

УДК 303.732.4:004.6:303.094.6
ББК

О ПРОГНОСТИЧЕСКИХ МЕТОДАХ, ИСПОЛЬЗУЮЩИХ ВЫСОКИЕ УРОВНИ ПОНИМАНИЯ ИНФОРМАЦИИ

**Аверин Г.В.¹, Константинов И.С.², Звягинцева А.В.³,
Иващук О.А.⁴**

*(Белгородский государственный национальный
исследовательский университет, Белгород)*

Показано, что в основу методов описания данных о состоянии и развитии общественных систем может быть положен принцип многомерного континуального представления количественной информации и принцип феноменологического описания полей вероятностей совместных событий, отражающих состояния объектов прогнозирования. Предлагаемые прогностические методы основаны на вероятностных моделях описания данных, использующих высокие уровни понимания информации. Данные методы предполагают применение знаний о феноменологических закономерностях развития объектов. Предложены зависимости для построения моделей описания данных в многомерных информационных пространствах и рассмотрены примеры создания таких моделей для некоторых классов социально-экономических систем.

Ключевые слова: прогностические методы, интеллектуальный анализ данных.

¹ Геннадий Викторович Аверин, доктор технических наук, профессор (averin@bsu.edu.ru).

² Игорь Сергеевич Константинов, доктор технических наук, профессор (konstantinov@bsu.edu.ru).

³ Анна Викторовна Звягинцева, кандидат технических наук, доцент (zviagintseva@bsu.edu.ru, anna_zv@ukr.net).

⁴ Ольга Александровна Иващук, доктор технических наук, доцент (ivaschuk@bsu.edu.ru).

1. Введение

Современный мир вошел в эпоху политической нестабильности, локальных войн, экономических кризисов и борьбы за ресурсы, поэтому стратегическое прогнозирование и планирование развития любой страны становится жизненно необходимым фактом. Протесты, цветные революции, внешние воздействия и смена власти в постсоветских странах и арабском мире говорят о том, что мир управляем, и любая страна может стать объектом масштабного политического эксперимента. Последнее время подобные методы и средства воздействий на государства и другие государственные и общественные образования получили название прогностических вооружений. Очевидно, что новые методы ведения информационных и экономических войн разрабатываются и апробируются на практике ведущими разведывательными и аналитическими службами ряда стран мира. В прогностическом аспекте эти методы тесно связаны с задачами интегрирования (data fusion) различных баз данных и их использования, экспертными технологиями, применением методов интеллектуального анализа данных (ИАД) и имитационным моделированием возможных сценариев развития событий.

Исторически так сложилось, что стратегическое прогнозирование в своей базовой методологии опирается преимущественно на экспертные методы, которые обладают многими недостатками. Однако, сегодня перспективный путь развития прогностических методов связан, в первую очередь, с развитием новых подходов к моделированию общественных процессов и объектов, использующих естественнонаучную методологию. Поэтому, одна из критических технологий, которая может возникнуть в будущем и позволит обеспечить получение качественно новых результатов в области национальной безопасности, стратегической разведки и в социально-экономической сфере — это создание математических методов и средств предсказания состояния и развития общественных систем при высокой степени достоверности прогнозов. Важным этапом на пути создания таких технологий является разработка методов, которые позволяют формировать представления о возможных путях эволюционного развития стран и регионов на среднесрочную и долго-

срочную перспективу, исходя из анализа множества показателей, изучения групповых закономерностей развития и с учетом возможных негативных воздействий на объекты прогнозирования. В целом, это не простая задача, решение которой может быть найдено на основе применения методов анализа и описания данных, которые могут быть реализованы на высоких уровнях понимания информации и позволяют учитывать закономерности развития многомерных систем. Цель таких исследований заключается в том, чтобы найти пути разработки наиболее вероятных прогнозов для стратегического планирования при существующих ограничениях и воздействиях – прогнозы, которые, по словам американского социолога У. Томаса, «формируют реальность» и относятся к так называемым «самоисполняющимся пророчествам».

Во многих странах процесс интеграции баз данных в различных сферах деятельности общества и поиск методов их анализа идет уже полным ходом. На этой основе с середины 90-х годов Агентство национальной безопасности США и разведывательные структуры ЦРУ разрабатывали модели экономических войн. В 2001 году военными США было признано, что «Сегодня актуальные и точные прогнозы в экономике имеют такое же военное значение, как разведывательные данные о состоянии войск нашего противника». Считают, что с этого момента начались исследования, связанные с разработками прогностического оружия. В 2012 году правительство США потратило на эти цели 4,9 млрд. долларов, в 2017 году на разработку средств прогнозирования по большим данным планируется потратить уже 7,2 млрд. долларов. Программы по созданию и развитию технологий анализа и прогнозирования для ключевых направлений государственной политики существуют в Японии, Австралии, Великобритании и т.д.

Сегодня уже очевидно, что создание новых прогностических методов и систем достаточно актуально, т.к. они позволят оценивать тенденции развития и осуществлять поддержку принятия решений, направленных на среднесрочную и долгосрочную перспективу в области развития государства и состояния национальной безопасности. В основе использования подобных информационно-аналитических средств должно лежать качест-

венно новое математическое обеспечение, направленное на решение задач прогнозирования и управления большими системами. Для создания подобного математического обеспечения предлагается новый подход, в основу которого положены методы, использующие высокие уровни понимания информации. Данные методы предполагают формирование знаний о феноменологических закономерностях изменения и развития объектов прогнозирования. Причем подход отличается универсальностью, т.к. применяется к данным, представленным в определенной форме, не зависимо от природы изучаемых объектов.

Целью работы является разработка новых прогностических методов, основанных на использовании феноменологических моделей описания данных о состоянии и развитии сложных систем.

2. Тенденции в области развития прогностических методов и прогностических средств и систем

За последние десятилетия наукой было рождено множество методов и технологий прогнозирования и комплексной оценки: классическая прогностика, модели системной динамики, имитационно-прогностические компьютерные модели, технологическое предвидение, форсайт, комплексная индикативная оценка, циклическое прогнозирование и прогнозирование по критериям стратегических рисков и т.д. [2, 6, 9, 10, 11, 13, 14]. Большинство этих технологий прогнозирования относятся к классу экспертных методов или требуют построения гипотетических моделей, которые формулируются экспертами. Однако, указанные методы неохотно используются практиками. Преобладающая ошибочность среднесрочных и долгосрочных прогнозов, низкая эффективность современных средств поддержки принятия решений, сложность и трудоемкость многих методов, неоднозначность оценок и субъективизм экспертов, а иногда даже наукообразие некоторых методов – основные причины для скептического отношения практиков к имеющимся технологиям прогнозирования.

Как следствие, научных прогнозов и комплексных оценок развития мира и отдельных стран не так уж и много. Среди них

следует выделить модель Форрестера, долгосрочную модель развития энергетики и состояния окружающей среды ЕС – VLEEM, модель Месаровича-Пестеля, прогноз Pricewaterhouse-Coopers «Мир в 2050 году», прогноз Дж. Ф. Коутса «2025: Сценарии развития США и мирового сообщества под воздействием науки и технологий», прогноз новых технологий университета им. Дж. Вашингтона, прогнозы глобальных климатических и экологических изменений и т.д. Практика обычно показывает низкую достоверность таких прогнозов и оценок, однако они имеют большое значение для развития методологии прогнозирования и комплексной оценки.

В настоящее время опубликовано множество работ, направленных на стандартизацию параметров и индикаторов для прогнозирования и комплексной оценки развития стран мира [4, 5, 16, 19], что позволяет создавать базы данных структурированной информации. Достаточно полный обзор научных и научно-методических исследований в области анализа развития стран мира выполнен в Докладах Программы развития ООН (ПРООН), Всемирного банка, аналитических структур ЦРУ и т.д. [4, 5, 21].

Сравнение военной мощи различных государств также основывается на различных индикаторах. Например, известный рейтинг Global Firepower Index (GFP) составляется для 106 стран на основе базы данных, куда входит более 50 показателей вооруженных сил стран мира [22]. Политическая, экономическая и военная силы и самодостаточность государств также имеют определенные количественные индикаторы. Существуют различные системы оценки «национальной силы» государств, позволяющие провести сравнительный анализ стран [3].

Основные направления и тенденции исследований в области прогнозирования развития стран и их регионов связаны с накоплением и созданием все более обширных баз данных показателей, применением новых методов визуализации и оценки данных, использованием статистических методов анализа информации, созданием национальных информационно-аналитических систем хранения, представления и обработки данных, развитием систем математического моделирования, совершенствованием методов ИАД и системной динамики.

Однако, несмотря на значительный научный прогресс, следует признать, что в настоящее время не существует фундаментальной теории, которая характеризовала бы социально-экономическое развитие регионов, стран, а тем более мира в целом. Тем не менее, во многих случаях использование ИТ-технологий позволяет осуществить обобщение базовых закономерностей, направленных на построение теорий, и дает возможность научно обосновать поиск закономерностей в массивах информации. Применение ИАД при решении многих сложных задач пока не имеет альтернативы, несмотря на ряд существенных недостатков. Среди таких недостатков следует указать слабую связь с особенностями прикладной области и невозможность учета фундаментальных закономерностей изучаемых систем, применение логических методов «слепого» поиска закономерностей, отсутствие средств анализа данных, использующих высокие уровни понимания информации, и т.д.

Одна из тенденций развития методов прогнозирования и ИАД связана с применением различных моделей для описания данных. Сегодня можно уже говорить о новых возможностях создания таких моделей применительно к некоторым видам больших баз данных. Например, компания Palantir Technologies в своих продуктах применяет онтологические модели данных, компания Quid Inc. использует в продукте анализа данных Quid модель, которая получила название «техноценоза» [7, 15].

Самые последние тенденции в разработках прогностических и поведенческих вооружений связаны с предсказанием развития и поведения групп объектов, которые характеризуются множеством показателей. Такие задачи сводятся к исследованию процессов развития объектов в многомерных информационных пространствах при различных воздействиях и определенных ограничивающих условиях. Например, основные исследования Агентства по перспективным исследованиям США (IARPA) связаны с разработкой методологии, математического аппарата и программных решений по прогнозированию сложных социальных и общественных систем и управлению групповым поведением. В математическом обеспечении упор делается не на применение методов математической статистики, а на использование теории категорий и функторов, методов

распознавания образов, нейронных вычислений, машинного обучения и построение различных моделей коллективного взаимодействия и развития объектов [8, 15, 20].

Выполненный обзор информации по изучаемой проблеме позволяет сделать следующие выводы:

- развитие прогностических систем идет по пути совершенствования IT-решений сбора, обработки, хранения и визуализации информации. Анализ данных осуществляется с помощью статистических методов, средств ИАД, а также применения алгоритмов поиска явных и латентных связей, требующих значительных ресурсов;
- при создании математического обеспечения прогностических систем пока слабо учитываются закономерности поведения и развития систем в прикладных областях, что часто связано с отсутствием возможности построения адекватных моделей описания данных и теорий их создания. Поэтому при использовании даже самых современных продуктов анализа данных требуется работа высококвалифицированных экспертов, которые принимают решения и делают окончательные выводы;
- начинают появляться прогностические системы, которые предполагают наличие моделей описания изучаемых объектов и процессов. Однако, эти модели пока носят скорее качественный, чем количественный характер.

Основные трудности, с которыми столкнулись разработчики прогностических систем, заключаются в сложности интегрирования информации из различных источников для получения структурированных данных и в значительных проблемах, связанных с созданием адекватных моделей описания данных. Формирование массивов данных для анализа изначально требует исключительно трудоемкой работы множества исследователей. Большие структурированные данные можно получить там, где имеются стандартизованные методики сбора и обработки данных и применяются различные модели для оценки их точности и достоверности, а также используются методы сглаживания, описания и ассимиляции данных и т.д. На базе таких данных уже могут создаваться модели, которые учитывают основные закономерности развития и поведения объектов.

Однако практикой пока не выработаны основные принципы создания прогностических моделей, позволяющих получать достоверные прогнозы развития многомерных систем.

3. Принципы, гипотезы и методы

Сущность прогностических методов, реализованных на высоких уровнях понимания информации, заключается в том, что алгоритмы анализа данных изначально основываются на математических моделях описания количественных данных применительно к некоторым видам проблемно-ориентированных баз данных. В данном случае под моделью описания данных понимаются логические, теоретические и феноменологические закономерности, которые позволяют предложить математические или алгоритмические методы и средства для моделирования данных в виде некоторого подхода или теории применительно к изучаемому классу систем.

Новые идеи в области прогнозирования поведения больших систем связаны с феноменологическими подходами анализа и описания фактических данных [1, 17, 18], представленных в виде измерений или наблюдений. Данное научное направление имеет большое значение, так как позволяет предложить теоретические методы исследования многомерных общественных, экологических и глобальных систем.

Предлагаемый подход ориентирован на построение моделей описания данных, представляемых в виде временных рядов измерений или наблюдений различных величин. Практика показывает, что для целого ряда систем, которые изменяются и развиваются во времени, возможно формирование таблично-временных массивов информации. Обычно такие массивы имеют структуру таблиц в виде матриц «объекты-атрибуты», причем множество таблиц (t) упорядочено по времени, например, годам, месяцам, часам и т.д. В качестве объектов выступают однотипные классы, например, общественные, социальные и экономические системы одного вида, предприятия, города, регионы, страны, социальные группы, различные категории людей, граждане и т.д.

В качестве атрибутов (показателей), отражающих свойства определенных видов систем, могут быть различные физические, химические, биологические, технологические, социально-экономические или идентификационные величины, имеющие количественное измерение. Для определенного объекта значения каждого показателя в таблично-временном массиве данных будут представлены временным рядом из опытных точек в количестве t , которые задаются с определенным лагом. В целом отметим, что при оценке состояния и развития объектов возможно использование любых разноплановых статистических данных, характеризующих социально-экономический, финансовый, промышленный, научно-технологический, военный, ресурсный или культурный потенциал.

При таком описании каждый объект в определенный момент наблюдения находится в некотором состоянии и характеризуется совокупностью показателей. Подобный общепринятый подход позволяет определить состояние объекта в многомерном пространстве показателей как совокупность его наблюдаемых свойств, параметры которых формируются в конкретный момент времени.

Предположим, что для r однотипных объектов, формирующих систему определенной природы, в таблично-временных массивах данных содержится количественная информация об n показателях, характеризующих множество самых различных свойств данной системы. Перечень атрибутивных показателей p_k ($k=1, \dots, n$) для моделирования определяется сложившимися в научном сообществе представлениями о поведении изучаемых объектов, корреляционным анализом данных или другими методами установления наиболее значимых переменных. Любое множество n переменных для параметров свойств задает n -мерное пространство Ω^n , где $p=(p_1, p_2, \dots, p_n)$, $p_k \in \Omega^n$. Точки этого пространства соответствуют n -мерным наборам значений всех переменных $p_1, p_2, \dots, p_n$. Таким образом, состояние любого объекта в n -мерном пространстве в каждый момент времени будет отображаться многомерной точкой $M=M(p_1, p_2, \dots, p_n)$, будем называть ее фигуративной точкой. Тогда процесс изменения состояния объекта во времени характеризуется кривой в n -

мерном пространстве, которая описывается фигуративной точкой $M = M(p_1, p_2, \dots, p_n)$ в этом пространстве.

Основные принципы, используемые при обработке количественной информации заключаются в следующем.

В основу методов описания данных положен континуальный принцип представления количественной информации в пространстве Ω^n . В соответствии с этим принципом информационная среда в виде многомерного пространства показателей считается непрерывной, бесструктурной, а каждый элемент пространства связан со всеми соседними элементами с учетом закономерностей, свойственных изучаемой предметной области. Это позволяет рассматривать имеющиеся опытные данные как некоторую выборку из сплошной информационной среды бесконечного количества состояний для объектов одного класса.

Второй принцип основан на гипотезе, что данные формируют в континуальном информационном пространстве некоторый «образ». Считается, что данный «образ» может быть описан на основе оценок статистических вероятностей событий, характеризующих местоположение каждой опытной точки относительно всей группы изучаемых объектов (всего облака точек M). Исходя из вероятностных представлений, предполагается возможность построения уравнений состояний объектов в виде многомерных распределений $w = F(p_1, p_2, \dots, p_n)$, при этом каждое состояние в информационном пространстве Ω^n представляется совместным событием A_j одновременного наблюдения показателей $p_1, p_2, \dots, p_n$. Такой подход позволяет искать связи и закономерности не между показателями объектов, а между вероятностями наблюдения различных событий, свойственных изучаемому массиву данных. При данном подходе основной объект моделирования – это состояния объектов (стран, регионов, городов, предприятий, граждан и т.д.), которые могут характеризоваться многомерным эмпирическим распределением.

Третий принцип связан с возможностью алгоритмического определения статистических вероятностей событий A_j одновременного наблюдения нескольких показателей. При этом считаем, что состояние определенного объекта в заданный момент

времени характеризуется не только совокупностью параметров показателей этого объекта, но и вероятностью наблюдаемого события A_j . Назовем эту статистическую вероятность вероятностью состояния изучаемого объекта. Статистические вероятности для сложного события A_j могут быть найдены с использованием различных алгоритмов перебора, группировки и подсчета частот благоприятных событий в общей выборке всех наблюдений [1]. С этой целью статистическая вероятность совместного события w определяется алгоритмически через относительные частоты событий A_j путем разбиения всего пространства Ω^n на многомерные параллелепипеды ω_k , исходя из заданного количества интервалов группирования для каждой переменной p_k (обычно одинакового для всех p_k). После этого подсчитываются относительные частоты событий, которые равны отношению количества опытных точек, которые попадают в заданные области группирования ω_k , к общему числу всех наблюдаемых точек. Статистическая вероятность точки в определенной области ω_k принимается в виде кумулятивных относительных частот изучаемых событий.

Четвертый принцип связан с возможностью феноменологического описания поля вероятностей событий A_j в пространстве Ω^n с помощью моделирующих функций. При известном виде таких функции $\theta = \theta(p_1, p_2, \dots, p_n)$ и значениях переменных $p_1, p_2, \dots, p_n$, применительно к континуальному пространству Ω^n можно сформулировать несколько постулатов, которые позволяют построить феноменологическую модель описания данных.

1. Пусть в пространстве состояний системы Ω^n каждой точке M поставлено в соответствие действительное число w ($0 \leq w \leq 1$), которое будем называть статистической вероятностью наблюдаемого состояния.

2. Величина $w(M)$ является функцией точки и образует скалярное вероятностное поле, которое является непрерывным в области Ω^n .

Для построения модели описания данных предположим, что скалярное поле величины w может быть аналитически описано в окрестности произвольной точки M . Для описания скалярного поля $w = w(M)$ как функции независимых переменных

$p_1, p_2, \dots, p_n$ необходимо определить функцию точки. Предположим, что в области Ω^n можно задать аналитическую непрерывную функцию $\theta(p_1, p_2, \dots, p_n)$, на основе которой будет формироваться математическая модель. Исходя из этого, в пространстве Ω^n можно построить еще одно скалярное поле, которое является основой для моделирования. Поэтому, для построения феноменологической модели описания данных сформулируем гипотезу.

3. Пусть в пространстве состояний некоторой системы Ω^n скалярные поля величин w и θ однозначно связаны между собой. Если в окрестности любой точки M объект системы осуществляет некоторый процесс l , то для линии процесса l справедливо соотношение $dw = c_l d\theta$, где c_l – эмпирические величины, которые являются функциями процесса.

Выберем в области Ω^n произвольную точку M . Будем считать, что вблизи данной точки осуществляется элементарный процесс изменения состояния некоторого объекта. Тогда элементарное изменение величины w можно представить в виде:

$$(1) \quad dw = \left(\frac{\partial w}{\partial \theta} \right)_{p_2, \dots, p_n} \cdot \left(\frac{\partial \theta}{\partial p_1} \right) dp_1 + \left(\frac{\partial w}{\partial \theta} \right)_{p_1, p_3, \dots, p_n} \cdot \left(\frac{\partial \theta}{\partial p_2} \right) dp_2 + \dots \\ \dots + \left(\frac{\partial w}{\partial \theta} \right)_{p_1, \dots, p_{n-1}} \cdot \left(\frac{\partial \theta}{\partial p_n} \right) dp_n,$$

$$(2) \quad dw = c_1 \cdot \frac{\partial \theta}{\partial p_1} dp_1 + c_2 \cdot \frac{\partial \theta}{\partial p_2} dp_2 + \dots + c_n \cdot \frac{\partial \theta}{\partial p_n} dp_n.$$

Величины c_k в общем случае зависят от показателей $p_1, p_2, \dots, p_n$.

Величина θ в области Ω^n может быть представлена в виде различных функциональных зависимостей относительно атрибутивных показателей: мультипликативными, степенными, аддитивными, экспертными или иными зависимостями, входящими в классы однородных или мультипликативных функций. В работе [1] установлено, что при этих условиях в пространстве Ω^n для построения моделей описания данных можно использовать квазилинейные многомерные уравнения в частных производных первого порядка, которые тесно связаны с уравнениями

Пфаффа вида (2). При этом, используя библиотеку функций $\theta(p_1, p_2, \dots, p_n)$ можно построить различные модели с разной точностью описания данных.

В связи с тем, что мы строим вероятностную среду моделирования, будем считать, что аналитическая функция $\theta(p_1, p_2, \dots, p_n)$ может быть представлена в виде геометрической вероятности $\theta = (p_1 \cdot p_2 \cdot \dots \cdot p_n) / (p_{1_m} \cdot p_{2_m} \cdot \dots \cdot p_{n_m})$ пространства состояний системы Ω^n , где p_{k_m} – максимально наблюдаемые значения величин p_k в группе объектов. Можно показать, что в данном случае для уравнения (2) существует интегрирующий делитель, который обращает данное уравнение в полный дифференциал. Покажем, что интегрирующим делителем уравнения (2) будет функция θ , принятая в виде геометрической вероятности пространства Ω^n . Подставив данную функцию в (2) и деля это уравнение на величину θ , получим:

$$(3) \quad ds = \frac{dw}{\theta} = c_1 \cdot \frac{dp_1}{p_1} + c_2 \cdot \frac{dp_2}{p_2} dp_2 + \dots + c_n \cdot \frac{dp_n}{p_n} dp_n.$$

Интегрируя уравнение (3) в окрестности состояния M и считая величины c_k условно постоянными, представим общий интеграл в виде:

$$(4) \quad s - s_0 = c_1 \cdot \ln\left(\frac{p_1}{p_{1_0}}\right) + c_2 \cdot \ln\left(\frac{p_2}{p_{2_0}}\right) + \dots + c_n \cdot \ln\left(\frac{p_n}{p_{n_0}}\right),$$

где $s_0, p_{1_0}, \dots, p_{n_0}$ – параметры некоторого опорного состояния в пространстве Ω^n .

Определим общий интеграл s как энтропию, исходя из аналогий с термодинамикой и теорией информации. Энтропия является характеристической функцией пространства состояний системы Ω^n . Как показано в источнике [1] в параметрическом представлении энтропия является длиной дуги векторной линии некоторого поля направлений, порождаемого скалярным полем статистической вероятности w . Этот вывод вытекает из вида уравнения Пфаффа (2) и его общих решений. Известно, что соответствующее поле направлений имеет вид:

$$(5) \quad \Gamma(p_1, p_2, \dots, p_n) = \frac{p_1}{n \cdot c_1} \cdot \mathbf{e}_1 + \frac{p_2}{n \cdot c_2} \cdot \mathbf{e}_2 + \dots + \frac{p_n}{n \cdot c_n} \cdot \mathbf{e}_n,$$

где $\mathbf{e}_k$ – единичные векторы, направленные соответственно по осям координат $p_1, p_2, \dots, p_n$ пространства состояний Ω^n . Векторные линии данного поля будут определяться обыкновенными дифференциальными уравнениями:

$$(6) \quad n \cdot c_1 \frac{dp_1}{p_1} = n \cdot c_2 \frac{dp_2}{p_2} = \dots = n \cdot c_n \frac{dp_n}{p_n} = \frac{dw}{\theta} = ds,$$

откуда легко получить выражение (3) для изменения энтропии состояния.

Для вектора Γ существует семейство поверхностей, ортогональных к линиям энтропии s . Уравнение таких поверхностей определятся из скалярного произведения $(\Gamma \cdot \mathbf{t}) = 0$, где $\mathbf{t} = \mathbf{e}_1 \cdot dp_1 + \mathbf{e}_2 \cdot dp_2 + \dots + \mathbf{e}_n \cdot dp_n$ – вектор, лежащий в касательной плоскости к исходной поверхности. Это уравнение в развернутом виде приводит к соотношению:

$$(7) \quad \frac{p_1}{c_1} dp_1 + \frac{p_2}{c_2} dp_2 + \dots + \frac{p_n}{c_n} dp_n = 0,$$

которое также является уравнением Пфаффа. Пфаффова форма, стоящая в левой части уравнения Пфаффа (7), при постоянных величинах c_k в окрестности точки M является полным дифференциалом, поэтому из уравнения (7) может быть определен потенциал $P(p_1, p_2, \dots, p_n) = C$, который представляет собой поверхность, ортогональную векторным линиям энтропии:

$$(8) \quad P = \frac{1}{2} \left(\frac{p_1^2 - p_{1_0}^2}{c_1} dp_1 + \frac{p_2^2 - p_{2_0}^2}{c_2} dp_2 + \dots + \frac{p_n^2 - p_{n_0}^2}{c_n} dp_n \right).$$

Потенциал в опорном состоянии принят равным нулю. Энтропия s и потенциал P могут быть использованы в качестве обобщенных критериев для прогнозирования группового поведения и комплексной оценки состояния объектов в многомерном пространстве Ω^n . Их наиболее важной особенностью является то, что данные величины являются функциями состояния системы при справедливости условия существования скалярного поля статистической вероятности w . Изменение данных

функций зависит только от начального и конечного состояния объекта и не зависит от пути его перехода между этими состояниями.

Таким образом, создание феноменологических моделей описания данных основано на постулировании существования многомерного скалярного поля величины w и описании процессов через связь величин w и θ вида $dw = c_l d\theta$. Величины c_l определяются по таблично-временным массивам данных, характеризующим изучаемую группу объектов.

Отсюда видно, что предлагаемый метод тесно связан с логикой построения теории термодинамики, так как изначально вводятся феноменологически определяемые величины c_l . Поэтому по аналогии на основе данного подхода возможно применение к описанию структурированных данных соответствующего развитого математического аппарата.

Все это позволяет предложить теорию для описания некоторых типов количественной информации. В целом, данный подход позволяет создавать модели описания массивов данных, представленных в виде временных рядов количественных показателей и характеризующих процессы изменения и развития социально-экономических и общественных систем, для которых имеется обширная совокупность результатов статистических наблюдений.

4. Пример оценки уровня человеческого развития стран

Полученные результаты позволяют предложить метод анализа развития стран мира и построить систему оценки их показателей, не опираясь на экспертные методы. Использование данного подхода покажем на примере разработки методики оценки человеческого развития, альтернативной методике расчета индекса человеческого развития ПРООН. С 2010 года в методике ПРООН для расчета индекса в качестве атрибутивных показателей стали использоваться следующие величины: средняя продолжительность обучения (p_1), лет; ожидаемая продолжительность обучения (p_2), лет; валовый национальный доход (ВНД) на душу населения в пересчете по паритету покупатель-

ной способности (ППС) в долларах США (p_3); ожидаемая продолжительность жизни (p_4), лет.

В качестве характерного события для определения вероятности состояния объектов выберем совместное событие одновременного наблюдения указанных выше четырех величин. Воспользуемся последними Докладами ПРООН [5], базы данных которых охватывают данные по странам мира с 1990 по 2014 годы. Алгоритмы подсчета частот благоприятных событий [1, 18] дают возможность найти значение вероятности характерного события для каждой страны мира, исходя из имеющегося массива опытных данных. Статистическая вероятность w подсчитывается во всей группе объектов (169 стран).

Для поиска нелинейных связей между переменными воспользуемся методом пробит-анализа [19]. С учетом (4) свяжем полученную вероятность w с распределениями атрибутивных показателей в массиве опытных данных, в результате чего будем иметь следующую регрессионную зависимость вероятности от энтропии состояния системы (рис. 1):

$$(9) \quad Prob = -3,050 + s; \quad w = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{Prob} \exp\left(-\frac{t^2}{2}\right) dt;$$

$$s = 0,338 \cdot \ln\left(\frac{p_1}{p_{1_0}}\right) + 0,861 \cdot \ln\left(\frac{p_2}{p_{2_0}}\right) + 0,165 \cdot \ln\left(\frac{p_3}{p_{3_0}}\right) + 2,403 \cdot \ln\left(\frac{p_4}{p_{4_0}}\right).$$

Коэффициент корреляции зависимости (9) составил 0,98. Переменные относились к значениям $p_{1_0}, p_{2_0}, p_{3_0}, p_{4_0}$, которые соответствуют выбранному опорному состоянию. В качестве опорного состояния были приняты показатели развития страны Нигер в 2008 году, как одной из самых слаборазвитых стран, которые, в частности, равны: $p_{1_0} = 1,4$ лет; $p_{2_0} = 4,3$ лет; $p_{3_0} = 675$ \$; $p_{4_0} = 52,5$ лет.

С учетом вероятностной оценки совместных событий, связанных с наблюдением показателей p_1, p_2, p_3, p_4 , были определены ранги развития для каждой страны мира. Результаты ранжирования стран мира приведены в таблице 1, исходя из оценки уровня человеческого развития стран в период 2008–2014 годов.

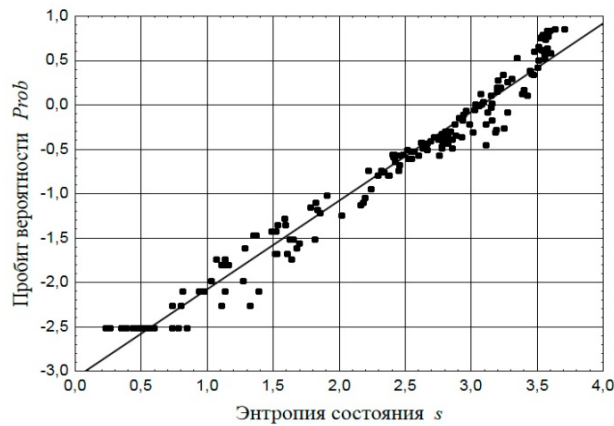


Рис. 1. Зависимость статистической вероятности состояния w от энтропии состояния системы s для совместно наблюдаемых значений показателей p_1, p_2, p_3, p_4

К первым десяти странам, имевшим самый высокий уровень развития в 2014 году, относятся: Катар, Кувейт, Лихтенштейн, Сингапур, Бруней, Норвегия, ОАЭ, Люксембург, Швейцария и Гонконг. Видно, что из Большой двадцатки стран мира (G20) в этом списке нет ни одной страны, а из ЕС – присутствует только Люксембург. По темпам развития в период 2008–2014 годов к десятке быстро развивающихся стран относятся: Катар, Кувейт, Сингапур, Бруней, Саудовская Аравия, Швейцария, Гонконг, Люксембург, Бахрейн и Норвегия.

К десяти странам, имевшим самый низкий уровень развития, относятся Нигер, Бурунди, Гвинея, ЦАР, Мозамбик, Либерия, Буркина-Фасо, Эфиопия, Гвинея-Бисау и Малави. Все эти страны находятся в Африке. По темпам развития в 2008–2014 годах самыми отстающими странами были: Барбадос, Лихтенштейн, Багамские острова, Греция, Ливия, Экваториальная Гвинея, Соломоновы острова, Малави, Гондурас и Того. В данном списке присутствуют Лихтенштейн и Греция, входящие в группу стран ЕС, у которых в изучаемый период времени произошло снижение удельного ВНД.

Таблица 1. Значения потенциала (P), его изменение (ΔP) и рейтинги стран в процессах развития стран мира в 2008–2014 годах

Страны мира	Потенциал развития P (2014)	Изменение потенциала ΔP (2008–2014)	Ранги развития стран		Ранги стран по ИЧР согласно методике ПРООН (2014)
			уровень развития (2014)	темпы развития (2008–2014)	
Норвегия	56437,1	10182,6	6	10	1
Германия	25922,0	9109,0	16	13	4
США	37549,5	7810,3	11	15	8
Канада	23899,3	3804,2	20	33	9
Сингапур	78282,2	46363,7	4	3	11
Гонконг	38923,4	11726,1	10	7	12
Швеция	27929,2	9573,2	13	11	14
Англия	20775,8	4257,6	23	31	15
Япония	18343,8	2130,2	27	51	20
Франция	19458,3	3601,6	26	37	22
Греция	8168,9	-2123,0	43	163	29
Эстония	8699	4555,2	41	24	30
Катар	201775,6	117780,9	1	1	33
Польша	7361,6	2989,9	44	39	36
Венгрия	7194,7	2922,5	45	40	44
Латвия	6807,0	4410,0	48	29	46
Россия	6865,1	3648,9	47	36	50
Беларусь	3921,2	1564,7	60	60	50
Румыния	4539,9	2171,9	57	50	52
Болгария	3407,1	1606,7	66	57	59
Казахстан	5993,4	4435,7	52	28	56
Турция	4730,0	2293,4	56	48	72
Бразилия	3156,4	1579,5	69	59	75
Грузия	905,9	365,1	99	97	76
Украина	1088,3	324	95	102	81
Китай	2180,9	1396,9	77	65	90
Узбекистан	588,6	312,7	108	103	114
Индия	444,6	270,5	114	104	130
Нигерия	427,7	333,4	116	100	152
Эфиопия	30,5	19,8	159	144	174
Нигер	6	6	166	156	188

Россия в рейтинге по уровню развития в 2014 году занимала 47 место, в свою очередь, Казахстан – 52, Белоруссия – 60, Украина – 95 место. По темпам развития в 2008–2014 годах Россия находилась на 36 месте, Казахстан на 28, Белоруссия на 60, а Украина на 102 месте. За этот же период Россия в темпах человеческого развития опережала Италию, Францию, Японию, Польшу, Чехию и Израиль, однако отставала от США, Германии, Швеции, Литвы, Латвии.

Рейтинги стран мира по уровню развития, определенные по методике расчета индекса человеческого развития ПРООН, существенно отличаются от аналогичного рейтинга, рассчитанного на основе данного подхода. Для большинства развитых стран методика ПРООН дает завышенные рейтинги уровня развития, а для многих развивающихся стран – заниженные рейтинги.

Индекс человеческого развития ПРООН дает «благоприятную» интегральную экспертную оценку для стран «золотого миллиарда» (США, Канада, Австралия, Япония, страны Евросоюза) и «менее благоприятную» для всех остальных стран мира.

5. Оценка человеческого развития регионов России

Использование данного метода покажем также на примере оценки человеческого развития регионов России. В качестве показателей для расчета индекса используем следующие величины: валовый внутренний продукт (ВВП) на душу населения в пересчете по паритету покупательной способности (ППС) в долларах США (p_1); ожидаемая продолжительность жизни (p_2), лет; уровень грамотности населения (p_3), %; доля охвата населения начальным, средним и высшим образованием (p_4), %. Для решения задачи воспользуемся базами данных Докладов развития человека ПРООН для регионов России [12], которые имеются в открытом доступе и являются на сегодняшний день самыми последними официальными статистическими данными (хотя это 2008–2010 годы).

Также, как и в предыдущем разделе, получим регрессионную зависимость вероятности от энтропии состояния системы:

$$(10) \quad \text{Pr ob} = -5,766 + s; \quad w = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^{\text{Pr ob}} \exp\left(-\frac{t^2}{2}\right) dt;$$

$$s = 0,618 \cdot \ln\left(\frac{p_1}{p_{1_0}}\right) + 8,968 \cdot \ln\left(\frac{p_2}{p_{2_0}}\right) + 36,476 \cdot \ln\left(\frac{p_3}{p_{3_0}}\right) + 3,028 \cdot \ln\left(\frac{p_4}{p_{4_0}}\right).$$

Коэффициент корреляции зависимости (10) составил 0,90, результаты обработки данных приведены на рис. 2. Из рисунка видно, что два города (Москва и Санкт-Петербург) выпадают из однородной группы регионов, что вполне естественно.

Атрибутивные показатели выбирались такими же, как и в методике ПРООН до 2010 года, и относились к значениям $p_{1_0}, p_{2_0}, p_{3_0}, p_{4_0}$, в качестве которых были приняты минимальные показатели развития регионов, в частности, равные: $p_{1_0} = 2822 \$$; $p_{2_0} = 59,65$ лет; $p_{3_0} = 56,0 \%$; $p_{4_0} = 46,7 \%$.

Согласно (8) был определен потенциал развития для каждого региона России в 2008 и 2010 годах. Результаты ранжирования по уровню и темпам развития для данного случая приведены в таблице 2.

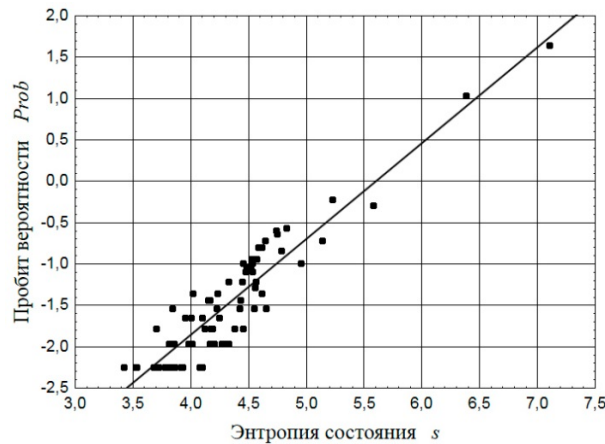


Рис. 2. Зависимость статистической вероятности состояния w от энтропии состояния системы s для совместно наблюдаемых значений показателей p_1, p_2, p_3, p_4

К первым пяти регионам, имеющим самый высокий уровень развития в 2010 году, относились: Тюменская область, город Москва, Сахалинская область, Чукотский автономный округ и Республика Татарстан. Видно, что от первых трех регионов остальные регионы отстают по развитию почти в два и более раза. В среднем по потенциалу развития разница в уровне развития наиболее развитых регионов и наименее развитых регионов достигает 70 раз.

По темпам развития к быстро развивающимся регионам относятся: Сахалинская, Тюменская области, Чукотский автономный округ, Красноярский край и Республика Саха-Якутия.

Города Москва и Санкт-Петербург по темпам развития занимали соответственно 28 и 8 место. Наиболее высокие темпы развития, превышающие средние темпы развития всех остальных регионов в 6–10 раз по изменению потенциала ΔP , наблюдались в Сахалинской и Тюменской областях и Чукотском автономном округе.

К пяти регионам, имеющим самый низкий уровень развития, относятся Республики Калмыкия и Тыва, Кабардино-Балкарская Республика, Чеченская республика и Республика Ингушетия.

По темпам развития самыми отстающими регионами были: Республика Алтай, Республика Ингушетия, Карачаево-Черкесская Республика, Чеченская Республика и Вологодская область.

Выполненный анализ данных указывает на существенную неоднородность уровня и темпов развития регионов России по показателям человеческого потенциала, что является в определенной степени дестабилизирующим фактором.

Полученные модели описания данных в области оценки человеческого потенциала позволяют провести анализ и прогнозирование развития регионов России.

Предложенный метод позволяет существенно повысить точность моделей описания данных, если феноменологические величины c_i , характеризующие развитие объектов, определять исходя не из среднестатистических тенденция развития всей группы объектов, а непосредственно для каждого объекта прогнозирования.

Таблица 2. Значения потенциала (P), его изменение (ΔP) и рейтинги регионов России в процессах развития в 2008–2010 годах

Регионы Российской Федерации	Потенциал развития P (2008)	Изменение потенциала ΔP (2008–2010)	Ранги регионов		Ранги регионов по ИЧР (2008)
			уровень развития (2008)	темпы развития (2008–2010)	
Город Москва	280,92	18,51	2	28	1
Тюменская обл.	475,25	234,01	1	2	2
Республ. Татарстан	72,54	36,39	5	11	4
Белгородская обл.	51,46	52,34	9	7	5
Томская область	47,22	34,94	14	12	6
Сахалинская обл.	165,45	358,49	3	1	7
Липецкая обл.	53,11	6,99	8	65	10
Красноярский край	50,25	91,86	10	4	11
Республ. Коми	59,27	60,05	7	6	12
Челябинская обл.	37,26	9,76	22	51	13
Омская обл.	34,78	17,99	24	29	15
Вологодская обл.	49,35	-3,10	12	80	16
Краснодарский край	22,59	16,53	39	30	22
Московская обл.	46,51	11,78	16	44	23
Ярославская обл.	28,18	12,98	29	39	24
Пермский край	39,88	19,69	18	26	26
Курская обл.	20,53	15,07	40	35	29
Ростовская обл.	16,21	8,23	48	55	30
Кемеровская обл.	49,72	34,22	11	13	31
Мурманская обл.	32,47	25,53	25	18	35
Воронежская обл.	14,55	6,25	54	69	37
Пензенская обл.	12,02	6,18	64	70	45
Хабаровский край	18,61	25,51	42	19	48
Тульская обл.	23,18	7,18	37	63	49
Тамбовская обл.	12,81	8,03	59	56	50
Республ. Карелия	23,76	16,06	35	32	51
Камчатский край	14,18	19,89	57	24	58
Курганская обл.	12,66	8,58	62	54	59
Приморский край	13,55	28,32	58	17	60
Чеченская Республ.	1,93	0,89	79	79	71
Республ. Ингушетия	0,10	2,96	80	77	75

6. Выводы и перспективы

Таким образом, показано, что при достаточном объеме опытных данных возможно построение феноменологических моделей для некоторых классов социально-экономических систем, в частности городов, регионов и стран. Практический опыт применения таких моделей в термодинамике и физике указывает на большие возможности, связанные с описанием и прогнозированием развития многих сложных систем в естествознании. Все это позволяет в прогнозных исследованиях применить объективный подход, отличающийся использованием методов, которые не зависят от воли и желаний субъекта, обеспечивают формализацию научной задачи в области предмета исследования и применяют адекватные модели для описания закономерностей. Важным является также возможность формирования обширной эмпирической базы для получения феноменологических описаний процессов и использование инструментов и средств для опытной проверки научных фактов на практике. Такой подход позволяет разработать принципиально новое математическое обеспечение для прогностических систем.

Предложенный метод, основанный на феноменологических подходах описания структурированных данных и учете закономерностей развития объектов прогнозирования, позволяет строить модели описания данных различной точности, в том числе и очень высокой, независимо от природы изучаемых объектов. Континуальный принцип представления количественной информации и вероятностные методы оценки значимых событий отличаются определенной универсальностью и дают возможности для описания закономерностей поведения объектов, если информация о них представляется в структурированном виде «объекты-показатели-время».

Данный метод может быть использован при изучении групповых особенностей развития и поведения объектов социально-экономической и общественной природы. Например, при изучении тенденций развития стран, регионов, городов и других государственных и общественных образований, при оценке особенностей и уровня налоговых поступлений в группах однородных субъектов налогообложения, при изучении поведения

социальных групп и различных категорий людей, исходя из анализа однородных значимых событий и среднестатистических особенностей и т.д.

Приведенные практические примеры создания моделей описания данных для оценки состояния и развития стран и регионов указывают на достаточную теоретическую проработку предлагаемых прогностических методов.

Литература

1. АВЕРИН Г.В. *Системодинамика*. – Донецк: Донбасс, 2014. – 405 с.
2. АКАЕВ А.А., САДОВНИЧИЙ В.А. *Математическое моделирование глобальной, региональной и национальной динамики с учетом воздействия циклических колебаний* // Моделирование и прогнозирование глобального, регионального и национального развития / Ред. А.А. Акаев, А.В. Коротаев, Г.Г. Малинецкий, С.Ю. Малков – М.: Либроком, 2011. С. 5–67.
3. *Глобальный рейтинг интегральной мощи 100 ведущих стран мира*. – 2-е изд., доп. – М.: Международная Академия исследований будущего, 2008. – 148 с.
4. *Доклады о мировом развитии*. Пер. с англ. публикации Всемирного банка (2000–2015). – М.: Весь мир. Сайт. – URL: www.worldbank.org (дата обращения 15.07.16).
5. *Доклады о человеческом развитии (1990–2015 гг.)*. Программа развития ООН: сайт. – URL: <http://hdr.undp.org/en/2015-report> (дата обращения 20.07.16).
6. ЗГУРОВСКИЙ М.З. *Глобальное моделирование процессов устойчивого развития в контексте качества и безопасности жизни людей (2005–2007/2008 годы)*. – К.: Политехника, 2008. – 331 с.
7. Компания Palantir: сайт. – URL: <http://www.palantir.com> (дата обращения 20.07.16).
8. ЛАРИНА Е. *Кибервойны XXI века* / Е. Ларина, В. Овчинский. – М.: Книжный мир, 2014. – 352 с.
9. *Методологические основы прогнозирования научно-технологического развития России до 2030 г. с использова-*

- нием критериев стратегических рисков / Координационный совет РАН по прогнозированию, 2009. – 27 с.
10. *Прогнозирование будущего: новая парадигма* / Под ред. Г.Г. Фетисова, В.М. Бондаренко. – М.: Экономика, 2008. – 283 с.
11. *Прогностика. Общие понятия. Объект и аппарат прогнозирования. Терминология.* – М.: Наука, 1978. – 32 с.
12. *Программа развития ООН: Развитие человеческого потенциала в регионах России в 2013 году*: сайт. – URL: <http://gtmarket.ru/news/2013/06/17/6014> (дата обращения 15.07.16).
13. *Проекты и риски будущего: Концепции, модели, инструменты, прогнозы* / Ред. А.А. Акаев, А.В. Коротаев, Г.Г. Малинецкий, С.Ю. Малков / Будущая Россия. – М.: Красанд, 2011. – 432 с.
14. САДОВНИЧИЙ В.А. *Моделирование и прогнозирование мировой динамики* / В.А. Садовничий, А.А. Акаев, А.В. Коротаев, С.Ю. Малков. – М.: ИСПИ РАН, 2012. – 359 с.
15. ЧЕРНЯК Л. *Аналитика двойного назначения* // Открытые системы СУБД. – 2013. – № 10. – С. 48 – 50.
16. *Экологические показатели и основанные на них оценочные доклады.* – Нью-Йорк, Женева: ООН. 2007. – 110 с.
17. AVERIN G., KONSTANTINOV I., ZVIAGINTSEVA A., TARASOVA O. 2015. *The Development of Multi-Dimensional Data Models Based on the Presentation of an Information Space as a Continuum.* International Journal of Soft Computing, Vol. 10(6): 458–461.
18. AVERIN G., ZVIAGINTSEVA A., KONSTANTINOV I., IVASHCHUK O. *Data Intellectual analysis means use for condition indicators assessment of the territorial and state formations* // Research Journal of Applied Sciences. – 2015. – no.10(8). – pp. 411–414.
19. BLISS C. *The method of probits* // Science 79(2037), 1934.– pp. 38–39.
20. *Central Intelligence Agency (US).* Available at: <http://www.cia.gov/library/publications/> (18.07.16).
21. *Global Firepower (GFP):* Available at: <http://www.globalfirepower.com> (15.07.16).

22. *World Development Indicators (issued annually)*. World Bank.
Available at: <http://data.worldbank.org/products/wdi> (15.07.16).

ABOUT THE PROGNOSTIC METHODS USING HIGH LEVELS OF UNDERSTANDING OF INFORMATION

Gennady Averin, Belgorod State National University, Belgorod, Doctor of technical science, professor (averin@bsu.edu.ru, averin.gennadiy@gmail.com).

Igor Konstantinov, Belgorod State National University, Belgorod, Doctor of technical science, professor (konstantinov@bsu.edu.ru).

Anna Zviagintseva, Belgorod State National University, Belgorod, Candidate of technical science, docent (zviagintseva@bsu.edu.ru, anna_zv@ukr.net).

Olga Ivaschuk, Belgorod State National University, Belgorod, Doctor of technical science, docent (ivaschuk@bsu.edu.ru).

It is shown that the principle of multidimensional continual representation of quantitative information and the principle of the phenomenological description of scalar fields of probabilities of the joint events, reflecting conditions of objects of forecasting, can be put in a basis of methods of the description of data on a state and development of public systems. The offered prognostic methods are based on probabilistic models of the data using high levels of understanding of information. These methods assume application of knowledge of patterns of development of objects, which are determined in the course of the analysis of data. The dependences for creation of models of data in multidimensional information spaces are offered and the examples of creation of such models for some classes of social and economic systems are considered.

Keywords: prognostic methods, data mining.

*Статья представлена к публикации
членом редакционной коллегии*

*Поступила в редакцию
Опубликована*