

# Методы и модели системного анализа

## Моделирование рационального развития Арктической зоны на основе анализа открытой информации (на примере Мурманской области РФ)\*

Ю.Э. Даник<sup>1</sup>, Д.А. Девяткин<sup>1</sup>, М.Г. Дмитриев<sup>1</sup>, М.И. Суворова<sup>1</sup>

<sup>1</sup> Институт системного анализа Федерального исследовательского центра «Информатика и управление» Российской академии наук, г. Москва, Россия

**Аннотация.** В статье описывается подход к моделированию социо-эколого-экономических процессов в Арктике и поиску рациональных решений, под которыми понимаются инвестиции в различные отрасли, на основе метода идеальной точки в пространстве критериев. Вводятся три критерия, связанные с оптимизацией динамики экономических, социальных и экологических переменных состояния. Отсутствие в полном объеме структурированных данных частично компенсируется автоматизированной обработкой открытой и неструктурированной информации. Выполнены модельные эксперименты для Мурманской области РФ, показывающие возможность использования подхода при прогнозировании основных показателей состояния экономики, природной среды и человеческого капитала.

**Ключевые слова:** инвестиции, моделирование развития, сценарий, Арктика, Мурманская область, переговоры.

**DOI:** 10.14357/20790279180405

### Введение

Имеется большое количество работ по математическому моделированию регионального развития. Первая группа связана с применением методологии факторного анализа к прогнозированию валового регионального продукта (ВРП) региона. Например, в работе [1] предложена многофакторная регрессионная модель расчета и прогноза ВРП, в которой основными факторами являются: денежные доходы населения; инвестиции в основной капитал; объем промышленной продукции; объемы продукции сельского хозяйства; объемы услуг; численность занятых. Также строятся прогнозные модели этих факторов и рассчитывается интегрированный по-

казатель устойчивого развития, основанный на динамических моделях оценки природно-ресурсного, производственного и трудового потенциалов региона. При этом в работе рассматриваются еще экологические показатели развития региона и предлагается экономический механизм финансирования природоохранных мероприятий – модель экологических сделок между предприятиями-загрязнителями.

В зарубежной литературе моделированию регионального развития с помощью многомерного статистического анализа посвящены, например, работы [2,3]. В [2] проведена оценка уровня регионального развития по экономическим, структурным и демографическим параметрам с использованием анализа ковариационных структур и факторно-аналитических моделей. В работе [3] был рассмотрен

\* Работа выполнена при финансовой поддержке РФФИ грант №15-29-06053 офи-м.

также и социальный фактор. Также можно выделить модели, основанные на статистических данных. В [4,5] представлена математическая модель экономики региона, включающая межотраслевой баланс и воспроизводственный процесс, определяемый инвестициями. Предложенная динамическая модель является развитием статистической модели «затраты — выпуск», в которую добавлен блок (подсистема) выпуска продукции конечного спроса, полученной за счет инвестиций, вкладываемых в развитие отрасли, как на уровне фирмы, так и на уровне региона, государства. Представлена реализация на тестовом примере моделирования экономики Приморского края, включающего пятнадцать отраслей на трехлетний период в соответствии с требованиями бюджетного кодекса. Результаты решения включают основные экономические показатели региона: валовой объем, валовой региональный продукт (ВРП), инвестиции (в том числе разбитые по отраслям), а также фонд заработной платы, налоги и другие.

В России тема социо-эколого-экономического моделирования регионов с позиции теории оптимального управления изучалась в работах проф. Гурмана В.И. и его учеников [6,7] при этом, одна из работ посвящена арктическим регионам РФ [8]. В них, на основе моделей оптимального управления и использования открытой структурированной региональной информации используется универсальный подход к сбору и анализу региональной информации, строятся прогнозы регионального развития.

В последние годы при моделировании социо-эколого-экономических процессов, в условиях неполных данных и нехватки некоторых взаимосвязей между ними, стали популярными модели, прототипами которых являются вычислимые модели общего равновесия (CGE), где основой поведения каждого агента являются целевые траектории, которые ориентируют действия агентов в направлении, обеспечивающем достижение поставленных целей [9]. Агент по результатам наблюдения траекторий формирует допустимые управляющие воздействия на производственный процесс, как результат решения оптимизационной задачи, в которой текущие траектории должны максимально приближаться к целевым траекториям.

Модели CGE и их применение к анализу регионального экономического развития рассмотрены также в [10–12]. Вопрос моделирования регионального развития в литературе тесно связан с экологическими моделями (см. E3MLAB (2010) <sup>1</sup>,

[13]). В последней работе оцениваются физические эффекты изменения климата на экономику.

И, наконец, отметим имитационные модели региональной динамики. В [14] приводится имитационная модель эколого-экономической системы, позволившая построить «экологическую карту» Республики Армения. Предлагаемый подход позволяет искать сценарии рациональной модернизации предприятий, являющихся основными источниками выбросов вредных веществ, с одновременным определением эффективной стратегии государственного регулирования.

Имеются примеры успешного применения математических моделей для прогноза загрязнений окружающей среды. Модельные расчеты работы [15] использовались при разработке соглашений между компаниями, в дискуссиях при подготовке Лондонской конвенции по предотвращению нефтяных загрязнений, Конвенции по морскому праву, Монреальского и Киотского протоколов, а результаты работы [16], связанные с моделью экономической эффективности добычи подводных природных ресурсов.

В тоже время, статистическая информация не всегда бывает достаточно детальной и полной для того или иного исследования. Данная проблема может быть частично решена, если обратиться к текстам из открытых источников. Анализ открытой неструктурированной информации широко используется в различных системах поддержки принятия решений и моделирования ситуаций. В работе [17] представлен подход, в котором разнородные данные интегрируются для моделирования политических процессов и процедур, связанных с реализацией государственного управления.

В [18] представлена модель развития арктического порта. В рамках этой модели в соответствии со своими целями и задачами действуют следующие группы агентов: федеральные, региональные и местные власти, администрация порта, частные и публичные компании, некоммерческие организации. Для описания текущей ситуации используется несколько групп индикаторов: физические (географическая локация), экономические, организационные (политические и административные), экологические. Для идентификации модели использовалась информация, извлеченная из таких текстовых ресурсов как новостные сайты, профильные журналы, веб-ресурсы государственных структур и компаний, а также статистические данные.

Анализ текстовых сообщений может использоваться также для учета влияния значимых событий, слабо отраженных в базах статистики, на

<sup>1</sup> [http://www.e3mlab.ntua.gr/manuals/Manual\\_of\\_GEM-E3.pdf](http://www.e3mlab.ntua.gr/manuals/Manual_of_GEM-E3.pdf)

развитие ситуации. В системе TwitterStand [19], предназначенной для сбора последних новостей по сообщениям социальных сетей, реализован простой байесовский классификатор для отделения новостных сообщений от прочей нерелевантной информации. Выявленные таким образом новости затем разбиваются на темы с помощью метрического метода кластеризации.

Для выявления финансово-экономических событий по новостным материалам используются лексико-семантические шаблоны, формируемые вручную, или с помощью методов машинного обучения с учителем [20]. В настоящей работе применен схожий подход, в котором для фильтрации текстов, релевантных экономическим или чрезвычайным событиям, применяется метод полнотекстового поиска, а для извлечения из текстов сообщений информации о финансировании используются лингвистические шаблоны.

### 1. Описание модели

Анализ и моделирование развития ситуаций в регионах Арктики предлагается вести с помощью решения многокритериальной задачи оптимального управления. Каждый агент имеет собственный набор интересов и целей в Арктике, и, следовательно, имеет представление об идеальном значении своей обобщенной цели/критерия. Действия агентов по реализации собственных интересов и целей при этом не должны наносить ущерб окружающей среде и условиям жизни коренных народов. В качестве региона моделирования рассматривается Мурманская область РФ.

**Экономический критерий.** В качестве экономических переменных состояния арктического агента в РФ, предлагается использовать объемы производства основных пятнадцати агрегированных отраслей в соответствии с общероссийским классификатором видов экономической деятельности (ОКВЭД) [ОК 029-2014]<sup>2</sup>.

В качестве единиц измерения объемов продукции каждой из 15 отраслей возьмем миллиарды рублей. Пусть  $x_j$  – стоимость объема производимой продукции  $j$ -той агрегированной отрасли, часть которого импортируется, а соответствующий

$n$ -мерный вектор-столбец  $x$  есть стоимостной вектор объема производимой продукции. В случае Мурманской области  $x_1, \dots, x_{15}$  – это стоимостные объемы производства валового регионального продукта (ВРП) по видам экономической деятельности. Продукт, относящийся к  $i$ -той агрегированной отрасли, можно разделить на четыре части. Первая часть расходуется на работу самой производственной системы, вторая идет на потребление, третья часть импортируется и четвертая идет на экспорт.

Введем вектор управления для экономического блока  $u_j, j = 1, \dots, n,$  – инвестиции в  $j$ -ую отрасль экономики в млрд. рублей.

Учитывая прирост выпуска в текущий момент времени, имеем модель динамики:

$$x(t+1) = x(t) + Bu(t), x(0) = x^0, \quad (1)$$

$$x(t) \geq 0, t = 0, 1, 2, \dots, N$$

при ограничениях межотраслевого баланса

$$x(t) = Ax(t) + w(t) + Im(t) + Exp(t), \quad (2)$$

где  $A$  – стоимостная матрица удельных затрат, коэффициенты  $a_{ji}$  которой означают долю затрат продукции  $i$ -той отрасли на производство продукции  $j$ -той отрасли стоимостью 1 млрд руб.;  $w = (w_1, w_2, \dots, w_n)^T$  – стоимостной вектор потребления продукции;  $B \in R^{n \times n}, b_{ij}$  – показывают долю прироста продукции  $i$ -той отрасли при увеличении инвестиций в  $j$ -тую отрасль на 1 млрд руб.;  $Im = (Im_1, Im_2, \dots, Im_n)^T$  – стоимостной вектор импорта продукции;  $Exp = (Exp_1, Exp_2, \dots, Exp_n)^T$  – стоимостной вектор экспорта продукции.

В качестве экономического критерия агента предлагается использовать следующий критерий:

$$I_1(u) = \frac{1}{2} \sum_{t=0}^{N-1} (x(t) - x_{зад}(t))^T Q (x(t) - x_{зад}(t)) + \quad (3)$$

$$+ u^T(t) Ru(t) \rightarrow \min_u,$$

где матрица  $Q$  – положительно полуопределенная матрица,  $R$  – положительно определенная матрица, а вектор  $x_{зад}(t), t = 0, \dots, N,$  отражает вектор целевых траекторий, как в методологии подхода CGE.

Вектор-функция инвестиций, вектор управлений  $u(t)$ , в каждый момент времени, имеет ограничения

$$u_{\min,j} \leq u_j(t) \leq u_{\max,j}, j = 1, \dots, n, \quad (4)$$

где  $u_{\max,j} = \alpha_j x_j$ , и  $\alpha_j$ , например, процент от выпуска в предыдущем году.

Отметим, что информация о верхних и нижних границах на координаты вектора управлений может зависеть от времени.

<sup>2</sup> 1.Сельское хозяйство, охота и лесное хозяйство; 2.Рыболовство, рыбоводство; 3.Добыча полезных ископаемых; 4.Обрабатывающие производства; 5.Производство и распределение электроэнергии, газа и воды; 6.Строительство; 7.Оптовая и розничная торговля, ремонт автотранспортных средств, бытовых изделий и предметов личного пользования; 8.Гостиницы и рестораны (объекты проживания и питания); 9.Транспорт и связь; 10.Финансовая деятельность; 11.Операции с недвижимым имуществом, аренда и предоставление услуг; 12.Государственное управление и обеспечение военной безопасности, социальное страхование; 13.Образование; 14.Здравоохранение и предоставление социальных услуг; 15. Предоставление прочих коммунальных, социальных и персональных услуг.

Итак, математическая модель (1) – (4) имеет вид задачи дискретного оптимального управления. Отметим, что здесь можно также использовать линейный и нелинейный варианты критерия максимизации приведенной стоимости суммарной продукции агрегированных отраслей. И подчеркиваем, что указанная дискретная модель, вообще говоря, должна содержать управления с запаздыванием, так как инвестиционные потоки с запаздыванием изменяют значения переменных состояния.

Предполагаем, что с этими же переменными состояниями можно определять управления для каждого агента и соответственно вычислять значения экономического критерия для всех агентов. Управления за прошлые периоды и значения переменных состояния, ограничения на управления – финансовая информация, которая извлекается в процессе обработки структурированной и неструктурированной информации из открытых источников. Если есть пробелы в полных статистических данных, необходимых для восстановления вектор-функций в ограничении (2), то, имея статистику по  $x(t), u(t)$ , можно пытаться определять матрицу  $B$  методом минимизации соответствующих невязок. Здесь, конечно, очень важна любая дополнительная информация, приводящая к восстановлению качественной картины взаимовлияния отраслей.

**Экологический критерий (загрязнение воздуха, отходы, загрязнение воды).** В качестве переменных состояния, отражающих экологическую ситуацию в конкретном арктическом регионе, предлагается использовать: 1. Выбросы твердых веществ в тыс. тонн. 2. Выбросы газообразных и жидких веществ (диоксида серы, оксиды азота, в пересчете на диоксид азота  $NO_2$ , оксида углерода, углеводородов без летучих органических соединений, летучих органических соединений в тыс. тонн. 3. Объем сброса сточных вод (млн кубических метров). 4. Отходы производства и потребления, млн. т. (за вычетом использованных и обезвреженных отходов).

Используя данные из Статистического ежегодника 2015<sup>3</sup> за период 2011-2015 гг. строим регрессионные зависимости переменных  $x_{16}, x_{17}, x_{18}, x_{19}$  от вектора управления, используя статистические данные об использовании природных ресурсов в регионе и соответствующих им данных о значениях компонент – вектора управления в экономическом критерии за этот же период. При этом переменные  $x_{16}(u), x_{17}(u), \dots, x_{19}(u)$  нормируем следующим образом:

$$\bar{x}_h(u) = \frac{x_h(u) - x_h^{\min}(u)}{x_h^{\max}(u) - x_h^{\min}(u)}, \quad h = 16, 17, \dots, 19,$$

где  $x_h^{\min}, x_h^{\max}$  – выбранные минимальное и максимальные значения.

Итак, экологический критерий есть функция четырех переменных состояния  $x_{16}(u), x_{17}(u), \dots, x_{19}(u)$ ,  $0 \leq x_h(t) \leq 1$ ,  $h = 16, 17, \dots, 19$ , каждую из которых, можно рассматривать в виде функций компонент – вектора управления в экономическом критерии. В итоге экологический критерий можно представить в виде:

$$I_2(u) = \sum_{t=0}^{N-1} x_{16}(u) + x_{17}(u) + \dots + x_{19}(u) \rightarrow \min_u, \quad (5)$$

где  $\sum_{j=1}^n u_j(t) \leq U_{\max}(t)$ ,  $0 \leq u_j(t)$ ,  $0 \leq x_h(t) \leq x_h^*$ ,  $h = 16, 17, \dots, 19$ ;

$U_{\max}(t)$  – предельный объем инвестиций.

Функции управляющих переменных зависят от времени и, выбирая то или иное управление, можно искать рациональное экономическое поведение с точки зрения состояния окружающей среды.

Фазовые ограничения на объемы выбросов веществ, сброса вод и образуемых отходов производства и потребления  $x_h^*$ ,  $h = 16, 17, 18, 19$  для экологических переменных состояния региона могут формироваться с помощью учета т.н. предельно допустимых концентраций. В качестве дополнительных экологических переменных можно рассматривать объемы нефтяного загрязнения, уровень радиактивного загрязнения, численности популяций видов-индикаторов арктической зоны.

**Социальный критерий** в идеале должен отражать состояние социальной инфраструктуры, уровень медицинского обеспечения, образования, правового и экономического регулирования. Здесь для простоты в качестве переменных состояния социального критерия в регионе Мурманской обл. выбираем следующие: среднедушевой денежный доход населения (руб. в месяц) и естественный прирост/убыль населения (тыс. чел.). Соответствующая статистика за период 2011-2015 гг. содержится в Статистическом ежегоднике 2015.

Как и выше, в случае, экологического критерия, здесь можно построить по имеющейся статистике зависимости, например, регрессионные, выбранных социальных показателей от управлений – инвестиционных потоков, по отраслям. Последние зависимости будут позволять получать представление социального критерия, как функции экономической активности в регионе.

<sup>3</sup> Федеральная служба государственной статистики, Территориальный орган Федеральной службы государственной статистики по Мурманской области. Мурманскстат, 2016 – 235 с.

Переменные нормируются так же, как и в случае экологического критерия. Здесь в качестве  $x_{20}^{\min}(u)$  можно взять прожиточный минимум, а  $x_{21}^{\min}(u)$  – задавать как показатель целевой минимальной величины естественного прироста.

Упрощенный социальный критерий может выглядеть так:

$$I_3(u) = -\sum_{t=0}^{N-1} x_{20}(u) + x_{21}(u) \rightarrow \min_u, \quad (6)$$

где  $\sum_{j=1}^n u_j(t) \leq U_{\max}(t), u_j(t) \geq 0,$

$x_p(t) \geq x_p^*, p = 20, 21.$

**Построение обобщенного критерия. Концептуальная модель согласования интересов агентов.**

Три приведенных критерия позволяют для конкретного арктического агента получить идеальную точку в пространстве критериев с учетом его инвестиционных возможностей и собственных представлений о траектории развития. Эти идеальные точки, естественно, теперь могут быть опорными при согласовании интересов агентов арктического сообщества. Теперь предполагаем, что переменные состояния выбраны и они одинаковы для всех агентов, т.е. будем считать, что все агенты принимают одну и ту же классификацию отраслей и управлений. Отметим, что, например, статистика в Норвегии<sup>4</sup> оперирует с 25-ю агрегированными отраслями, которые легко свести к 15. Итак, каждый из  $m$  агентов имеет набор из трех критериев –  $I_k^i(u^i), i = 1, \dots, m, k = 1, 2, 3$ . Здесь  $k=1$  отвечает экономическому,  $k=2$  – социальному,  $k=3$  – экологическому критериям,  $I_k^i(u^i)$  – значение  $k$ -го критерия  $i$ -го агента. Вектор управления  $u^i$   $i$ -го агента имеет размерность  $n=15$  и определяется следующим образом:  $u^i = (u_1^i, \dots, u_n^i)^T$ , где  $u_j^i, j = 1, n, i = 1, m$  –  $j$ -ая координата вектора управления  $i$ -го агента, а именно, вложения в  $j$ -ую отрасль  $i$ -го агента. Вектор состояния  $i$ -го агента имеет размерность  $l=21$  и определяется следующим образом:  $x^i = (x_1^i, \dots, x_l^i)^T$ , где  $x_s^i, s = 1, l, i = 1, m$  –  $s$ -ая координата вектора состояния  $i$ -го агента. Далее последовательно, для каждого из агентов отдельно, оптимизируются значения всех трех критериев с учетом введенных ограничений и находятся  $I_k^{i \min}, k = 1, 2, 3$ . Идеальная точка агента в пространстве экономического и социального критериев отражает наилучшие значения соответствующих критериев

$$I^{i*}(u) = \begin{pmatrix} I_1^{i \min}(u) \\ I_2^{i \min}(u) \end{pmatrix} \text{ для } i\text{-го агента. Для согласо-}$$

вания развития отдельных стран, регионов (агентов и сообществ агентов) в Арктике идеально было бы вести согласованную инвестиционную экономическую политику. В качестве связующего критерия для всех агентов вводится общий экологический критерий  $I_e$  как взвешенная сумма квадратичных отклонений текущих значений индивидуальных экологических критериев агентов от их идеальных значений, где весовые коэффициенты

$$\beta_i > 0, \sum_{i=1}^m \beta_i = 1, \text{ т.е. получаем задачу минимизации } I_e = \sum_{t=0}^{N-1} \sum_{i=1}^m \beta_i (I_3^i(u^i) - I_3^{i \min})^2 \rightarrow \min_{u^1, \dots, u^i, \dots, u^m},$$

где  $I_3^{i \min}$  – идеальное значение экологического критерия  $i$ -го агента, а  $I_3^i(u^i)$  – текущее значение этого критерия.

В итоге, с учетом  $I^{i*}(u)$  и  $I_e$ , получаем следующую задачу оптимального управления:

$$I(u) = \sum_{t=0}^{N-1} \sum_{i=1}^m (I^{i*}(u) - I^i(u))^T \quad (7)$$

$$D^i (I^{i*}(u) - I^i(u)) \rightarrow \min_{u^1, \dots, u^i, \dots, u^m},$$

$$\text{где } I^i(u) = \begin{pmatrix} I_1^i(u) \\ I_2^i(u) \\ I_3^i(u) \end{pmatrix}, \quad D^i = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}$$

где  $D^i \in R^{3 \times 3}$  – весовая матрица,  $0 < \gamma_i < \beta_i$  – коэффициент, устанавливающий превосходство экологического критерия над другими. Цель заключается в минимизации значения критерия (6) по  $u^1, \dots, u^i, \dots, u^m$  с учетом ограничений  $u^i \in U^i, x^i(u^i) \in G^i$ , где  $U^i, G^i$  – наборы допустимых значений переменных управления и состояния для каждого агента. Результатом оптимизации является вектор инвестиций  $u^* = (u^1, \dots, u^i, \dots, u^m)$  и соответствующие ему

$$\text{вектора } x^{i*}(u^{i*}) = \begin{pmatrix} x_1^i(u^{i*}) \\ \vdots \\ x_l^i(u^{i*}) \end{pmatrix}, i = \overline{1, m} \text{ в про-}$$

странстве переменных состояния. Полученная задача решается сведением ее к задаче нелинейного программирования, т.к. критерий (6) можно считать функцией координат вектора управления.

Присутствие связующего критерия позволяет формировать согласованную точку зрения на направление дальнейшего развития рассматриваемой арктической территории, и надеяться на возможность достижения некоего компромиссного рацио-

<sup>4</sup> [https://www.ssb.no/nasjonalregnskap-og-konjunkturer/oa/\\_attachment/101117?\\_ts=13d4491c260](https://www.ssb.no/nasjonalregnskap-og-konjunkturer/oa/_attachment/101117?_ts=13d4491c260).

нального решения. Полученные из (6) конкретные значения всех компонент векторов  $x^1, \dots, x^i, \dots, x^m$  являются квотами-ограничениями агентов в динамике для этих переменных состояния, а задача поиска траекторий развития каждого агента сводится к задаче определения инвестиционных программ  $u^1, \dots, u^i, \dots, u^m$ , приближающих сообщество агентов к согласованному оптимуму.

Применение данного подхода к одному региону арктической зоны позволяет построить стратегию развития региона, ориентированную на сбалансированное развитие с просмотром различных сценариев развития, определяемых теми или иными целевыми установками и проверкой этих сценариев на реализуемость, что иллюстрируется приводимыми ниже расчетами на примере Мурманской области.

## 2. Извлечение информации из неструктурированных источников

Описанный подход требует наличия разнообразной информации, которая может отсутствовать в официальной статистике, либо быть недостаточно актуальной и детализированной. Поэтому, помимо структурированной информации, предлагается обрабатывать тексты на естественном языке, применяя метод автоматизированного оценивания ситуации, основанный на полнотекстовом семантическом поиске [21].

Для проведения экспериментов использовался корпус текстов об Арктике из разнообразных открытых источников информации, описанный в [21,22]. Корпус содержит тексты, относящиеся ко всей Арктической зоне, однако в рамках исследования рассматривались только сообщения, относящиеся к региону Баренцева моря и Мурманской области. Фильтрация корпуса производилась следующим образом:

1. Из ресурса [wikidata](https://www.wikidata.org/w/api.php)<sup>5</sup> загружались наименования географических локаций находящихся в Мурманской области.
2. Для извлеченных наименований по формуле *tf-idf* вычислялись их веса в сообщениях. Наименования с весом ниже определенного порога, т.е. составленные только из общеупотребимой лексики, были отфильтрованы (Большая, Восточный, Белая, Черная, Широкая).
3. Выполнялся поиск [23] упоминаний Баренцева моря и оставшихся наименований локаций Мурманской области в экспериментальном корпусе текстов. Тексты, не содержащие этих упоминаний, отфильтровывались из корпуса.

Далее оставшиеся материалы распределя-

лись согласно отраслям экономики, которым они посвящены. Отрасли определялись в соответствии с Общероссийским классификатором видов экономической деятельности (ОКВЭД 2014).

Для классификации сообщений использовались слова и словосочетания, извлеченные вручную из описаний отраслей в ОКВЭД. К сожалению, в исходном виде эта лексика не позволила выполнить классификацию с удовлетворительной точностью и полнотой, поэтому была выполнена предобработка по следующему алгоритму:

1. Выполнить синтаксический анализ [24] текстовых описаний для каждой отрасли и построить деревья зависимостей. Примененный анализатор был обучен на размеченном корпусе SynTagRus [25].
2. Из полученных синтаксических деревьев зависимостей извлечь списки именных групп длиной до двух слов.
3. Для извлеченных именных групп вычислить веса по формуле *tf-idf*. Отсеять именные группы с весом ниже определенного порога, т.е. общеупотребимые словосочетания.

После формирования списков слов и словосочетаний, релевантных отраслям, производилась метрическая классификация сообщений экспериментального корпуса. Для этого для сообщений корпуса были автоматически построены ключевые слова [26]. В качестве фоновой коллекции при построении ключевой лексики сообщений использовался исходный, неотфильтрованный корпус статей. Для каждой отрасли статьи были ранжированы по степени сходства ключевых слов и словосочетаний сообщений с ключевой лексикой отраслей. Перед сопоставлением была выполнена нормализация именных групп (слова именных групп приведены к нормальной форме) [27]. В качестве меры при сопоставлении использовалось косинусное расстояние.

Для извлечения информации о финансировании отраслей был доработан подход, представленный ранее в [22]. Этот подход позволяет выявлять информацию о самом факте вложения средств, сумме инвестирования, организации-инвесторе и географической локации, в которой расположен объект финансирования.

Для построения связей между выявленными суммами организациями и локациями, а также выявления их ролей в процессе финансирования использовались результаты семантического анализа текстов сообщений [28]. На рис. 1 представлена предикатно-аргументная структура, построенная для одного и того же высказывания, содержащего информацию о финансировании на русском и английском язы-

<sup>5</sup> <https://www.wikidata.org/w/api.php>

Табл. 1

Коэффициенты линейных отраслевых инвестиционных программ

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a	0	0.593	3.206	2.062	0.522	0	0.585	0.068	1.523	0	-0.047	3.601	0	0.116	1.32
b	0.18	2.202	17.607	2.374	3.24	0.18	0.18	0.09	7.584	0.18	1.854	6.789	0.63	0.93	2.851

Табл. 2

Коэффициенты регрессионных зависимостей

	d	c <sub>1</sub>	c <sub>2</sub>	c <sub>3</sub>	c <sub>4</sub>	c <sub>5</sub>	c <sub>6</sub>	c <sub>7</sub>	c <sub>8</sub>	c <sub>9</sub>	c <sub>10</sub>	c <sub>11</sub>	c <sub>12</sub>	c <sub>13</sub>	c <sub>14</sub>	c <sub>15</sub>
x <sub>16</sub>	0.25	0.25	-0.025	1.6·e-3	-5.3·10 <sup>-4</sup>	-0.017	-1.3·10 <sup>-3</sup>	-0.01	-0.012	0.025	-1.5·10 <sup>-3</sup>	-2.5·10 <sup>-3</sup>	-5.236·10 <sup>-3</sup>	-2.4·10 <sup>-3</sup>	-4.9·10 <sup>-3</sup>	-3·10 <sup>-3</sup>
x <sub>17</sub>	0.499	-2.1·10 <sup>-3</sup>	0.235	-6.2·10 <sup>-3</sup>	0.017	-0.155	5.3·10 <sup>-3</sup>	-0.081	-0.156	0.049	-8.9·10 <sup>-3</sup>	0.035	-0.041	-0.021	-0.041	0.027
x <sub>18</sub>	0.412	0.253	-0.28	-0.036	0.314	-0.039	0.247	0.123	0.493	0.058	0.405	0.222	-0.149	0.211	0.166	0.048
x <sub>19</sub>	0.409	0.253	-0.286	-0.035	0.315	-0.039	0.246	0.121	0.481	0.056	0.405	0.234	-0.154	0.211	0.165	0.048
x <sub>20</sub>	0.405	0.253	-0.275	-0.047	0.319	-0.04	0.246	0.123	0.484	0.037	0.406	0.237	-0.145	0.211	0.165	0.049
x <sub>21</sub>	0.406	0.25	-0.228	-0.05	0.336	-7.1·10 <sup>-3</sup>	0.301	0.141	0.412	0.016	0.422	0.245	-0.162	0.215	0.209	0.083

ках. Необходимо отметить, что семантические роли аргументов при предикатном слове, такие как «Донор» – источник финансирования, «Дестинатив» – объект назначения и «Поссесив» – объект владения (как правило, сумма финансирования) могут быть использованы для определения ролей извлеченных сущностей в процессе финансирования. В ходе дальнейших исследований для связи сумм, организаций, локаций и целей, упоминаемых в разных предложениях текста, планируется также использование коррелентных связей.



Рис. 1. Результаты семантического анализа сообщений о финансировании на русском и английском языках

Информация о финансировании отраслей, извлеченная из неструктурированных источников, использовалась далее в ходе численного эксперимента для оценки нижней границы координат вектора управлений модели (1) – (4):  $u_{min,j} \leq u_j(t), j = 1, \dots, n$ .

### 3. Численный эксперимент

Расчеты выполнялись для региона Мурманской обл., т.е.  $m=1$  в среде MathCad 15.0. При этом прогнозный период равнялся пяти годам ( $N=5$ ) с 2016 по 2020 г., нормированные значения переменных состояния были в пределах отрезка  $[0,1]$ . Матрица Q в (3) выбиралась диагональной, где все элементы равны 1, кроме элементов для отраслей 3,4,7,9,12, для которых значения равны 100,10,10,100 и 10, соответственно.

Рассматривался сценарий, где имеется рост ВРП на 3% ежегодно, и, соответственно, вычислялся вектор целевых траекторий  $x_{зад}(t)$  по отраслям. Считая, что информация по межотраслевому балансу отсутствует, в расчете на основе статистики инвестиций за предыдущие пять лет (2011-2015 гг.), определялись с помощью минимизации суммы квадратов невязок в (1) коэффициенты матрицы B, считая, что диагональные элементы больше нуля, а половина суммы внедиагональных элементов в каждой строке матрицы B меньше ее соответствующего диагонального элемента. Найденная матрица B достаточно хорошо воспроизводит отсутствующий при расчетах механизм взаимодействия отраслей, расчетный выпуск всех отраслей при сравнении с известными данными 2016 г. Т.е. полученная модель (1) качественно воспроизводит модель (1), (2) без учета (2). Для уменьшения числа неизвестных в соответствующей задаче нелинейного программирования вектор инвестиций формировался в виде  $u(t) = at + b$  и начальное приближение для методов локального спуска формировалось, предполагая вектора a и b единичными. Т.е. для каждой отрасли инвестиции искались в виде линейных функций и в итоге оптимизации регрессионные зависимости переменных  $x_{16}(u), x_{17}(u), \dots, x_{21}(u)$  получились следующими  $x_j(u) = cu + d, j = 16, \dots, 21, c \in R^{1 \times 15}$  (см. Табл. 2).

Из полученных траекторий переменных состояния можно наблюдать снижение выпуска в отраслях: сельское хозяйство, охота и лесное хозяйство; финансовая деятельность. Также имеется область временного снижения в отрасли – производство и распределение электроэнергии, газа и воды. При этом особенно быстро растут отрасли: рыболовство, рыбководство; строительство; государственное управление и обеспечение военной

безопасности; социальное страхование; образование; здравоохранение и предоставление социальных услуг; предоставление прочих коммунальных, социальных и персональных услуг. Это происходит на фоне годовых инвестиций с неизменным темпом в следующие отрасли: добыча полезных ископаемых; оптовая и розничная торговля; ремонт автотранспортных средств, бытовых изделий и предметов личного пользования; гостиницы и рестораны; здравоохранение и предоставление социальных услуг, а в отраслях: с/х, охота и лесное хозяйство; производство и распределение электроэнергии, газа и воды; строительство; финансовая деятельность; операции с недвижимым имуществом, аренда и предоставление услуг; государственное управление и обеспечение военной безопасности; социальное страхование; образование – на фоне инвестиций с повышенным темпом.

Найденные необходимые инвестиции по этому сценарию приводят к росту населения в регионе на 40% и повышению за пять лет среднедушевого дохода в 1,5 раза. Но часть расчетных показателей загрязнений при этом значительно растут (объемы сброса сточных вод, отходов производства и потребления), а часть – выбросы твердых, газообразных и жидких веществ в атмосферу – незначительно падают.

### Заключение

В статье представлен подход к моделированию социо-эколого-экономических процессов в Арктике и поиску рациональных решений, под которыми понимаются инвестиции в различные отрасли. Вводятся три критерия, связанные с оптимизацией динамики экономических, социальных и экологических переменных состояния. Отсутствие в полном объеме структурированных данных частично компенсируется автоматизированной обработкой открытой и неструктурированной информации. Выполнены модельные эксперименты для Мурманской области РФ, показывающие возможность использования подхода при прогнозировании основных показателей состояния экономики, природной среды и человеческого капитала, а также для проверки реализуемости сценариев развития при различных инвестиционных стратегиях.

### Литература

1. *Игнатьева О.В.* Разработка концептуальной модели устойчивого развития региона (на материалах Восточно-Казахстанской области) [Текст]: автореф. дис. ... канд. эконом. наук (27.11.2010) / Игнатьева Оксана Викторовна; Университете «Туран». Алматы, 2010.
2. *Cziraky D. et al.* A multivariate methodology for modelling regional development in Croatia // *Croatian International Relations Review*. 2002. Т. 8. №. 26/27. P. 35-52.
3. *Cziraky D. et al.* Regional development assessment: A structural equation approach // *European Journal of Operational Research*. 2006. Т. 174. №. 1. P. 427-442.
4. *Машунин Ю.К., Машунин И.А.* Прогнозирование развития экономики региона с использованием таблиц «Затраты выпуск» // *Экономика региона*. 2014. №. 2 (38). С. 276-289.
5. *Машунин И.А., Машунин Ю.К.* Организация управления, моделирование и прогнозирование развития экономики региона // *Региональная экономика и управление: электронный научный журнал*. 2016. № 1 (45). С. 29-58.
6. *Гурман В.И. и др.* Сценарные расчеты стратегий развития региона // *Вестник Бурятского государственного университета. Экономика и менеджмент*. 2014. №. 1. С. 60-73.
7. *Гурман В.И. и др.* Использование социо-эколого-экономической модели региона для оценки эффективности инвестиционных проектов // *Проблемы и перспективы развития социально-экономического потенциала российских регионов*. 2016. С. 377-384.
8. *Гурман В.И. и др.* Возможности применения математических моделей и методов к исследованию проблем устойчивого развития регионов на примере арктической зоны // *Программные системы: теория и приложения*. 2016. Т. 7. №. 2. С. 105-125.
9. *Цыбатов В.А.* Стратегирование регионального развития: методы, модели, информационные технологии // *Региональная экономика: теория и практика*. 2015. № 27 (402). С. 36-52.
10. *Partridge M.D., Rickman D.S.* Computable general equilibrium (CGE) modelling for regional economic development analysis // *Regional Studies*. 2010. Т. 44. №. 10. P. 1311-1328.
11. *Allan G.J. et al.* Computable General Equilibrium Modelling in Regional Science // *Regional Research Frontiers-Vol. 2.* – Springer International Publishing, 2017. P. 59-78.
12. *Taylor L.* CGE applications in development economics // *Journal of Policy Modeling*. 2016. Т. 38. №. 3. P. 495-514.
13. *Ciscar J.C. et al.* The integration of PESETA sectoral economic impacts into the GEM-E3 Europe model: methodology and results // *Climatic Change*. 2012. Т. 112. №. 1. P. 127-142.

14. *Акопов А.С. и др.* Моделирование региональной эколого-экономической системы с механизмом государственного регулирования на примере Республики Армения // *Экономическая наука современной России*. 2016. №. 1. С. 109-119.
15. *Antrim C.* Converting Competition to Collaboration: Creative Applications of Models in the Law of the Sea Negotiations. In *Diplomacy Games. Formal Models and International Negotiations*. Edited by Rudolf Avenhaus & I. William Zartman. Springer 2007.
16. *Nyhart J.D., Antrim L., Capstaff A., Kohler A., Leshaw D.* A cost model of deep ocean mining and associated regulatory issues. MIT Sea Grant Report, 1978, 78-4.
17. *Rumm N., Ortner B., Löw H.* Approaches to integrate various technologies for policy modeling // *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*. – 2014.
18. *Gritsenko D., Efimova E.* Policy environment analysis for Arctic seaport development: the case of Sabetta (Russia) // *Polar Geography*. 2017. P. 1-22.
19. *Sankaranarayanan J. et al.* Twitterstand: news in tweets // *Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems*. ACM, 2009. P. 42-51.
20. *Hogenboom F.* Automated detection of financial events in news text. – 2014. – №. EPS-2014-326-LIS.
21. *Osipov G.S. et al.* The concept of the decision support system for international negotiations in the arctic region // *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*. 2016. Т. 1. P. 461-467.
22. *Ананьева М.И. и др.* Автоматическое извлечение финансово-экономической информации из текстов на русском языке // *Труды Института системного анализа Российской академии наук*. 2018. Т. 68. №. 1. С. 23-30.
23. *Osipov G. et al.* Relational-situational method for intelligent search and analysis of scientific publications // *Proceedings of the Integrating IR Technologies for Professional Search Workshop*. 2013. P. 57-64.
24. *Nivre J. et al.* MaltParser: A language-independent system for data-driven dependency parsing // *Natural Language Engineering*. 2007. Т. 13. №. 2. P. 95-135.
25. *Дяченко П.В. и др.* Современное состояние глубоко аннотированного корпуса текстов русского языка (СинТагРус) // *Сборник «Национальный корпус русского языка: 10 лет проекту»*. Труды Института русского языка им. В.В. Виноградова. М., 2015. Вып. 6. С. 272-299.
26. *Девяткин Д.А., Суворов Р.Е., Соценков И.В.* Метод тематической кластеризации масштабных коллекций научно-технических документов // *Информационные технологии и вычислительные системы*. 2013. №. 1. С. 33.
27. *Сокирко А.В.* Морфологические модули на сайте [www.aot.ru/](http://www.aot.ru/) // *Диалог-2004. Компьютерная лингвистика и интеллектуальные технологии: Труды междунар. конф.* – 2004.
28. *Shelmanov D.* Devyatkin Semantic Role Labeling with Neural Networks for Texts in Russian // *Annual International Conference “Dialog”*, Vol. 1, pp. 245-257, 2017.

**Даник Юлия Эдуардовна.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН. Младший научный сотрудник. Количество печатных работ: 26. Область научных интересов: системный анализ, теория управления, нелинейный синтез, Паде аппроксимации. E-mail: [yuliadanik@gmail.com](mailto:yuliadanik@gmail.com)

**Девяткин Дмитрий Алексеевич.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН. Главный специалист. Количество печатных работ: 30. Область научных интересов: машинное обучение, классификация и кластеризация текстов, методы извлечения информации. E-mail: [devyatkin@isa.ru](mailto:devyatkin@isa.ru)

**Дмитриев Михаил Геннадьевич.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН. Главный научный сотрудник. Количество печатных работ: свыше 250. Область научных интересов: асимптотические методы в теории оптимального управления, нелинейный синтез, моделирование социально-экономических систем. E-mail: [mdmitriev@mail.ru](mailto:mdmitriev@mail.ru)

**Суворова Маргарита Игоревна.** Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН. Главный специалист. Количество печатных работ: 16. Область научных интересов: компьютерная лингвистика, дискурсивный анализ текстов, корпусная лингвистика. E-mail: [suvorova@isa.ru](mailto:suvorova@isa.ru)

## Modelling of rational development of the Arctic using open information (on the example of the Murmansk region)

Y.E. Danik<sup>1</sup>, D.A. Devyatkin<sup>1</sup>, M.G. Dmitriev<sup>1</sup>, M.I. Suvorova<sup>1</sup>

<sup>1</sup> Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

**Abstract.** The article describes an approach to the modeling of socio-ecological and economic processes in the Arctic and the support of decision making, which means the decision to invest in one or the other industry. Such an approach is based on the ideal point method in the space of criteria. Three criteria are introduced. They are related to the optimization of the dynamics of economic, social and environmental state variables. Absence of some structured data can be partly compensated for by automated processing of unstructured data. Model experiments were performed for the Murmansk region of the Russian Federation. Their results showed that the approach can be applied to forecasting the state of the economy, the natural environment and human resources.

**Keywords:** *sinvestment, development modelling, scenario, the Arctic, Murmansk region.*

**DOI:** 10.14357/20790279180405

### References

1. *Ignatieva O.V.* 2010. Razrabotka kontseptual'noy modeli ustoychivogo razvitiya regiona (na materialakh Vostochno-Kazakhstanskoy oblasti) [Development of a conceptual model of sustainable development of the region (on materials of the East Kazakhstan region)] PhD thesis. Almaty: University "Turan".
2. *Cziraky D. et al.* 2002. A multivariate methodology for modelling regional development in Croatia. *Croatian International Relations Review*, 8(26/27), pp.35-52.
3. *Cziraky D. et al.* 2006. Regional development assessment: A structural equation approach. *European Journal of Operational Research*, 174(1), pp.427-442.
4. *Mashunin Yu.K., Mashunin I.A.* 2014. Forecasting the development of regional economy on the basis of input – output tables. *Economy of Region*. № 2 (38), pp. 276-289.
5. *Mashunin I.A. and Mashunin Yu.K.* 2016. Organization of management, modeling and forecasting of development of economy of the region. *Regional economics and management: electronic scientific journal*. № 1 (45), pp. 29-58.
6. *Gurman V.I. et al.* 2014. Scenario calculations of regional development strategies. *Bulletin of the Buryat State University. Economics and management*. №. 1. pp. 60-73.
7. *Gurman V.I. et al.* 2016. Kontseptual'naya osnova razrabotki kompleksa sotsio-ekologo-ekonomicheskikh modeley regiona [Conceptual basis for the development of the complex of socio-ecological and economic models of the region]. *Economics of nature management*, (4), pp.44-52.
8. *Gurman V.I. et al.* 2016. Possibilities of mathematical models and methods application to study of regional sustainable development problems on the case of the Arctic zone Program Systems: Theory and Applications, 7 (2 (29)). pp. 105-125.
9. *Tsybatov V.A.* 2015. Strategic planning of regional development: methods, models, information technology *Regional Economics: Theory and Practice*. № 27 (402). C. 36-52.
10. *Partridge M.D. and Rickman D.S.* 2010. Computable general equilibrium (CGE) modelling for regional economic development analysis. *Regional studies*, 44(10), pp.1311-1328.
11. *Allan G.J. et al.* 2017. Computable General Equilibrium Modelling in Regional Science. In *Regional Research Frontiers-Vol. 2* (pp. 59-78). Springer, Cham.
12. *Taylor L.* 2016. CGE applications in development economics. *Journal of Policy Modeling*, 38(3), pp.495-514.
13. *Ciscar J.C. et al.* 2012. The integration of PESETA sectoral economic impacts into the GEM-E3 Europe model: methodology and results. *Climatic Change*, 112(1), pp.127-142.
14. *Akopov A.S. et al.* 2016. Modelling the Regional Ecological-Economic System with the Mechanism of the Government Regulation for the Case-Study of the Republic of Armenia. *Economics of Contemporary Russia*. № 1 (72). pp.109-119.
15. *Antrim C.L.* 2007. Converting Competition to Collaboration: Creative Applications of Models in the Law of the Sea Negotiations. In *Diplomacy Games* (pp. 211-228). Springer, Berlin, Heidelberg.
16. *Nyhart J.D. et al.* 1978. A cost model of deep ocean mining and associated regulatory issues.
17. *Rumm N., Ortner B. and Löw H.* 2014. Approaches to integrate various technologies for policy modeling. In *Handbook of Research on Advanced*

- ICT Integration for Governance and Policy Modeling (pp. 272-295). IGI Global.
18. *Gritsenko D. and Efimova E.* 2017. Policy environment analysis for Arctic seaport development: the case of Sabetta (Russia). *Polar Geography*, 40(3), pp.186-207.
  19. *Sankaranarayanan J. et al.* 2009, November. Twitterstand: news in tweets. In Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems(pp. 42-51). ACM.
  20. *Hogenboom F.* 2014. Automated detection of financial events in news text (No. EPS-2014-326-LIS).
  21. *Osipov G.S. et al.* 2016. The concept of the decision support system for international negotiations in the Arctic region. *International Multidisciplinary Scientific GeoConference: SGEM: Surveying Geology & mining Ecology Management*, 1, pp.461-467.
  22. *Ananyeva M.I. et al.* 2018. Extraction of financial and economic information from texts in Russian Proceedings of the Institute for Systems Analysis of the Russian Academy of Sciences, 68 (1), pp. 23-30.
  23. *Osipov G. et al.* 2013. March. Relational-situational method for intelligent search and analysis of scientific publications. In Proceedings of the Integrating IR Technologies for Professional Search Workshop (pp. 57-64).
  24. *Nivre J. et al.* 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2), pp.95-135.
  25. *Dyachenko P.V. et al.* 2015. A deeply annotated corpus of Russian texts (SynTagRus): contemporary state of affairs. Proceedings of the V.V. Vinogradov Russian Language Institute, (6), pp.272-300.
  26. *Devyatkin D.A., Suvorov R.E. and Sochenkov I.V.* 2013. A method for topic clustering for large science publication collections. *Information Technologies and Computing Systems*, (1), p.33.
  27. *Sokirko A.V.* 2004. Morfologicheskiye moduli na sayte www.aot.ru [Morphological modules on the site www.aot.ru]. In Proceedings of the conference "Dialogue-2004".
  28. *Shelmanov A., Devyatkin D.* 2017. Semantic Role Labeling with Neural Networks for Texts in Russian. Annual International Conference "Dialog". (1), pp. 245-257.

**Danik Yulia.** Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (FRC CSC RAS). Junior researcher. Number of publications – 26. Research interests: system analysis, control theory, nonlinear synthesis, Padé approximation. E-mail: yuliadanik@gmail.com

**Devyatkin Dmitry.** Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (FRC CSC RAS). Chief Specialist. Number of publications – 30. Research interests: machine learning, classification and clustering of texts, methods of information retrieval. E-mail: devyatkin@isa.ru

**Dmitriev Mikhail.** Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (FRC CSC RAS). Chief Researcher. The number of publications is more than 250. Research interests: asymptotic methods in the theory of optimal control, nonlinear synthesis, modeling of socio-economic systems. E-mail: mdmitriev@mail.ru

**Suvorova Margarita.** Institute for Systems Analysis, Federal Research Center "Computer Science and Control" of Russian Academy of Sciences (FRC CSC RAS). Chief Specialist. Number of publications – 16. Research interests: computer linguistics, discourse analysis of texts, corpus linguistics. E-mail: suvorova@isa.ru