на некоторый товар в зависимости от степени его технологичности. Эта метрика может быть применена в качестве одного из факторов при выявлении и оценке перспективности точек роста.

В [21] анализируется целесообразность выхода на международный рынок (начала экспорта) для конкретной фирмы на основе показателей ее прошлой активности, в том числе участие в экспорте, конкурентная среда и т. п. Среди интересных заключений стоит отметить слабое влияние отраслевой государственной поддержки экспорта на его фактические объемы.

В [22] рассматривается связь топологии сети международной торговли (СМТ, International Trade Network, ITN) между странами в целом с топологиями сетей в рамках каждой товарной группы. Предлагается интересная методология для исследования динамики изменения структуры нескольких неоднородных сетей, представляющих торговые потоки между странами по отдельным товарным группам. В результате идентифицированы наиболее активные экспортеры и импортеры по отдельным группам, в том числе в динамике.

В [23] авторами предлагается моделировать структуру и динамику сети международной торговли с помощью классической задачи о выборе шаров из урн. Анализ выполняется на уровне стран в целом и реализуется принцип предпочтительного присоединения (preferential attachment, «богатые богатеют, бедные беднеют»).

В [24] предлагается моделировать структуру и динамику СМТ с помощью гамильтоновой системы. Авторы описывают динамику СМТ в терминах гамильтониана, а также делают предположение о том, что основные положения из области статистической физики будут также применимы и к СМТ.

В [25, 26] рассматривается сеть международной торговли на уровне стран и товаров. С помощью анализа потоков в графе и статистик на вершинах, авторы делают ряд заключений, относящихся к специализации стран, а также доминированию развитых стран в смысле многообразия экспортируемой продукции (принцип предпочтительного присоединения). Среди прочего, авторы эмпирически подтверждают факт, что продукты-товары в основном реализуются между наиболее

близко расположенными странами, в то время как высокотехнологичные товары распространяются фактически по всему миру. Также авторы выявляют страны с аномальным профилем импорта, который может говорить о ряде экономических проблем. Предложенная авторами методика может быть полезна для данной работы в качестве одного из факторов.

В [27] анализируется экспорт в сфере услуг на примере фирм в Германии с целью выявления зависимости направлений, способа экспорта (или использования международного партнерства вместо экспорта) от различных характеристик экспортируемых услуг. Используются закрытые данные Дойчебанка. Помимо прочего, авторы выявили такие стилизованные факты как «экспорт более предпочтителен в страны с более высоким уровнем доходов (для стран с меньшим уровнем доходов более предпочтительно международное партнерство)»; «при продаже в более удаленных странах международное партнерство более выгодно».

Согласно [28], одними из наиболее сильных факторов, определяющих выход или не выход на рынок некоторой другой страны, является объем невозвратных затрат, а также наличие уже каких-то других фирм, экспортирующих в том же направлении. Частично это подтверждается и отечественными учеными, которые посредством опросов и экспертных оценок получили [29] более детальное представление о сложностях выхода на международный рынок для российских производителей. К основным негативным факторам они относят высокие расходы, отсутствие информации о зарубежных рынках, уровне спроса, а также программам господдержки экспортеров. Авторами также была разработана методика многокритериальной оценки экспортного потенциала предприятия, которая может быть применена и для оценки перспективности отраслей и групп товаров в целом (на основе оценки соответствующих предприятий).

В работе иркутских коллег [30] рассматривается выявление новых географических направлений экспорта куриных яиц. Для решения поставленной задачи авторы оценивают насыщенность рынка в различных странах (в основном Ближнего Востока, Юго-Восточной Азии и Тихоокеанского региона), отношение объемов внутреннего производства, цены в странах-кандидатах, их объемы импорта, а также стоимость доставки из Иркутска. Значимыми препятствиями являются законодательство стран в области качества и санитарных требований к импортиру-

емой продукции и развитость деловых и политических связей.

К третьей группе можно отнести следующие работы.

В [31] описывается ВАСИ – база данных по международной торговле на уровне товаров, которая включает более 200 стран и 5000 товаров с 1994 по 2007 год. Благодаря достаточно высокой детализации, есть возможность построить качественные гравитационные модели. Данные доступны для скачивания по подписке COMTRADE.

В [32] описывается набор данных, представляющий экспортные пошлины в разных странах по отдельным категориям продуктов и оценке их влияния на объем торговли.

Подробный список баз торговой статистики приведен на сайте экономиста Всемирного торгового центра и автора ежегодного аналитического отчета о конкурентоспособности малого и среднего бизнеса (SME Competitiveness Outlook) О. Соллердер [33]. Рассмотрим некоторые базы подробнее.

FAOSTAT [3] – база данных Продовольственной и сельскохозяйственной организации ООН. Содержит информацию об уровне производства, экспорта и импорта более чем 90 категорий продовольственной и сельскохозяйственной продукции в 245 странах за период с 1961 года по настоящее время, однако информация об основных торговых контрагентах стран отсутствует.

UN Comtrade Database [34] – база международной торговой статистики ООН. Содержит ежемесячную информацию о величинах торговых оборотов между странами-участниками ООН.

База данных международной торговли французского Центра перспективных исследований и международной информации (CEPii) [35] содержит детализированную информацию о величинах торговых оборотов между странами по 71 категории товаров за период с 1967 года по настоящее время.

База Всемирного Банка [36] предоставляет уровень цен на различные группы промышленных и сельскохозяйственных товаров при двусторонних торговых взаимоотношениях государств. База содержит информацию по 178 странам за 1995-2014 годы.

База таможенной статистики внешней торговли РФ [37] содержит детальную ежемесячную информацию о внешнеторговых операциях компаний-резидентов РФ с указанием групп товаров, величин торговых оборотов и регионов, в которых зарегистрированы эти компании.

Торговая статистика Всемирного торгового центра [38] содержит информацию о величине торговых оборотов между 220 странами. База предоставляет информацию по 5300 категориям товаров.

Патентные базы Федерального института промышленной собственности [39] содержат информацию о более чем 700 тыс. патентов на изобретения, полезные модели и промышленные образцы, выданных в РФ с 1991 года.

Обобщая рассмотренные подходы, сформулируем далее задачу интеллектуального анализа данных для выявления точек роста.

2. Постановка задачи интеллектуального анализа данных для выявления точек роста

Задача интеллектуального анализа данных для выявления точек роста заключается в нахождении таких сочетаний $\langle Product_i, Country_j \rangle$, для которых экспорт продукта $Product_i$ из России в страну $Country_j$ экономически целесообразен, где под $Product_i$ понимается либо отдельный продукт, либо группа продуктов; под $Country_j$ понимается страна, либо группа стран.

Результатом решения этой задачи будет список пар $\langle Product_i, Country_j \rangle$, отсортированный по убыванию оценки «целесообразности». Эта оценка должна быть многофакторной, т. е. (косвенно) учитывать:

- объемы импорта страной $Country_j$ продуктов $Product_i$;
- емкость внутреннего рынка страны-импортера и динамику емкости (она не должна значительно сокращаться в последние несколько лет);
- соотношение объемов импорта и емкости (как оценку разрыва между внутренним производством и импортом);
- развитость торговых и дипломатических связей со страной $Country_j$;
- уровень конкуренции при импорте в страну $Country_j$ (количество поставщиков, насколько они отличаются друг от друга);
- наличие налаженного экспорта товара $Product_i$ из РФ куда-либо (это означает, что $Product_i$ уже производится);
- конкурентность цены (необходимо сравнить среднюю стоимость продукта $Product_i$ в РФ со средней стоимостью в стране $Country_j$);
- другие критерии.

Решение этой задачи требует использования современных методов семантического поиска в коллекциях научно-технических документов, обработки больших данных, машинного обучения и интеллектуального анализа данных статистики международной торговли.

Для идентификации значений многих перечисленных критериев, например прогнозирования динамики емкости рынка, могут применяться методы

машинного обучения, с помощью которых строятся регрессионные модели, основанные на ансамблях деревьев решений или нейронных сетях [40].

Наличие научно-технологических заделов является важным фактором выявления новых точек роста экспортного потенциала. Методы семантического поиска в патентных коллекциях могут применяться для определения возможности наращивания или организации производства определенной группы товаров в России. Благодаря развитию методов автоматической обработки текстов на естественных языках [41, 42] полнотекстовый поиск с учетом синтаксиса и семантики запроса стал наиболее перспективным подходом к анализу коллекций научно-технических документов.

Автоматизированное выявление точек роста предполагает формирование агрегированной статистики по результатам поиска в различных гетерогенных источниках информации. Распространенным подходом для построения подобной статистики является применение технологии mapReduce [43], позволяющей обрабатывать большие массивы данных с помощью распределенных вычислительных систем.

Заключение

Перспективы развития экономики РФ напрямую зависят от ее способности вписаться в будущие продовольственные рынки и сформировать на них сильное торговое предложение. К сожалению, задаче поиска точек роста и анализу научно-технических решений и технологий для расширения направлений продовольственного экспорта уделяется недостаточное внимание. При этом имеющиеся статистические базы данных не являются достаточно полными для проведения исследований по этой теме и их сложно интегрировать друг с другом ввиду особенностей классификации и кодификации товаров. К основным недостаткам рассмотренных работ можно отнести: отсутствие интеграции данных из различных источников; отсутствие связи с имеющимся технологическим заделом. Выделенные недостатки затрудняют и без того сложную задачу объективного сравнения альтернатив при диверсификации структуры российского экспорта.

В условиях сложной макроэкономической ситуации использование возможностей интеллектуального анализа больших объемов статистической информации и создания новых методов объективной оценки состояния и динамики научно-технических решений и технологий для потребностей экономики РФ позволят не только конструировать новые сценарии развития рынка, но и предоставят возможность

определить стратегические направления политики и проводить оценку эффективности предпринимаемых страной мероприятий по поддержке отечественных производителей-экспортеров.

Основным направлением дальнейших исследований является создание объединенной методики выявления точек роста экспортного потенциала на основе интеллектуального анализа баз статистики международной торговли вместе с семантическим поиском в коллекциях научно-технических документов, содержащих описания технологий (в первую очередь, патентные коллекции), а также ее экспериментальная апробация на примере продовольственного рынка Российской Федерации.

Литература

1. *Экспорт-импорт важнейших товаров за январь-декабрь 2015 года* Веб-сайт. Официальный сайт таможенной службы [Электронный ресурс] /- Режим доступа: http://www.customs.ru/index.php?option=com_content&view=article&id=22570#_ftn1, свободный.
2. *Россия обогнала США по экспорту зерна*, [Электронный ресурс], 2016 Режим доступа: <http://vz.ru/news/2016/7/20/822641.html>
3. *FAOSTAT* [Электронный ресурс] / Продовольственная и сельскохозяйственная организация ООН. – Веб-сайт. – Режим доступа: <http://faostat.fao.org/beta/en/#data>, свободный.
4. *Иванова Л.Н., Тверская Г.А.* Точки роста и драйверы роста: к вопросу о содержании понятий // Журнал институциональных исследований. – 2015. – Т. 7, № 2. – С.120-133.
5. *Yoon B.* On the development of a technology intelligence tool for identifying technology opportunity //Expert Systems with Applications. – 2008. – Vol. 35(1). – pp. 124-135.
6. *Daim T. U. et al.* Forecasting emerging technologies: Use of bibliometrics and patent analysis //Technological Forecasting and Social Change. – 2006. – Vol. 73(8). – pp. 981-1012.
7. *Andersen B.* The hunt for S-shaped growth paths in technological innovation: a patent study //Journal of Evolutionary Economics. – 1999. – Vol. 9(4). – pp. 487-526.
8. *Green R. T., Allaway A. W.* Identification of export opportunities: A shift-share approach //The Journal of Marketing. – 1985. – pp. 83-88.
9. *Duenas M., Fagiolo G.* Modeling the International-Trade Network: a gravity approach //Journal of Economic Interaction and Coordination. – 2013. – Vol. 8(1). – pp. 155-178.

10. *Snijders T. A. B.* Models for longitudinal network data //Models and methods in social network analysis. – 2005. – Vol. 1. – pp. 215-247.
11. *Jaud M., Kukenova M., Strieborny M.* Financial Development and Sustainable Exports: Evidence from Firm product Data //The World Economy. – 2015. – Vol. 38(7). – pp. 1090-1114.
12. *Сердюкова Ю. С., Усенко Н. И.* Стратегические приоритеты интеграционного взаимодействия России и Беларуси с позиции обеспечения продовольственной безопасности //Экономические и социальные перемены: факты, тенденции, прогноз. – 2013. – №. 3 (27).
13. *Приходченко О. И.* Выявление экспортных и импортных приоритетов Республики Беларусь на основе моделирования межотраслевых связей. – 2002.
14. *Закшевская Е., Литвиненко Т.* Мировые тенденции в производстве и сбыте мяса КРС //Международный сельскохозяйственный журнал. – 2016. – №. 5.
15. *Grater S. et al.* Linking export opportunities of products and services: the case of South Africa.
16. *Cuyvers L. et al.* A decision support model for the planning and assessment of export promotion activities by government export promotion institutions—the Belgian case //International Journal of Research in Marketing. – 1995. – Vol. 12(2). – p. 173-186.
17. *Cuyvers L.* Identifying export opportunities: the case of Thailand //International Marketing Review. – 2004. – Vol. 21(3). – pp. 255-278.
18. *Cuyvers L., Viviers W. (ed.)*. Export Promotion-A Decision Support Model Approach. – AFRICAN SUN MeDIA, 2012.
19. *Sgrignoli P.* The World Trade Web: A Multiple-Network Perspective //arXiv preprint arXiv:1409.3799. – 2014.
20. *Lall S., Weiss J., Zhang J.* The “sophistication” of exports: a new trade measure //World Development. – 2006. – Vol. 34(2). – pp. 222-237.
21. *Bernard A. B., Jensen J. B.* Why some firms export //Review of Economics and Statistics. – 2004. – Vol. 86(2). – p. 561-569.
22. *Barigozzi M., Fagiolo G., Garlaschelli D.* Multinetwork of international trade: A commodity-specific analysis //Physical Review E. – 2010. – Vol. 81(4). – p. 46-104.
23. *Peluso S. et al.* International Trade: a Reinforced Urn Network Model. – 2016. – №. 1601.03067.
24. *Fronczak A.* Structural Hamiltonian of the international trade network //No. – 2012. – Vol. 1. – No. arXiv: 1205.4589. – pp. 31-46.
25. *Shen B., Zhang J., Zheng Q.* Exploring multi-layer flow network of international trade based on flow distances //arXiv preprint arXiv:1504.02361. – 2015.
26. *Shi P. et al.* Hierarchicality of trade flow networks reveals complexity of products //PloS one. – 2014. – Vol. 9(6). – p. e98247.
27. *Kelle M. et al.* Cross border and Foreign Affiliate Sales of Services: Evidence from German Microdata //The World Economy. – 2013. – Vol. 36(11). – pp. 1373-1392.
28. *Choquette E., Meinen P.* Export spillovers: Opening the black box //The World Economy. – 2015. – Vol. 38(12). – pp. 1912-1946.
29. *Мосейко В. О., Азмина Ю. М.* Многофакторная оценка экспортного потенциала малых и средних предприятий региона //Вестник Волгоградского государственного университета. Серия 3: Экономика. Экология. – 2012. – №. 2.
30. *Калинина Л. А.* Перспективные направления экспорта яиц и яйцепродуктов, произведенных в иркутской области //Вестник Восточно-Сибирского государственного университета технологий и управления. – 2015. – Т. 1. – №. 52. – С. 201591.
31. *Gaulier G., Zignago S.* Baci: international trade database at the product-level (the 1994-2007 version). – 2010.
32. *Solleder O. et al.* Panel Export Taxes (PET) Dataset: New Data on Export Tax Rates //Graduate Institute of International and Development Studies Working Paper. – 2013. – Vol. 7. – p. 2013.
33. *Olga Solleder.* Market access and trade – free data from ITC/UNCTAD/WTO [Электронный ресурс] / Olga Solleder. – Веб-сайт. – Режим доступа: <http://olga.solleder.org/data-on-trade-and-market-access.html>, свободный.
34. *UN Comtrade:* International Trade Statistics [Электронный ресурс] / ООН. – Веб-сайт. – Режим доступа: <http://comtrade.un.org/data/>, свободный.
35. *CHELEM – International Trade Database* [Электронный ресурс] / СЕПІІ. – Веб-сайт. – Режим доступа: <http://www.cepii.fr/%5C/anglaisgraph/bdd/chelem/internatrade/itpresent.htm>, свободный.
36. *Trade Costs Dataset* [Электронный ресурс] / Всемирный банк. – Веб-сайт. – Режим доступа: <http://data.worldbank.org/data-catalog/trade-costs-dataset>, свободный.
37. *АСД «Доступ ТСБТ»* [Электронный ресурс] / Федеральная таможенная служба. – Веб-сайт. – Режим доступа: <http://stat.customs.ru>, свободный.
38. *Trade Map – Trade Statistics for International Business Development* [Электронный ресурс] /

- Международный торговый центр ООН. – Веб-сайт. – Режим доступа: <http://trademap.org/>, свободный.
39. *Федеральный институт промышленной собственности* [Электронный ресурс] / – Веб-сайт. – Режим доступа: http://www1.fips.ru/wps/wcm/connect/content_ru/ru.
40. *Patel J. et al.* Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques // *Expert Systems with Applications*. – 2015. – Vol. 42(1). – pp. 259-268.
41. *Sheremetyeva S.* Natural language analysis of patent claims // *Proceedings of the ACL-2003 workshop on Patent corpus processing – Vol. 20. – Association for Computational Linguistics, 2003. – pp. 66-73.*
42. *Девяткин Д.А., Смирнов И.В., Соченков И.В., Тихомиров И.А.* Современные методы компьютерной лингвистики для патентного поиска и анализа информации // *Интеллектуальная собственность. Промышленная собственность. Специальный выпуск 2016*, С.71-77
43. *Dean J., Ghemawat S.* MapReduce: simplified data processing on large clusters // *Communications of the ACM*. – 2008. – Vol. 51(1). – pp. 107-113.

Суворов Роман Евгеньевич. Младший научный сотрудник ИСА ФИЦ ИУ РАН. Окончил в 2012 г. Рыбинский государственный авиационный технический университет им. П.А. Соловьева. Количество печатных работ: 22. Область научных интересов: машинное обучение, интеллектуальный анализ данных, анализ текстов на естественном языке, извлечение информации, распределенные вычисления. E-mail: rsuvorov@isa.ru

Девяткин Дмитрий Алексеевич. Младший научный сотрудник ИСА ФИЦ ИУ РАН. Окончил в 2011 г. Рыбинскую государственную авиационную технологическую академию. Количество печатных работ: 20. Область научных интересов: машинное обучение, классификация и кластеризация текстов, методы обработки больших данных, методы анализа патентной и наукометрической информации. E-mail: devyatkin@isa.ru

Усенко Наталья Ивановна. Доцент, ведущий научный сотрудник Новосибирского государственного университета. Окончила в 1976 г. Новосибирский государственный университет. Кандидат экономических наук. Количество печатных работ: 120. Область научных интересов: продовольственный рынок, анализ поведения потребителей, государственное управление. E-mail: n.i.usenko@yandex.ru

Отмахова Юлия Сергеевна. Заведующая лабораторией Новосибирского государственного университета. Окончила в 2001 г. Московский государственный университет коммерции (КФ). Кандидат экономических наук. Количество печатных работ: 80. Область научных интересов: моделирование социально-экономических процессов, системы поддержки принятия решений, продовольственный рынок. E-mail: otmakhovajs@yandex.ru

Review of Methods for Data-Driven Export Potential Analysis

Suvorov R., Devyatkin D., Usenko N., Otmakhova Ju.

Abstract. The paper presents a survey of approaches and data sources that may be useful for forecast of export growth potential of a country. We considered main studies in the area of analysis of international trade flows. The observed materials were divided into three main groups according to their intended purpose. We suggested a data-driven problem of discovering «growth points» to improve export potential of a country. Conclusions were drawn on further directions of the research using mining of the databases of international trade statistics and semantic search on collections of scientific and technical documents.

Keywords: *export growth potential, patent search, data mining, international trade, customs statistics, growth points, technologies.*

References

1. Export-import of basic commodities in January-December 2015. Official website of the Russian Customs Service / – Available at: http://www.customs.ru/index.php?option=com_content&view=article&id=22570#_ftn1, (accessed December 23, 2016).
2. Russia has overtaken the United States on the export of grain, 2016 / Available at: <http://vz.ru/news/2016/7/20/822641.html>, (accessed December 23, 2016).
3. FAOSTAT / UN Food and Agriculture Organization. – Available at: <http://faostat.fao.org/beta/en/#data> (accessed December 23, 2016).
4. *Ivanova L., Tverskaya G.* Tochki rosta i drajvery rosta: k voprosu o sodержanii ponyatij [Points of growth and drivers of growth: the question of the content of the concepts] // Zhurnal institucionalnyh issledovanij [Journal of Institutional Research]. – 2015. – Vol. 7(2). – pp. 120-133.
5. *Yoon B.* On the development of a technology intelligence tool for identifying technology opportunity // Expert Systems with Applications. – 2008. – Vol. 35(1). – pp. 124-135.
6. *Daim T. U.* et al. Forecasting emerging technologies: Use of bibliometrics and patent analysis // Technological Forecasting and Social Change. – 2006. – Vol. 73(8). – pp. 981-1012.
7. *Andersen B.* The hunt for S-shaped growth paths in technological innovation: a patent study // Journal of Evolutionary Economics. – 1999. – Vol. 9(4). – pp. 487-526.
8. *Green R. T., Allaway A. W.* Identification of export opportunities: A shift-share approach // The Journal of Marketing. – 1985. – pp. 83-88.
9. *Duenas M., Fagiolo G.* Modeling the International-Trade Network: a gravity approach // Journal of Economic Interaction and Coordination. – 2013. – Vol. 8(1). – p. 155-178.
10. *Snijders T. A. B.* Models for longitudinal network data // Models and methods in social network analysis. – 2005. – Vol. 1. – pp. 215-247.
11. *Jaud M., Kukenova M., Strieborny M.* Financial Development and Sustainable Exports: Evidence from Firm-product Data // The World Economy. – 2015. – Vol. 38(7). – pp. 1090-1114.
12. *Serdyukova Yu., Usenko N.* Strategicheskie priorityety integracionnogo vzaimodejstviya Rossii i Belarusi s pozicii obespecheniya prodovolstvennoj bezopasnosti [Strategic priorities of the integration between Russia and Belarus in terms of ensuring food security.] // Ehkonomicheskie i socialnye peremeny fakty tendencii prognoz [Economic and social changes: facts, trends and outlook]. – 2013. – No. 3 (27).
13. *Prihodchenko O.* Vyyavlenie ehksportnyh i importnyh priorityetov respubliky Belarus na osnove modelirovaniya mezhotraslevykh svyazej [Detection of export and import priorities of the Republic of Belarus on the basis of modeling of inter-branch relations] // Belorusskij zhurnal mezhdunarodnogo prava i mezhdunarodnyh otnoshenij [Belarusian Journal of International Law and International Relations]. – 2002.
14. *Zakshevskaya E., Litvinenko T.* Mirovyte tendencii v proizvodstve i sbyte myasa KRS [Global trends in the production and marketing of beef] // Mezhdunarodnyj selskohozyajstvennyj zhurnal [International Journal of Agricultural]. – 2016. – No. 5.
15. *Grater S.* et al. Linking export opportunities of products and services: the case of South Africa.
16. *Cuyvers L.* et al. A decision support model for the planning and assessment of export promotion activities by government export promotion institutions—the Belgian case // International Journal of Research in Marketing. – 1995. – Vol. 12(2). – pp. 173-186.
17. *Cuyvers L.* Identifying export opportunities: the case of Thailand // International Marketing Review. – 2004. – Vol. 21(3). – pp. 255-278.
18. *Cuyvers L., Viviers W.* (ed.). Export Promotion-A Decision Support Model Approach. – AFRICAN SUN MeDIA, 2012.

19. *Sgrignoli P.* The World Trade Web: A Multiple-Network Perspective //arXiv preprint arXiv:1409.3799. – 2014.
20. *Lall S., Weiss J., Zhang J.* The “sophistication” of exports: a new trade measure //World Development. – 2006. – Vol. 34(2). – pp. 222-237.
21. *Bernard A. B., Jensen J. B.* Why some firms export //Review of Economics and Statistics. – 2004. – Vol. 86(2). – pp. 561-569.
22. *Barigozzi M., Fagiolo G., Garlaschelli D.* Multinetwork of international trade: A commodity-specific analysis //Physical Review E. – 2010. – Vol. 81(4). – pp. 046104.
23. *Peluso S.* et al. International Trade: a Reinforced Urn Network Model. – 2016. – No. 1601.03067.
24. *Fronczak A.* Structural Hamiltonian of the international trade network //No. – 2012. – Vol. 1. – №. arXiv: 1205.4589. – pp. 31-46.
25. *Shen B., Zhang J., Zheng Q.* Exploring multi-layer flow network of international trade based on flow distances //arXiv preprint arXiv: 1504.02361. – 2015.
26. *Shi P.* et al. Hierarchicality of trade flow networks reveals complexity of products //PloS one. – 2014. – Vol. 9(6). – p. e98247.
27. *Kelle M.* et al. Cross-border and Foreign Affiliate Sales of Services: Evidence from German Microdata //The World Economy. – 2013. – Vol. 36(11). – pp. 1373-1392.
28. *Choquette E., Meinen P.* Export spillovers: Opening the black box //The World Economy. – 2015. – Vol. 38(12). – pp. 1912-1946.
29. *Mosejko V., Azmina Yu.* Mnogofaktornaya ocenka ehksportnogo potentsiala malyh i srednih predpriyatij regiona [Multifactor assessment of the export potential of small and medium-sized enterprises in the region]// Vestnik volgogradskogo gosudarstvennogo universiteta seriya 3 ehkonomika. ehkologiya [Journal of Volgograd State University. Series 3: The Economy. The Ecology.]. – 2012. – No. 2.
30. *Kalinina L.* Perspektivnye napravleniya ehksporta yaic i yajceproduktov proizvedennyh v irkutskoj oblasti [Promising directions for export eggs and egg products produced in the Irkutsk region]// Vestnik Vostochno-sibirskogo gosudarstvennogo universiteta tekhnologij i upravleniya [Journal of the East Siberian State University of Technology and Management]. – 2015. – Vol. 1(52). – pp. 201-591.
31. *Gaulier G., Zignago S.* Baci: international trade database at the product-level (the 1994-2007 version). – 2010.
32. *Solleder O.* et al. Panel Export Taxes (PET) Dataset: New Data on Export Tax Rates //Graduate Institute of International and Development Studies Working Paper. – 2013. – Vol. 7. – p. 2013.
33. *Solleder O.* Market access and trade – free data from ITC/UNCTAD/WTO / Olga Solleder. – Available at: <http://olga.solleder.org/data-on-trade-and-market-access.html>, (accessed December 23, 2016).
34. UN Comtrade: International Trade Statistics / UN. – Available at: <http://comtrade.un.org/data/>, (accessed December 23, 2016).
35. CHELEM – International Trade Database / CEPII. – Available at: <http://www.cepii.fr/%5C/anglaisgraph/bdd/chelem/internatrade/itpresent.htm> (accessed December 23, 2016).
36. Trade Costs Dataset / World Bank. – Available at: <http://data.worldbank.org/data-catalog/trade-costs-dataset>, (accessed December 23, 2016).
37. ASD «Dostup TSVT» [AAS “Access CSFT”] / Federal Customs Service. – Available at: <http://stat.customs.ru>, (accessed December 23, 2016).
38. Trade Map – Trade Statistics for International Business Development / UN International Trade Centre. – Available at: <http://trademap.org/> (accessed December 23, 2016).
39. Federal Institute of Industrial Property – Available at: http://www1.fips.ru/wps/wcm/connect/content_ru/ru (accessed December 23, 2016).
40. *Patel J.* et al. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques // Expert Systems with Applications. – 2015. – Vol. 42(1). – pp. 259-268.
41. *Sheremetyeva S.* Natural language analysis of patent claims //Proceedings of the ACL-2003 workshop on Patent corpus processing – Vol. 20. – Association for Computational Linguistics, 2003. – pp. 66-73.
42. *Devyatkin D., Smirnov I., Sochenkov I., Tikhomirov I.* Sovremennye metody kompyuternoj lingvistiki dlya patentnogo poiska i analiza informacii [Natural language processing methods for patent search and patent mining] //Intellektualnaya sobstvennost promyshlennaya sobstvennost specialnyj vypusk [Intellectual property, industrial property – special issue] 2016, p.71-77.
43. *Dean J., Ghemawat S.* MapReduce: simplified data processing on large clusters //Communications of the ACM. – 2008. – Vol. 51(1). – pp. 107-113.

Roman Suvorov. Researcher at ISA FRC CSC RAS. Graduated from Rybinsk State Aviation Technology University after Pavel Solovyov in 2012. Authored 22 scientific papers. Main research interests are machine learning, data mining, natural language processing, information extraction and distributed computations. E-mail: rsuvorov@isa.ru

Dmitry Devyatkin. Researcher at ISA FRC CSC RAS. Graduated from Rybinsk State Aviation Technology Academy after Pavel Solovyov in 2011. Authored 20 scientific papers. Main research interests are machine learning, full-text clustering, data mining, scientometrics. E-mail: devyatkin@isa.ru

Natalya Usenko. Ph.D., leading researcher at Novosibirsk State University. Graduated from Novosibirsk State University in 1976. Authored more than 120 scientific papers. Main research interests are food market, consumer behavior analysis, and public administration. E-mail: n.i.usenko@yandex.ru

Julia Otmakhova. Ph.D., head of laboratory at Novosibirsk State University. Graduated with honors from the Moscow State University of Commerce in 2001. Authored more than 80 scientific papers. Main research interests are modeling of socio-economic processes, decision support systems, food market. E-mail: otmakhovajs@yandex.ru