

methane concentration in the outgoing air stream flowing from the coal face. Level and dynamics of the methane concentration adequately characterize the fact of implementation of and duration of the outburst prevention measures, and application of appropriate sensors automates monitoring of their implementation. The experimental results support a conclusion that gas content in the coal seam is one of the informative indicators of effectiveness of the outburst prevention measure implementation.

Keywords: coal seam, outburst, gas emission, working, methane concentration.

Статья поступила в редакцию 20.08.2013

Рекомендовано к публикации д-ром техн. наук Т.В. Бунько

УДК 303.732.4:332.1

Г.В. Аверин, д-р техн. наук, профессор,
А.В. Звягинцева, канд. техн. наук, доцент
(ГВУЗ «ДонНТУ»)

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАНЫХ ПРИ ОЦЕНКЕ РАЗВИТИЯ УКРАИНЫ

Г.В. Аверін, д-р техн. наук, професор,
Г.В. Звягінцева, канд. техн. наук, доцент
(ДВНЗ «ДонНТУ»)

ЗАСТОСУВАННЯ МЕТОДІВ ІНТЕЛЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ПРИ ОЦІНЦІ РОЗВИТКУ УКРАЇНИ

G.V. Averin, D. Sc. (Tech.), Professor,
A.V. Zviyagintseva, Ph.D. (Tech.), Associate Professor
(Donetsk National Technical University)

APPLICATION OF DATA MINING TECHNIQUES FOR THE ASSESSMENT OF UKRAINE DEVELOPMENT

Аннотация. Показана возможность применения методов интеллектуального анализа данных (ИАД), отличающихся высоким уровнем понимания информации, для изучения социально-экономического развития стран мира. В качестве исходной информации использована база данных Программы развития ООН. Построена модель развития стран мира на основе применения метода многомерного шкалирования по атрибутивным индикаторам системы и алгоритмов ИАД для вычисления статистической вероятности распределения наблюдаемых в совокупности показателей развития стран. Предложен метод определения энтропии и потенциала развития объектов. Проведены вычислительные эксперименты и определены ранги развития стран мира, а также ранги темпов развития. Выполнена оценка развития Украины по сравнению с основными странами мира. Полученные результаты могут быть использованы при стратегическом прогнозировании и планировании развития страны.

Ключевые слова: интеллектуальный анализ данных, модели развития стран, стратегическая оценка, ранги стран мира, темпы развития Украины.

В настоящее время поиск закономерностей развития стран связывают с глобалистикой, прогностикой и другими междисциплинарными научными направлениями, изучающими развитие общества. Глобалистика выявляет сущность, тенденции и причины процессов глобализации, а также анализирует последствия глобальных процессов для человека и биосферы. Прогностика представляется как наука для предсказания будущего. Практическим результатом данных областей знаний являются научные методы стратегического прогнозирования и планирования. Сегодня во многих странах растет понимание того, что стратегическое планирование и национальная безопасность неразрывно связаны между собой, и эти составляющие государственной политики должны быть направлены на достижение долгосрочных целей, которые определяли бы место страны в бурно меняющемся мире XXI века. Однако реализация такой политики не мыслима без эффективной системы стратегического прогнозирования. Теоретические работы прогностики в области политических, экономических и экологических наук, а также в области глобальных исследований (Global studies) часто сводятся к гипотезам и обобщениям, оторванным от реальной статистической базы и систематического изучения фактов.

В области теоретических исследований существует несколько проблем, которые не позволяют многочисленным научным обобщениям превратиться в общепринятые теории. Во-первых, не редко изначально формулируются противоречивые и явно непознаваемые теоретические концепции. «Прогнозировать будущее можно только из будущего» – данная идея, которая претендует на новую парадигму прогнозирования и особую методологию познания [1], вряд ли может быть подтверждена или опровергнута на данном этапе науки и практики. Во-вторых, очень часто теоретические гипотезы высказываются на основе «озарений» и поверхностного обобщения данных, при этом современные методы анализа информации не используются, а результаты не проверяются на массивах информации. Все это приводит к «валу» частных и противоречивых моделей или, вообще, к множеству различных качественных описаний общественных процессов и явлений, которые не могут быть обобщены. В-третьих, уже почти пятнадцать лет существуют обширные и общедоступные базы данных о развитии стран и регионов мира, и только в последние годы наметился явный интерес исследователей к данной информации.

Как следствие, научных прогнозов развития мира, при составлении которых задействованы коллективы исследователей, не так уж и много. Среди них следует выделить модель Форрестера, модель Месаровича-Пестеля, прогноз PricewaterhouseCoopers «Мир в 2050 году», долгосрочную модель развития энергетики и состояния окружающей среды ЕС-VLEEM, прогноз Дж. Ф. Коутса «2025: Сценарии развития США и мирового сообщества под воздействием науки и технологий», прогнозы глобальных климатических и экологических изменений и т.д. Время обычно показывает низкую достоверность таких прогнозов,

однако они имеют большое значение для развития методологии прогнозирования в области изучения глобальных процессов.

Следует признать, что в настоящее время не существует фундаментальной теории, которая характеризовала бы социально-экономическое развитие и экологические изменения стран и регионов, а тем более мира в целом. Современная наука должна дать лицам, принимающим решения, понимание закономерностей развития общества и предложить новую парадигму прогнозистики. Методология фундаментальной теории может быть сформулирована только на основе использования объективного подхода, проведения междисциплинарных исследований и установления количественных закономерностей. Решение данной задачи возможно с помощью методов интеллектуального анализа данных (ИАД) – области науки, которая бурно развивается.

Сегодня анализ данных социально-экономического развития стран и регионов мира не мыслим без использования методов ИАД. В этой области при изучении процессов и объектов исследователь оперирует массивами данных, которые содержат сотни статистических показателей. Современная карта мира включает около 200 стран, многие из которых имеют административное деление на десятки регионов, республик, областей, округов, районов, штатов, провинций, земель и т.д. В свою очередь, ретроспективная глубина данных может составлять десятки лет по каждому объекту с разбивкой на кварталы и даже месяцы. Известно, что процесс обобщения результатов наблюдений является первым этапом в построении любой теории. Современные инструменты для анализа, статистической обработки и визуализации подобных массивов информации имеются пока только в ИАД. В этой области могут лежать ответы на многие актуальные вопросы о путях развития мирового сообщества и его будущего.

Постановка задачи. Цель статьи – показать возможности применения методов ИАД для изучения социально-экономического развития стран мира. Это одна из актуальных задач прогнозистики в области глобальных исследований.

В качестве исходной информации использована для анализа база данных Программы развития ООН (<http://hdr.undp.org/en/data>). База данных Программы развития ООН (ПРООН) включает статистические таблицы данных почти по 100 странам в период 1975 – 1980 годов и по 187 странам в период 2011 – 2013 годов. База данных содержит около 100 индикаторов, по которым определяются несколько индексов, характеризующих различные аспекты человеческого развития, в том числе и индекс человеческого развития (ИЧР). На Web-сайте организации предоставлены инструменты визуализации и работы с данными для чаще всего используемых индикаторов (около 50). Пользователь может создавать свои таблицы и конвертировать их в Excel формат. База данных уже несколько лет находится в открытом доступе.

Информация ПРООН не является абсолютно полной. Например, данные по детской смертности в возрасте до 5 лет приведены для 1970, 1980, 1990, 2000, 2005, 2007–2009 годов. Данные для удельного потребления энергии странами представлены только для 1980 и 2003 годов. Однако большинство показателей стран ме-

няется достаточно медленно. К примеру, детская смертность в Украине с 1970 по 2010 годы изменилась с 34 до 13 смертей на 1 тысячу живорожденных (в целом по миру в среднем она снизилась с 146 до 75 смертей). В свою очередь, младенческая смертность в Украине с 1970 по 2010 год снизилась с 22 до 11 смертей на одну тысячу живорожденных, а потребление энергии в целом по миру с 1980 по 2003 год в среднем возросло с 1573 до 2490 кВтч на душу населения. Анализ показывает, что за 30 – 40 лет для большинства показателей стран мира наблюдаются изменения значений в среднем не более, чем в 1,5 – 2,5 раза. С 2000 года информация по основным индикаторам развития является достаточно полной, позволяющей применить методы интеллектуального анализа данных.

Построение модели системы. В качестве атрибутов системы выберем принятые при расчете ИЧР показатели: ожидаемую продолжительность жизни (ρ_1), уровень грамотности населения (ρ_2), коэффициент охвата населения средним и высшим образованием (ρ_3), ВВП страны в пересчете по ППС (ρ_4), а также дополнительно долю городского населения (ρ_5). Первые четыре показателя входят в методику ПРООН как атрибутивные переменные.

Для анализа данных используем метод многомерного шкалирования по индикаторам системы. Процесс шкалирования предполагает построение шкал абсолютного индекса развития системы и эмпирических индексов развития:

$$T = a \cdot \frac{\rho_1 \cdot \rho_2 \cdot \dots \cdot \rho_n}{\rho_{10} \cdot \rho_{20} \cdot \dots \cdot \rho_n} = a \cdot \xi_1 \cdot \xi_2 \cdot \dots \cdot \xi_n, \quad t_j = b \cdot \frac{\rho_j^*}{\rho_{j0}^*} = b \cdot \xi_j^*. \quad (1)$$

Здесь T – абсолютный индекс развития системы; t_j – эмпирические индексы развития; a, b – постоянные шкалирования; ρ_k – геометрическая вероятность распределения некоторого индикатора, принятого в качестве атрибутивной переменной; ρ_j^* – геометрическая вероятность распределения индикатора, который не является атрибутивной переменной и отражает тенденции развития для некоторого компонента развития системы ($j \leq m$); ρ_{k0}, ρ_{j0}^* – геометрические вероятности распределения индикаторов, которые соответствуют некоторому опорному состоянию; m – число переменных, отражающих различные компоненты развития.

Для одномерной случайной величины геометрическая вероятность находится согласно уравнения:

$$\rho_k = \frac{I_k - I_{\min}}{I_{\max} - I_{\min}}; \quad 0 \leq \rho_k \leq 1, \quad (2)$$

где I_k – некоторый индикатор; I_{\max}, I_{\min} – соответственно максимальное и минимальное значение данного индикатора в изучаемой группе стран (в данном

классе объектов, наблюдаемых в опыте) за некоторый период, например, 2004 год или за некоторый диапазон времени, например, 1980 – 2010 гг.

Построим шкалы атрибутивных переменных. С этой целью база данных индикаторов нормируется путем выбора опорного состояния (базового объекта с заданными в выбранный момент времени индикаторами). Учитывая специфику данной статьи в качестве опорного состояния системы приняты индикаторы развития Украины в 2004 году. Параметры опорного состояния в дальнейшем будем обозначать дополнительным индексом «0». В результате имеем безразмерные переменные вида $\xi_k = \rho_k / \rho_{k0}$, где ρ_{k0} – геометрические вероятности распределения индикаторов, которые соответствуют опорному состоянию.

Цель построения абсолютного индекса – создать моделирующую среду (построить поле величины), которая даст возможность разработать аналитическую модель системы.

В результате данного подхода строится среда моделирования в виде пространства координат Ω , где координатные оси соответствуют шкалам атрибутивных переменных $\xi_1, \xi_2, \dots, \xi_n$ для изучаемой базы данных. На практике мы обычно имеем ограниченное количество статистических точек, отражающих данные наблюдений о параметрах развития стран мира. Статистическая база данных опытных точек $M_i \{ \xi_1, \xi_2, \dots, \xi_n \}$ отражает в пространстве Ω некоторую область Ω_n , которая определена наблюдаемыми значениями атрибутивных переменных. Область Ω_n будем рассматривать как общее пространство наблюдаемых состояний системы. Каждой точке $M \{ \xi_1, \xi_2, \dots, \xi_n \}$ данного пространства поставим в соответствие значение абсолютного индекса $T, (T_j)$, которое находится согласно уравнения (1) для величины T . В свою очередь, каждой опытной точке $M_i \{ \xi_1, \xi_2, \dots, \xi_n \}$ также может быть поставлен в соответствие определенный набор значений абсолютных $T, (T_j)$ и эмпирических t_j индексов развития. Таким образом, Ω_n – многомерное пространство точек M , в свою очередь, $T = T(M)$, $T_j = T_j(M)$ – непрерывное скалярное поле абсолютного индекса системы, которое мы будем называть, по аналогии с термодинамикой, полем идеальных состояний (объектов) системы. На первом этапе анализа данных задача сводится к изучению возможности построения регрессионных зависимостей для описания скалярного поля величины T , исходя из связи для каждого компонента системы эмпирического и абсолютного индексов, т.е. $\varphi_j(t_j) = T_j$. Таким образом, может быть построен набор эталонных шкал абсолютного индекса системы $\varphi_j(t_j) = T_j$, которые характеризуют множества различных состояний системы. Построение системы из m уравнений вида $\varphi_j(t_j) = T_j$ указывает на наличие значимых связей в изучаемой базе данных.

Для построения общей теории развития данного класса систем сделаем предположение, что в пространстве состояний Ω_n процессы развития стран мира описываются многомерными непрерывными кривыми, соединяющими

между собой различные состояния. Однако, процессы развития определенной страны, которые могут осуществляться между наблюдаемым состоянием M и любым другим возможным состоянием в области Ω_n , будут отличаться между собой по интенсивности взаимодействия объекта с глобальным социально-экономическим, биосферным, информационным и другим окружением, т.е. с глобальной окружающей средой. Для того, чтобы иметь возможность характеризовать взаимодействие системы с окружающей средой для каждого компонента системы введем некоторую переменную величину W_j , непосредственно связанную с опытными данными. Данная величина характеризует реальные процессы развития стран мира в пространстве многомерных переменных Ω_n и определяет интенсивность взаимодействия системы с окружающей средой. Будем называть величину W_j статистическим индексом развития системы по выбранному компоненту системы. Другими словами, мы принимаем гипотезу о существовании множества общих мер для различных форм материального движения и взаимодействия социальных систем, которыми являются страны мира. Далее для упрощения индекс j при определении величин T и W будем опускать, имея в виду, что все сказанное выше справедливо как для системы в целом, так и для каждого ее компонента. Для обоснования величины W сделаем предположение, что она связана со статистическими вероятностями наблюдаемых состояний системы.

Будем также предполагать, что функция $W = W(M)$ имеет непрерывные частные производные по всем переменным. Из теории известно, что при справедливости указанных выше условий изменение статистического индекса dW в окрестности любой точки M при условии осуществления некоторого процесса l может быть представлено в виде:

$$dW = W_1 d\xi_1 + W_2 d\xi_2 + \dots + W_n d\xi_n, \quad (3)$$

где $\xi_1, \xi_2, \dots, \xi_n$ – атрибутивные переменные системы; W_1, W_2, \dots, W_n – функции этих переменных, причем $W_k = \partial W / \partial \xi_k$.

Будем считать, что в окрестности любой точки M при бесконечно малом изменении состояния социальной системы в произвольном процессе l существует линейная связь между индексами W и T , которую можно представить в виде отношения элементарного приращения индекса W к соответствующему приращению индекса T в процессе l :

$$c_l = dW_l / dT_l. \quad (4)$$

Величина c_l определяется на основе опыта и зависит как от положения точки $M(\xi_1, \xi_2, \dots, \xi_n)$, так и от направления процесса развития системы в пространстве состояний Ω_n .

Согласно уравнений (3) и (4) в окрестности точки M имеем следующие соотношения:

$$dW = c_1 \frac{\partial T}{\partial \xi_1} d\xi_1 + c_2 \frac{\partial T}{\partial \xi_2} d\xi_2 + \dots + c_n \frac{\partial T}{\partial \xi_n} d\xi_n. \quad (5)$$

Если учесть уравнение (1), то из зависимости (5) при условии существования аналитической зависимости для абсолютного индекса T получим:

$$ds = \frac{dW}{T}; \quad ds = c_1 \frac{d\xi_1}{\xi_1} + c_2 \frac{d\xi_2}{\xi_2} + \dots + c_n \frac{d\xi_n}{\xi_n}, \quad (6)$$

где величину s согласно (6) определим как *энтропию* системы.

Для установления связей модели с показателями развития стран выберем различные индикаторы, отражающие состояние основных компонентов системы, а именно: удельное потребление энергии; младенческую и детскую смертность; фертильность; случаи заболевания туберкулезом и ВИЧ; количество врачей на 1 тыс. человек; количество пользователей сети Интернет и абонентов сотовых сетей на 1 тыс. человек и т.д.

Для построения системы измерения индекса T установим корреляционные связи между абсолютным индексом системы, зависящим от ее атрибутивных переменных, и всеми эмпирическими индексами развития t_j , отражающими различные аспекты развития системы. Среди полученных уравнений выберем одну или несколько наиболее значащих зависимостей и построим шкалу (шкалы) измерений. Общий подход несколько подобен построению шкал температур в термодинамике. Результаты анализа данных приведены в таблице 1.

Таблица 1 – Зависимости различных индикаторов от абсолютного индекса развития стран мира в 2004 году

Индикаторы	Регрессионное уравнение Вида $T = t_j$	Атрибутивные переменные	Коэф. корреляции
Удельное потребление энергии, t_e	$T = 1,514 \cdot t_e^{1,178}$	ρ_2, ρ_4, ρ_5	0,92
Младенческая смертность, $t_{ml,s}$	$T = 1,400 \cdot \exp(-0,233 \cdot t_{ml,s})$	ρ_1, ρ_4	0,91
Абоненты сотовых телефонов, t_{mob}	$T = 0,846 \cdot t_{mob}^{1,208}$	ρ_4, ρ_5	0,90
Количество врачей, t_{doc}	$T = 3,279 \cdot t_{doc}^{1,714}$	$\rho_1, \rho_2, \rho_3, \rho_4, \rho_5$	0,91
Пользователи Интернет, t_{int}	$T = 0,750 \cdot t_{int}^{1,140}$	ρ_1, ρ_4, ρ_5	0,89

Следующим шагом разрабатывается система оценки значений статистического индекса W , что позволяет совместно с уравнением (4) найти коэффици-

енты c_k для каждого компонента в процессе ее развития, исходя из среднестатистических тенденций. Для этого оценку значений индекса W будем проводить путем определения вероятности состояния системы в пространстве Ω_n на основании статистических данных. Определим статистическую вероятность состояния системы w исходя из относительных частот событий, которые связаны с распределениями атрибутивных переменных в массиве опытных данных, когда атрибутивные переменные наблюдаются в совокупности одновременно. Алгоритмы определения вероятности в этом случае предложены в работе [2].

В окрестности любой точки M свяжем статистический индекс системы W теоретической линейной зависимостью со статистической вероятностью w :

$$W = \alpha \cdot w/w_0, \quad (7)$$

где w_0 – вероятность состояния системы для условий принятого опорного состояния; α – некоторый коэффициент пропорциональности, который можно принять равным единице. Для построения системы оценки статистического индекса W установим связи индекса с энтропией системы, которая согласно (6) может быть представлена в виде:

$$s = c_1 \cdot \ln\left(\frac{\rho_1}{\rho_{10}}\right) + c_2 \cdot \ln\left(\frac{\rho_2}{\rho_{20}}\right) + \dots + c_n \cdot \ln\left(\frac{\rho_n}{\rho_{n0}}\right). \quad (8)$$

Здесь принято, что $s_0 = 0$ для условий опорного состояния.

Таким образом, обработка данных о развитии стран мира ведется путем определения значения энтропии s по атрибутивным переменным согласно уравнение (8) и установления значений величин c_k по известным данным статистической вероятности w , для событий наблюдаемых в совокупности значений атрибутивных переменных. Согласование теоретической модели с опытными данными сводится к построению регрессионного уравнения вида:

$$s = \Psi(W), \quad (9)$$

где также принято, что $s_0 = 0$ для опорного состояния.

Данное уравнение является аналогом уравнения Больцмана в физике $s = k \cdot \ln W$ применительно к изучаемой задаче, при этом вид функции Ψ определяется исходя из особенностей изучаемой системы.

Следующим этапом для каждой страны следует определить функции состояния системы (потенциал P и энтропию s), которые выступают в качестве обобщенных переменных. Функции состояния определяют многомерные криволинейные координаты поля величины W , при этом каждая страна в пространстве своего развития в процессе Ω_n будет занимать некоторое положение относи-

тельно этих координат. Это позволяет объективно определить ранг страны в иерархическом множестве других стран по различным компонентам системы. Энтропия будет определять направление развития страны относительно опорного состояния, потенциал – принадлежность точки некоторой поверхности уровня, ортогональной линиям энтропии, при условии определенной вероятности состояния системы (постоянной энтропии). При известных абсолютном индексе T и энтропии системы s определяется уравнение для всего массива опытных данных – аналог уравнения сохранения энергии [2]:

$$T \cdot ds = du + \sum_{k=1}^{n-1} \alpha_k \cdot \xi_1 \cdot \dots \cdot \xi_{k-1} \cdot \xi_{k+1} \cdot \dots \cdot \xi_n d\xi_k, \quad (10)$$

где коэффициенты α_k находятся путем регрессионного анализа, а $du = c_n dT$.

В свою очередь, согласно [2] потенциал страны относительно опорного состояния будет равен:

$$P = \frac{\xi_1^2 - 1}{c_1} + \frac{\xi_2^2 - 1}{c_2} + \dots + \frac{\xi_n^2 - 1}{c_n}, \quad (11)$$

где принято, что величина P равна нулю для опорного состояния. Потенциал является наиболее удобной величиной для обобщенного определения ранга страны при ее развитии в пространстве Ω_n .

После построения общей модели развития системы определяется статус Украины, который комплексно характеризует ее состояние среди стран мира. Проводится также сравнение положения страны относительно некоторых индикативных объектов, в качестве которых взяты страны, имеющие определенный интерес при проведении оценки. На заключительном этапе анализа определяются рейтинги для каждой страны (ранг развития страны), проводится визуализация актуальных данных, делаются выводы и разрабатываются предложения для оптимизации процесса развития.

Изложение результатов оценки. Покажем общую процедуру построения системы измерения человеческого развития на примере, когда в качестве индикатора эмпирической шкалы принимают удельное потребление энергии странами мира. С этой целью на основе опытных данных для стран мира определим вероятность многомерного распределения, когда совместно наблюдаются значения атрибутивных переменных ρ_2, ρ_4, ρ_5 (табл. 1). После этого свяжем значение установленной вероятности совместных событий со значениями атрибутивных переменных ρ_2, ρ_4, ρ_5 , отнесенных, в свою очередь, к величинам $\rho_{20}, \rho_{40}, \rho_{50}$, которые соответствуют опорному состоянию. Это позволит нам согласно (8) определить функцию энтропии системы и величины c_k на множестве опытных данных, характерных для стран мира. В связи с тем, что многомерная нелинейная функция вероятности описывается приближенной зависимостью, то в

совокупности будет наблюдаться определенный разброс опытных точек (рис. 1).

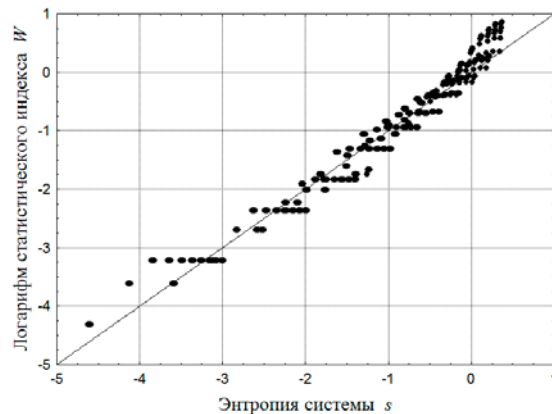


Рисунок 1 – Зависимость статистического индекса (W) от энтропии системы (s) для распределения совместно наблюдаемых значений атрибутивных величин (ρ_2, ρ_4, ρ_5) для 177 стран мира

Зная значения абсолютного индекса T и величин c_k , легко определить энтропию системы s и уравнение вида (10). Для этого находится регрессионная зависимость между переменными. Результаты обработки данных приведены на рис. 2, а соответствующее уравнение в дифференциальной форме имеет вид:

$$T \cdot ds = du + 1,2938 \cdot \xi_4 \cdot \xi_5 \cdot d\xi_2 - 1,1834 \cdot \xi_2 \cdot \xi_5 \cdot d\xi_4, \quad (13)$$

где $ds, du, d\xi_2, d\xi_4$ – приращения величин относительно опорного состояния. Коэффициент корреляции для уравнения (13) составляет 0,985, средняя относительная ошибка – 11%.

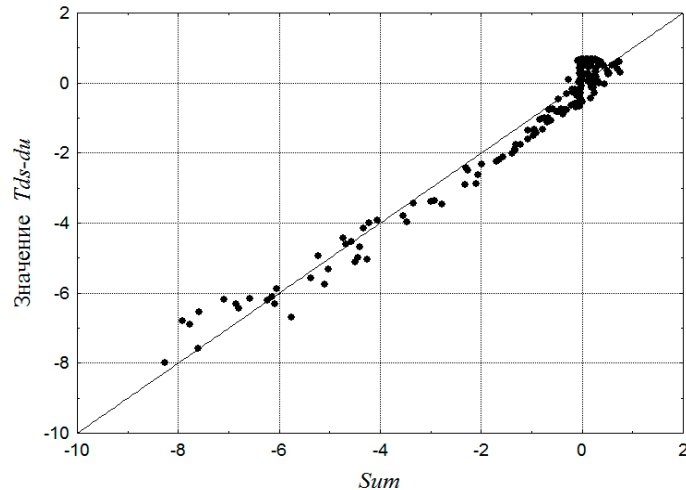


Рисунок 2 – Обработка данных для уравнения (10) относительно опорного состояния:

$$Sum = 1,2938 \cdot \xi_4 \cdot \xi_5 \cdot d\xi_2 - \dots - 1,1834 \cdot \xi_2 \cdot \xi_5 \cdot d\xi_4$$

На основе полученных данных определены энтропия и потенциал для каждой страны мира, исходя из оценки событий, связанных с удельным потреблением энергии. Результаты ранжирования стран мира для данного случая приве-

дены в таблице 2. Аналогичным образом, для примера, проведено ранжирование стран по факту событий, связанных с уровнем младенческой смертности и количеством абонентов сотовых телефонов, при этом была установлена очень высокая устойчивость измерений по всем трем эмпирическим шкалам.

Результаты анализа развития стран по различным индикаторам указывают на то, что, несмотря на различия в потенциалах развития, наблюдается высокая устойчивость рангов стран, которые практически очень мало отличаются от значений приведенных в таблице 2 для оценок по различным шкалам. Полученные результаты позволяют предложить объективный метод оценки развития стран мира и построить систему прогнозирования их параметров и индикаторов. В целом суть метода основывается на гипотезе существования абсолютно-го индекса $T = T(\xi_1, \xi_2, \dots, \xi_n)$ и связи скалярного поля этого индекса с опытными данными по распределению статистической вероятности состояния системы. При справедливости этой гипотезы в пространстве наблюдаемых состояний системы Ω_n можно построить криволинейные координаты, которые определяют некоторое поле направлений, отражающее среднестатистические тенденции в развитии системы. Все это позволяет предложить методику оценки человеческого развития как альтернативу известной методике ПРООН. Особенностью методики является использование объективного подхода и отсутствие необходимости построения экспертных шкал для оценки. Результаты дают также возможность разработать методы, позволяющие получить ответы на ряд актуальных вопросов. Например, может ли Украина за десять лет попасть в тридцать наиболее развитых стран мира? Какой выбрать эффективный путь развития страны? По каким компонентам страна имеет наиболее низкие (высокие) темпы развития, и как эти темпы сопоставимы с темпами развития отдельных стран?

При условии индикативной оценки развития стран по факту потребления энергии на вопрос о возможности войти Украине в число 30 наиболее развитых стран мира можно ответить отрицательно.

Несмотря на наличие в стране атомной энергетики (4 станции), более 25 тепловых электростанций и 10 гидроэлектростанций уровень развития Украины в 2004 году по удельному показателю потребления энергии был невысокий – всего 85 ранг. Кроме того, в 2004 – 2008 годах 108 стран имели более высокие темпы развития. Ранг Украины при оценке развития по данному показателю даже ниже, нежели ранг человеческого развития (77 ранг) по методике ПРООН. В 2004 – 2008 годах самые высокие темпы развития имели Катар, ОАЭ, Кувейт, Норвегия и Сингапур, в свою очередь, самые низкие темпы – Исландия, Ирландия, Кипр, Тонга и ЮАР. За этот период Украина по темпам развития занимала 109 место и существенно отставала от России и Белоруссии, а также от большинства стран бывшего СССР.

Таблица 2 – Значения энтропии (S) и потенциала (P) и рейтинги стран мира при построении шкалы оценки по факту потребления энергии

Страны мира	Значения в 2004 году		Ранги стран согласно предложенной методики		Ранги стран по ИЧР согласно методике ПРООН (2004 г.)
	S	P	уровень развития (2004 г.)	темпы развития (2004– 2008 гг.)	
Норвегия	0,1280	1382,7	4	5	1
Исландия	0,2681	1008,3	5	153	2
Швеция	0,1905	795,1	16	15	5
Канада	0,1521	896,6	10	14	6
Япония	-0,0067	778,3	18	21	7
США	0,1606	1475,9	2	13	8
Франция	0,1156	781,3	17	24	16
Италия	0,0038	718,8	20	148	17
Англия	0,2406	870,3	13	30	18
Германия	0,1004	725,7	19	16	21
Греция	-0,1566	428,4	27	22	24
Словения	-0,2247	375,7	28	25	27
Чехия	0,0785	316,9	34	44	30
Венгрия	-0,0128	226,9	39	145	35
Польша	-0,0666	118,3	48	29	37
Эстония	0,0381	159,6	44	49	40
Литва	0,0030	121,7	47	61	41
Словакия	-0,1281	161,3	43	19	42
Латвия	0,0174	87,84	54	68	45
Болгария	0,0077	22,24	65	50	54
Румыния	-0,2276	28,00	63	43	60
Россия	0,0700	52,60	59	31	65
Македония	-0,0547	2,511	82	55	66
Беларусь	0,0536	7,08	79	34	67
Бразилия	-0,0364	24,31	64	56	69
Украина	0,0000	0,00	85	109	77
Казахстан	-0,1346	12,85	74	54	79
Армения	-0,0494	-21,28	112	77	80
Китай	-0,6252	-6,06	91	71	81
Турция	-0,2427	17,35	70	35	92
Грузия	-0,2176	-28,75	119	74	97
Азербайджан	-0,2443	-21,17	110	48	99
Кыргызстан	-0,5896	-32,27	137	100	110
Египет	-1,0223	-21,09	109	70	111
Узбекистан	-0,5589	-32,44	139	86	113
Молдова	-0,3566	-32,65	145	83	114
Индия	-1,7407	-27,89	117	99	126
Пакистан	-2,0035	-31,71	132	94	134
Кения	-1,7646	-34,10	160	102	152
Нигерия	-1,0868	-33,84	156	93	159
Эфиопия	-3,4512	-34,55	176	116	170

В целом трудоемкость оценки развития стран определяется не методической сложностью метода, а необходимостью построения множества зависимостей для индикаторов, входящих в базу данных о развитии стран мира. Если база данных ПРООН содержит несколько десятков индикаторов, требующих анализа, то база данных Всемирного банка – это уже более тысячи индикаторов. На фоне 186 стран, данные о которых уже есть для 2010 года, видна явная необходимость автоматизации процесса вычислений и поиска закономерностей в базах данных. Это возможно только при создании IT-систем для анализа данных в области Global studies.

Таким образом, международные базы данных индикаторов в комплексе с методами интеллектуального анализа данных позволяют установить закономерности развития стран мира. Все это говорит о возможности создания общей математической теории оценки развития стран по логике построения аналогичной той, которая применяется сегодня в термодинамике.

СПИСОК ЛИТЕРАТУРЫ

1. Прогнозирование будущего. Новая парадигма. / Под ред. Г.Г. Фетисова и В.М. Бондаренко. - М.: Экономика, 2008. – 283 с.
2. Аверин, Г.В. Фундаментальные модели в общей теории систем: закон перехода количественных изменений в качественные // Збірник наук. праць ДонНТУ: Системний аналіз та інформаційні технології в науках про природу та суспільство. - Донецьк: ДонНТУ, №1(2) – 2(3), 2012. – С. 6 – 52.

REFERENCES

1. *Prognozirovaniye budushchego. Novaya paradigma* [Forecasting of the future. New Paradigm] (2008), G.G. Fetisov and V.M. Bondarenko. Eds., Economica, Moscow, Russiya.
2. Averin, G.V. (2012), «Fundamental models in common theory of systems: the law of quantity to quality transition», *Zbirnik naukovirh prac DNTU: Sistemniy analiz ta informatsiyini tekhnologii v naukakh pro prirodu ta suspilstvo*, no. 1(2) – 2(3), – pp. 6 – 52.

Об авторах

Аверин Геннадий Викторович, доктор технических наук, профессор, заведующий кафедры компьютерных систем мониторинга Государственного высшего учебного заведения «Донецкий национальный технический университет» (ГБУЗ «ДонНТУ»), Донецк, Украина, averin@donntu.edu.ua

Звягинцева Анна Викторовна, кандидат технических наук, доцент, доцент кафедры компьютерных систем мониторинга Государственного высшего учебного заведения «Донецкий национальный технический университет» (ГБУЗ «ДонНТУ»), Донецк, Украина, anna_zv@ukr.net

About the authors

Averin Gennadiy Viktorovich, Dr.Sc., professor, head of the computer systems for monitoring department of Donetsk National Technical University, Donetsk, Ukraine, averin@donntu.edu.ua

Zviagintseva Anna Victorovna, PhD, associate professor with the department of computer systems for monitoring of Donetsk National Technical University, Donetsk, Ukraine, anna_zv@ukr.net

Анотація. Показана можливість застосування методів інтелектуального аналізу даних (ІАД), які відрізняються високим рівнем розуміння інформації, для вивчення соціально-економічного розвитку країн світу. В якості вихідної інформації використана для аналізу база даних Програми розвитку ООН. Побудована модель розвитку країн світу на основі застосування методу багатовимірного шкалювання за атрибутивними індикаторами системи та алгоритмів ІАД для обчислення статистичної ймовірності розподілу спостережуваних в сукупності показників розвитку країн. Запропоновано метод визначення ентропії та потенціалу розвитку об'єктів. Проведено обчислювальні експерименти і визначено ранги розвитку країн світу, а також ранги темпів розвитку. Виконана оцінка розвитку України в порівнянні з основними країнами світу. Отримані результати можуть бути використані при стратегічному прогнозуванні та плануванні розвитку України.

Ключові слова: інтелектуальний аналіз даних, моделі розвитку країн, стратегічна оцінка, ранги країн світу, темпи розвитку України.

Abstract. The possibility of applying data mining techniques (DM) is shown that differs by a high level of understanding of information for the study of the socio-economic development of countries. As the initial information the database of the United Nations Development Program was used for the analyzes. A model of countries' development was built using the method of multidimensional scaling on attribute indicators of the system and application of DM algorithms to calculate the statistical probability distribution of observed in the set of indicators of countries' development. A method for determining the entropy and potential of the development of objects was proposed. Computational experiments were carried out and countries' development ranks were determined as well as the ranks of the pace of development. The evaluation of the development of Ukraine was carried out in comparison with the major countries of the world.

The results obtained can be used for strategic planning and forecasting of the development of Ukraine.

Keywords: data mining, country development models, strategic assessment, countries' ranks, Ukraine development pace.

Стаття постуила в редакцію 25.02. 2013

Рекомендовано к публікації д-ром техн. наук Т.В. Бунько