Математическое моделирование

Тененев Валентин Алексеевич

профессор, доктор физико-математических наук

Системное моделирование

Задачи обработки данных с целью извлечения новых знаний сопровождают системное и математическое моделирование поведения объектов самой различной природы. Системный подход к анализу данных дает общую методологию обработки, независимо от природы объектов.

Классическое определение системы:
«система – совокупность элементов, организованных каким-либо образом и образующих целостность и органическое единство».

Элемент – предел разбиения системы с точки зрения аспекта рассмотрения, решения конкретной задачи, поставленной цели.

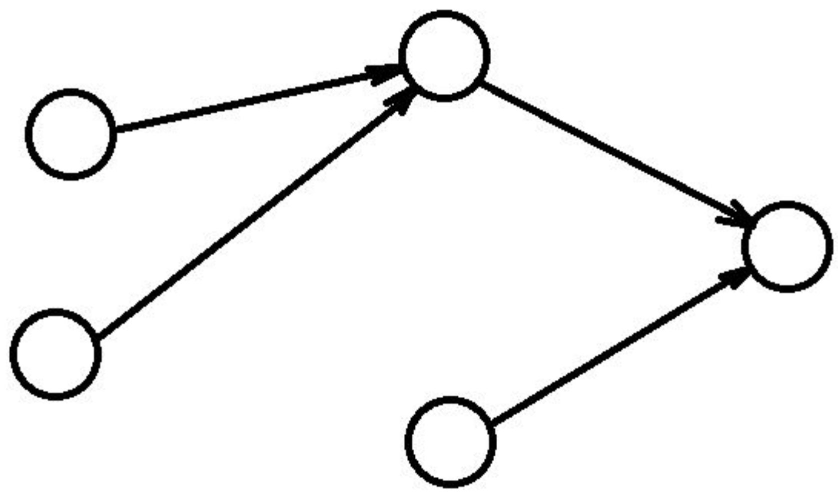
Связь определяют как ограничение степени свободы элементов.

Определение модели

Модель – созданная или выбранная исследователем система, воспроизводящая существенные для целей познания характеристики изучаемого объекта.
Исследование этой системы служит опосредованным Способом получения информации об этом объекте.

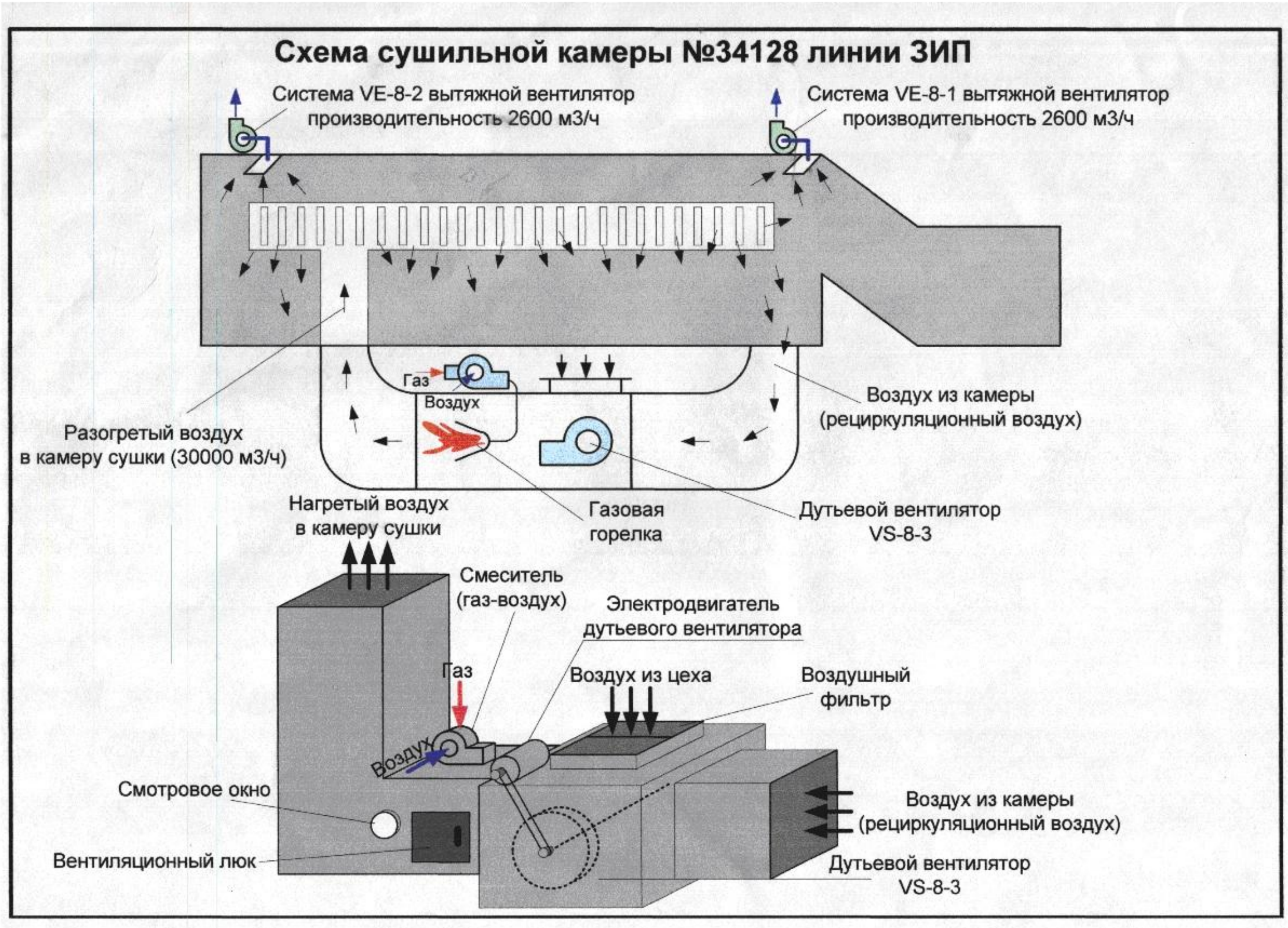
Моделирование – способ оперирования объектом, при котором исследуется не сам объект, а вспомогательная система, находящаяся с ним в объективном соответствии, и которая дает необходимую информацию.

Виды связей:
1. Алгебраические функции и выражения.
2. Дифференциальные и интегро-дифференциальные уравнения.
3. Логические правила и системы логического вывода.



Направления моделирования сложных систем

Численное моделирование взрывоопасной ситуации в сушильной камере



Модель венчурного инвестирования

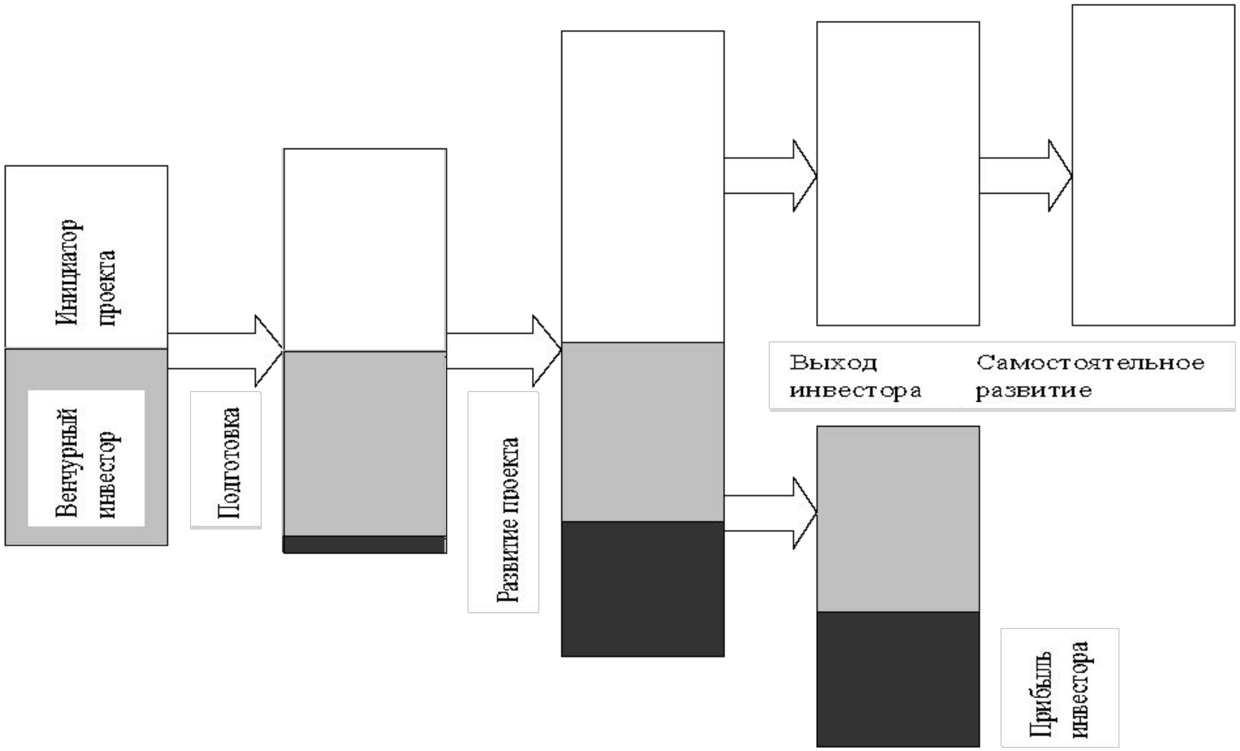


Схема венчурного финансирования

Модель развития венчурного инвестиционного проекта:

$$t = 0:$$

$$k(0) = \frac{Z_{IV}^{0}}{L} u_0, \quad q(0) = \frac{Z_{IV}^{0}}{L}(1 - u_0);$$

$$t \in (0, \tau_{TP}):$$

$$\frac{dk}{dt} = \frac{C(t - \tau_K, T_{t-\tau_K}) - Z}{L} u(t - \tau_K) - \mu k - B(t),$$

$$\frac{dq}{dt} = (1 - u(t)) \frac{C(t, T_t) - Z}{L},$$

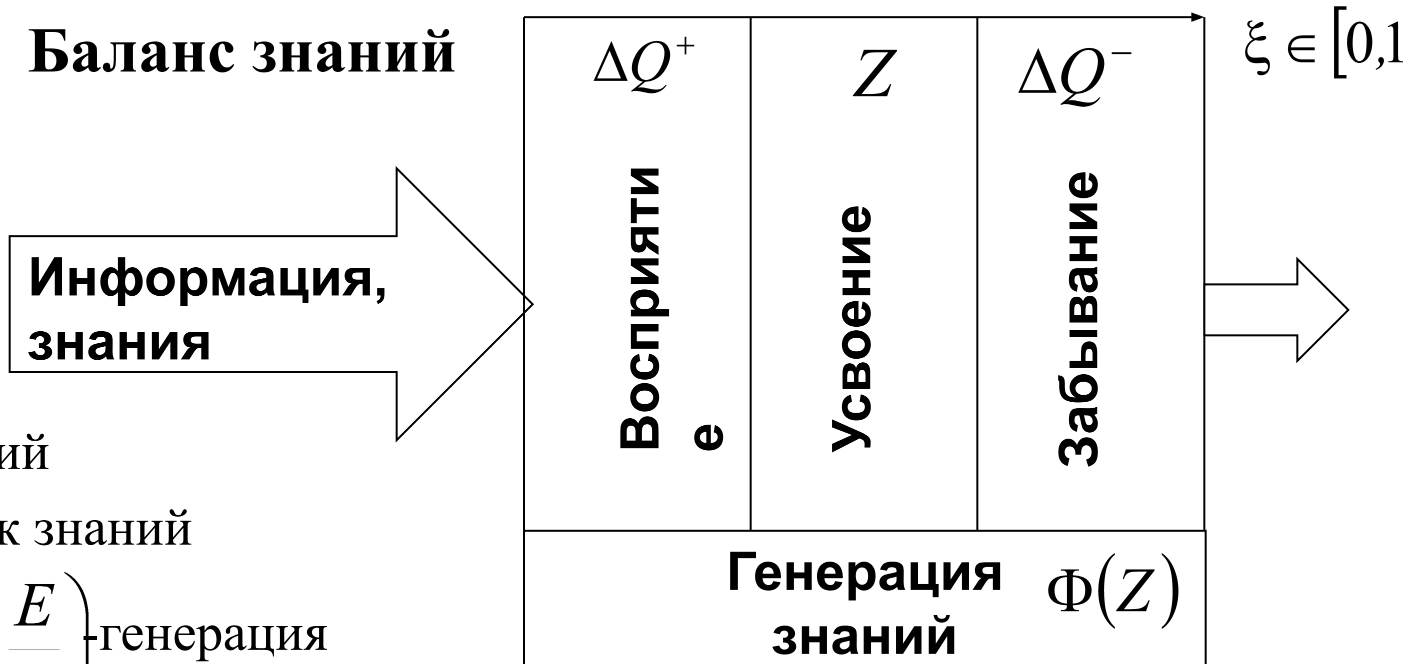
$$t \geq \tau_{TP}:$$

$$\frac{dk}{dt} = (1 - s) y(t - \tau_K)(u(t - \tau_K) - d(t)) + \frac{C(t - \tau_K, T_{t-\tau_K})}{L} u(t - \tau_K) - \mu k - B(t),$$

$$\frac{dq}{dt} = (1 - s)(1 - u(t)) y + (1 - u(t)) \frac{C(t, T_t)}{L},$$

Математическая модель процесса обучения

Баланс знаний



Z - уровень знаний

$Q \sim A\Delta Z$ - поток знаний

$\Phi(Z) = K\,exp\left(-\dfrac{E}{Z}\right)$ -генерация новых знаний

$\Delta Z = \left(Q^{+} - Q^{-}\right)\Delta t + \Phi(Z)\Delta t$ - баланс

При $\Delta t \to 0$

$$\frac{\partial Z}{\partial t} = A\frac{\partial^2 Z}{\partial \xi^2} + \Phi(Z) \tag{1}$$

Начальные условия:

$t = 0, \quad Z = Z_0$

.

Граничные условия:

$\xi = 0, \quad B_L\left(Z_{00} - Z\right) = -\Lambda\dfrac{\partial Z}{\partial \xi}$

$\xi = 1, \quad B_R\left(Z_N - Z\right) = \Lambda\dfrac{\partial Z}{\partial \xi}$

Динамическое моделирование ценовой политики в условиях конкуренции

Двухкритериальная задача оптимального управления

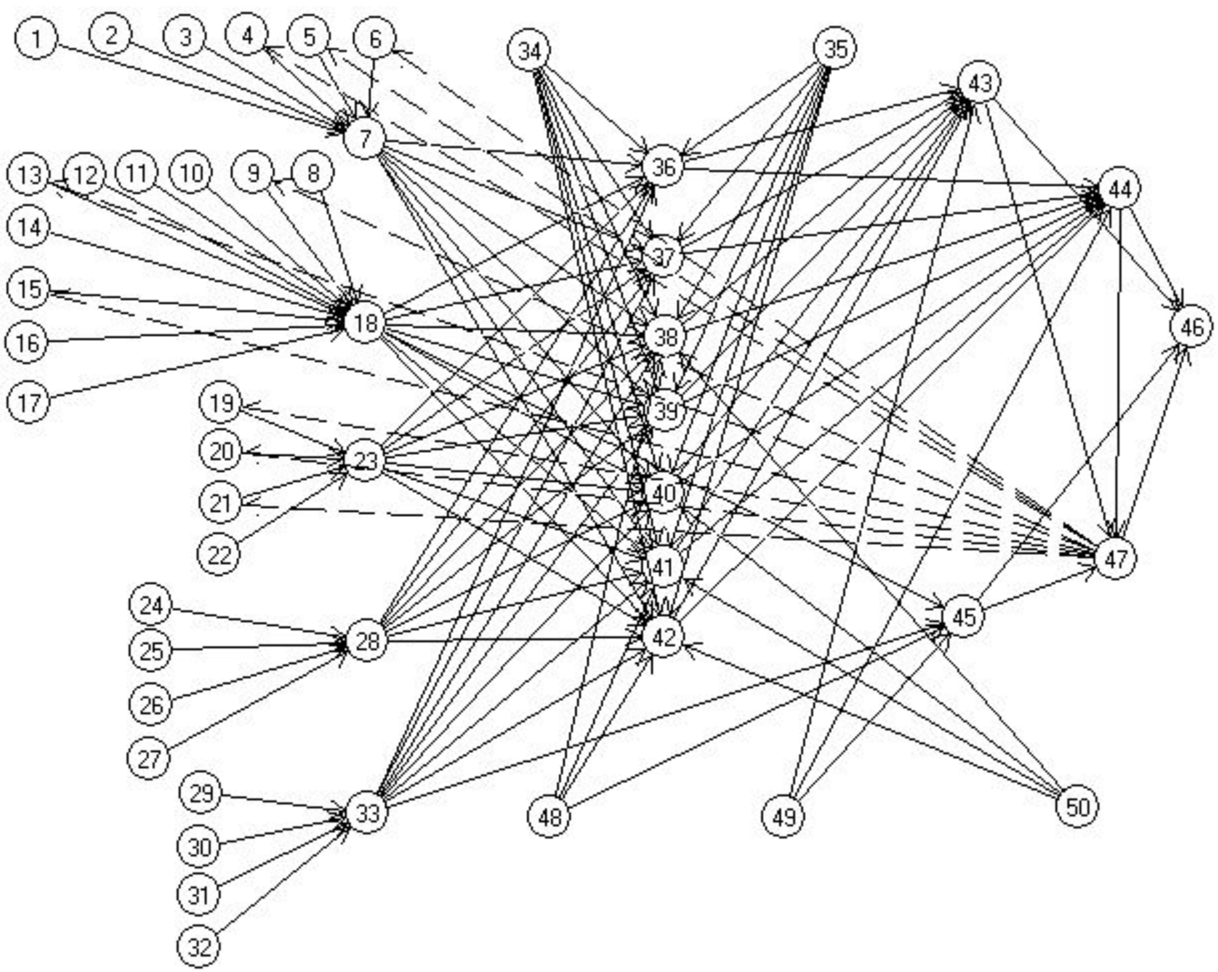
$$P_i(C_i) = \int_0^T x_i(t) C_i(t) dt \rightarrow \max, \quad i = 1,2$$

$$\frac{dx_1}{dt} = x_1 \left[V_1(C_1(t)) - a_{11} x_1 - a_{12} x_2 \right]$$

$$\frac{dx_2}{dt} = x_2 \left[V_2(C_2) - a_{21} x_2 - a_{22} x_1 \right]$$

$$x_1(0) = x_1^0, \quad x_2(0) = x_2^0$$

Нечеткая модель производства продукции



Структурная схема системы управления капиталом предприятия

Модель государственного управления экономикой

$$\frac{dx_i}{dt} = S_i^1 y_i - \mu_{xi} x_i$$

$$\frac{dq_i}{dt} = (S_i^2 + r_i) y_i - \mu_{qi} q_i$$

$$\frac{dQ_i}{dt} = (S_i^3 + R_i) y_i - \mu_{Qi} Q_i \quad i = \overline{1, n}$$

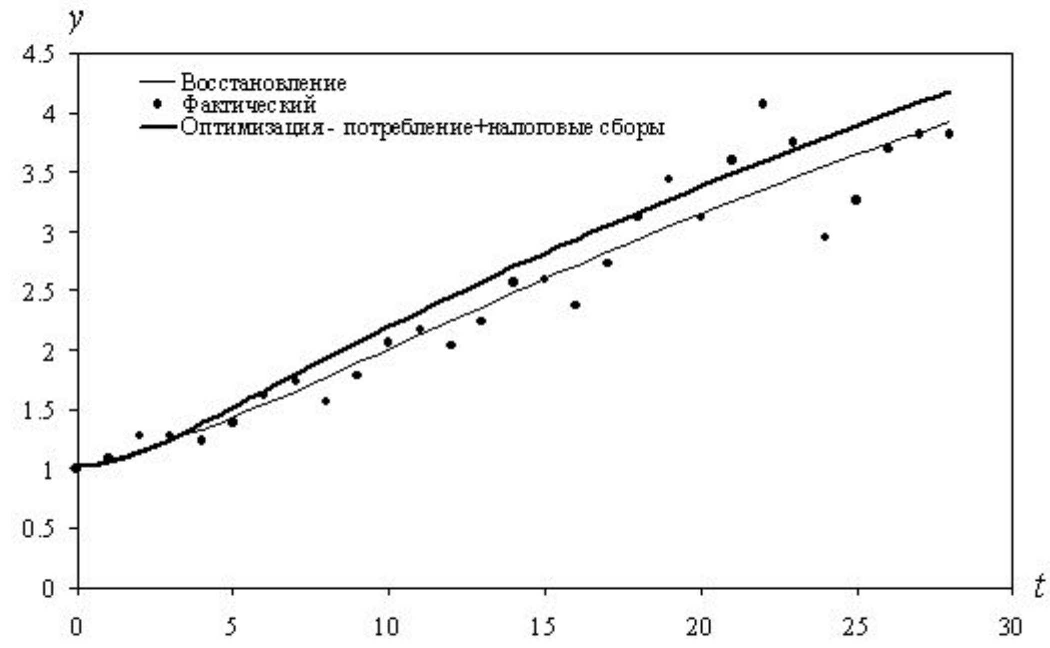
Дифференциальные уравнения, описывающие прирост производственных фондов, научно-образовательного потенциала и потенциала здоровья

$$\Phi_0 = \int_0^T \left(\sum_{i=1}^n H_i(t) y_i(t) - \sum_{i=1}^n r_i(t) y_i(t) - \sum_{i=1}^n R_i(t) y_i(t) \right) dt \to \max$$

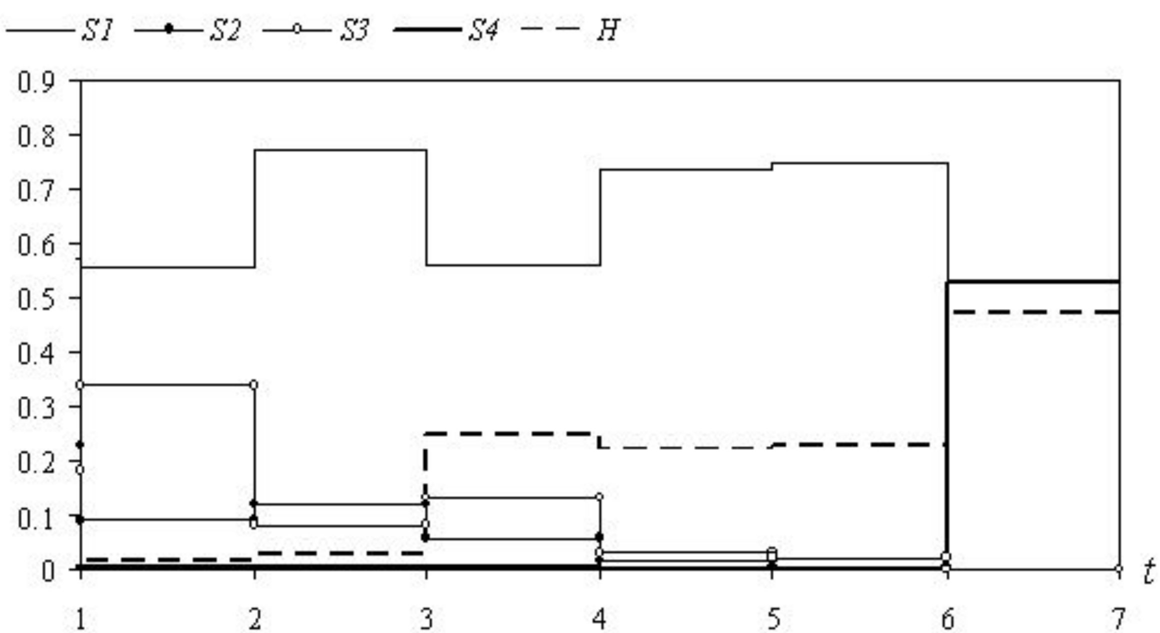
Целевым критерием государства является получение максимального налога за вычетом средств, направляемых на науку и здравоохранение

$$\Phi_i = \int_0^T S_i^4(t) y_i(t) dt \to \max$$

Для производственных элементов критерием является максимальное потребление



Изменение ВВП во времени



Изменение управляющих функций с =1 год

Продолжение временных рядов

Временной ряд или последовательность преобразуется в матрицу, с помощью сдвига по времени или лага длиной *k*.

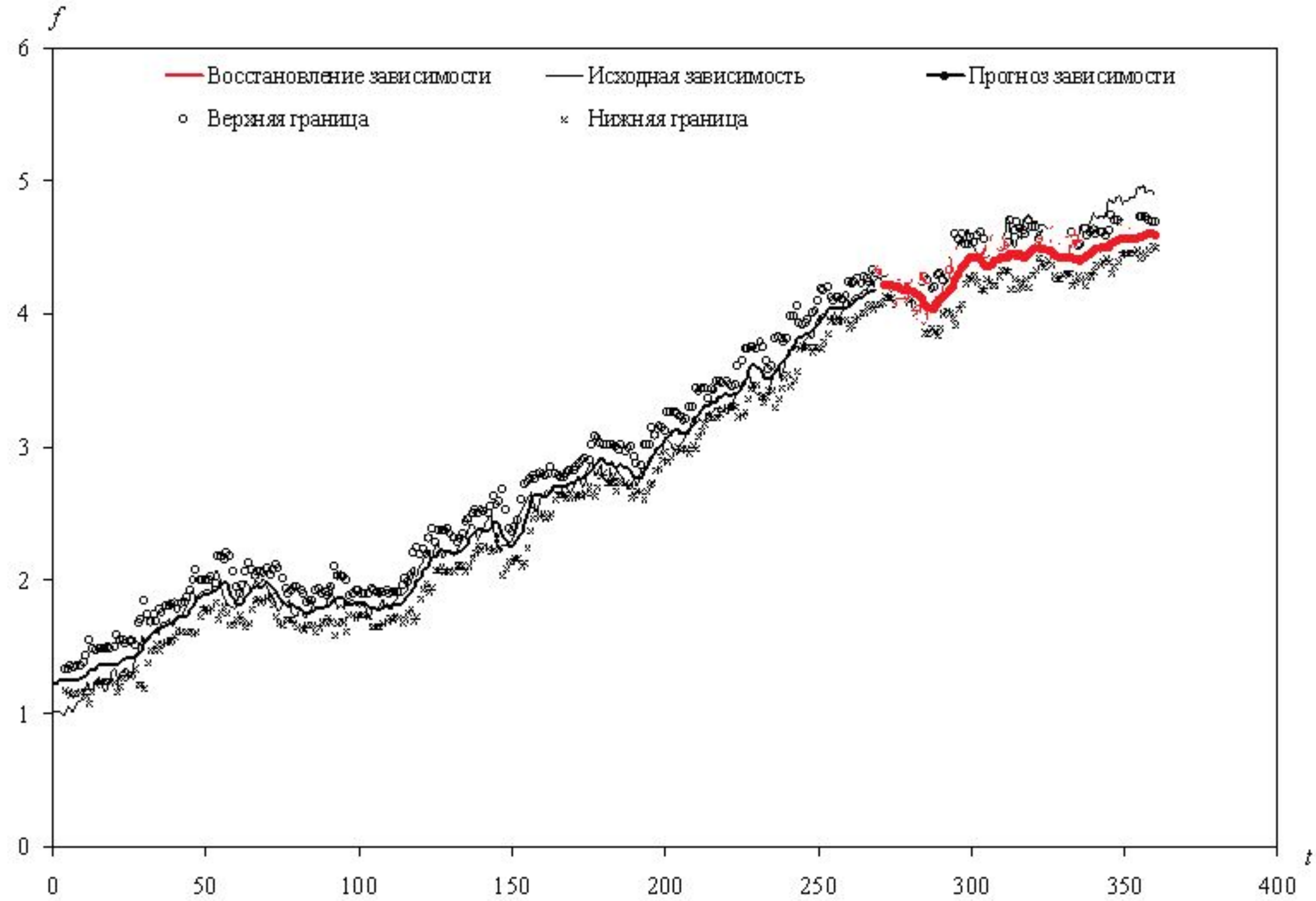
$$\left[\mathbf{Q}_j\right]=\begin{bmatrix} f_0 & f_1 & \cdots & f_{N-k+1} \\ f_1 & f_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ f_{k-1} & f_k & \cdots & f_N \end{bmatrix}, j=\overline{0,p};\ p=N-k+1$$

Последовательность временного ряда переведена в набор обучающих данных, $(\mathbf{x},y)^j,\ j=\overline{0,p}$

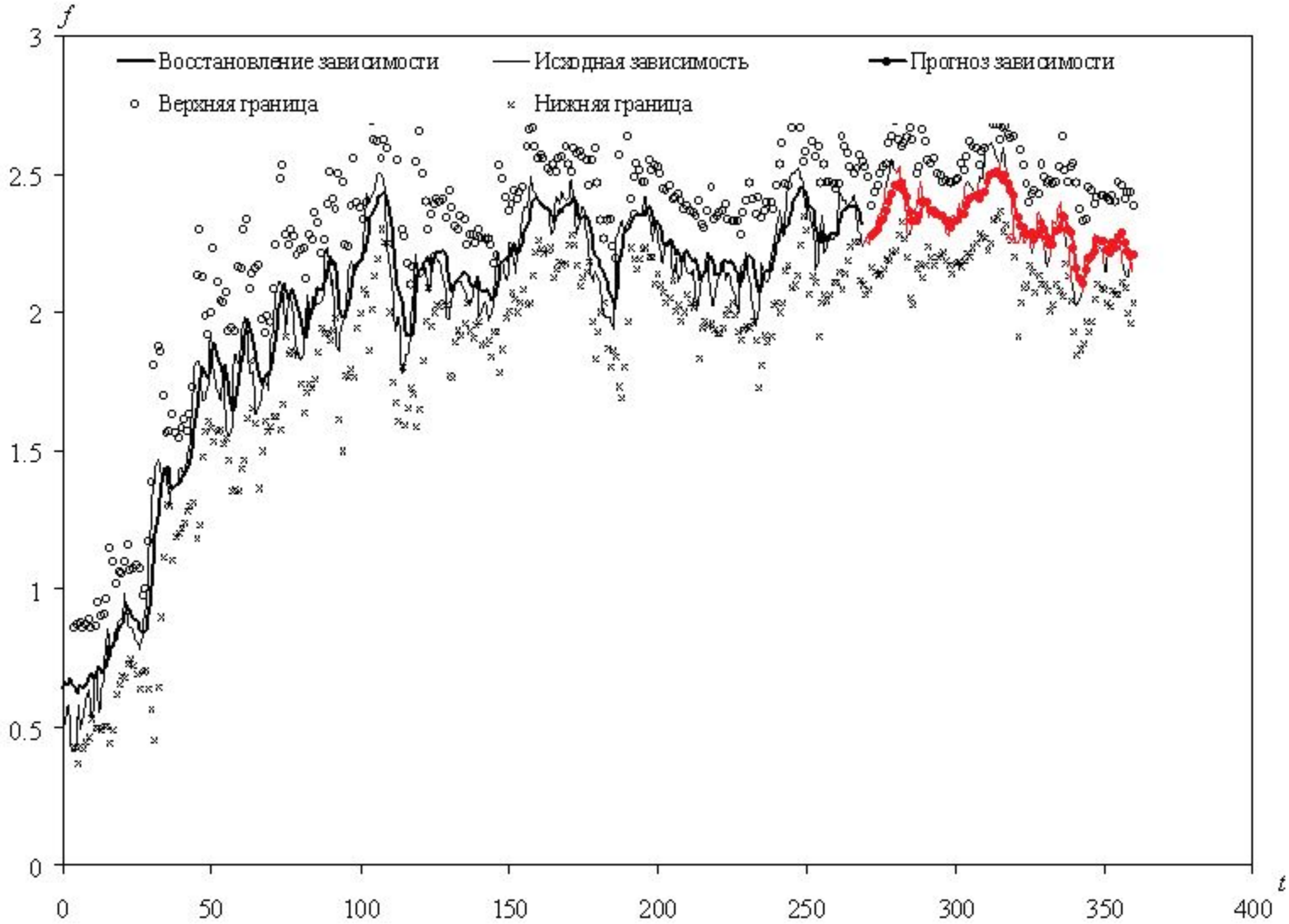
содержащий ***p*** точек.

$$\mathbf{x}^j=\begin{pmatrix} f_j \\ \vdots \\ f_{j+k-2} \end{pmatrix} \qquad y=f_{j+k-1}$$

$$\nu_i^2=\frac{1}{(k-1)}\sum_{j=1}^{k}\left(f_{i-j}-f\left(\mathbf{x}_{i-j}\right)\right)^2, i=\overline{k,N}$$ дисперсия



Восстановление временного ряда поступлений вкладов (лаг =5)



Восстановление временного ряда выдачи (лаг =5)

Задача скоринга

Для обработки использовался набор данных, состоящий из 10000 записей с 17 полями (атрибутами). 16 атрибутов представляют вектор входных данных. К ним относятся:

размер запрашиваемого кредита;
срок кредита;
доход клиента;
характер работы;
рабочий стаж;
уровень образования;
место проживания;
продолжительность регистрации по месту жительства;
наличие в собственности квартиры или другой недвижимости;
наличие движимого имущества;
возраст клиента и его пол;
наличие и размер текущего счета в данном банке;
наличие бравшихся кредитов в данном банке.

Семнадцатый атрибут представляет выходную переменную и свидетельствует о своевременном или несвоевременном возврате кредитов.

Выходная переменная является лингвистической: «возврат в срок», «просроченный или неполный возврат», «невозврат».
Выходной переменной соответствуют три класса с номерами 0, 1, 2.

K	0	1	2
δ_{0k},%	99.51	0.49	0
δ_{1k},%	49.02	34.73	16.25
δ_{2k},%	4.49	11.54	83.97
N_k	4465	124	131
N_k	4487	357	156

Результаты классификации клиентов нейронной сетью

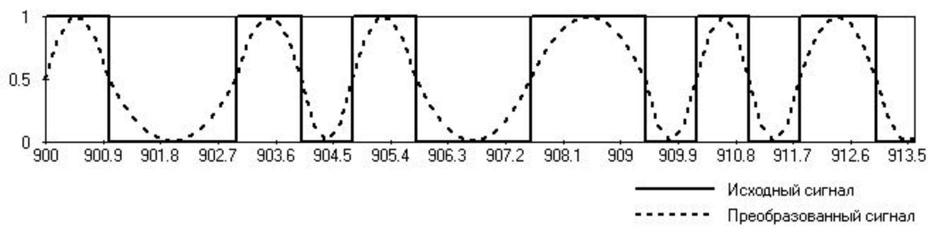
Задача интерпретации геофизических данных

Задача интерпретации геофизических данных

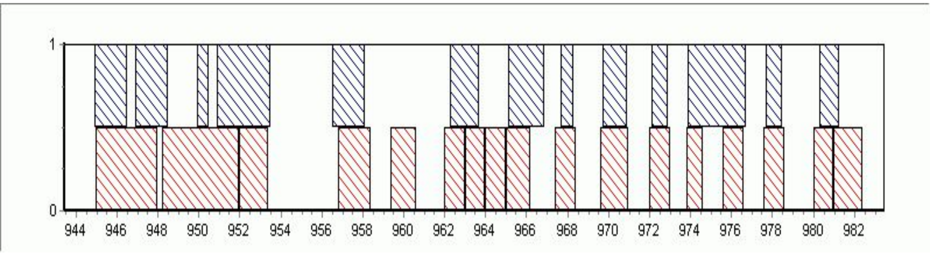
Разделение разреза на пласты нейросетевыми методами

$$\mathbf{X} = (w_{ij}^{k}, A_{j}^{k}, B_{j}^{k}, Y_{k}, K)$$

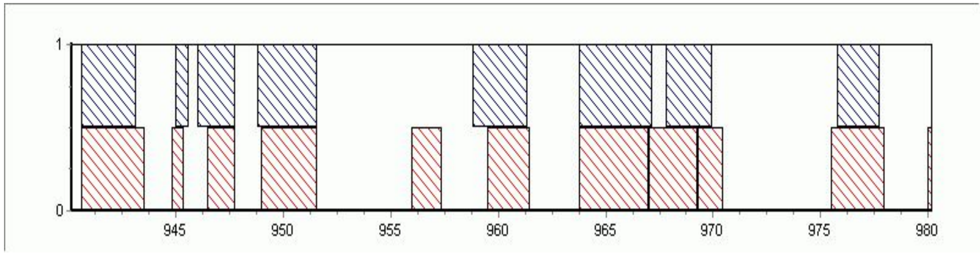
$$\Phi(\mathbf{X}) = \frac{n}{N} \Rightarrow \max$$ - оптимизационная функция для обучения



Кодирование выходного сигнала нейронной сети



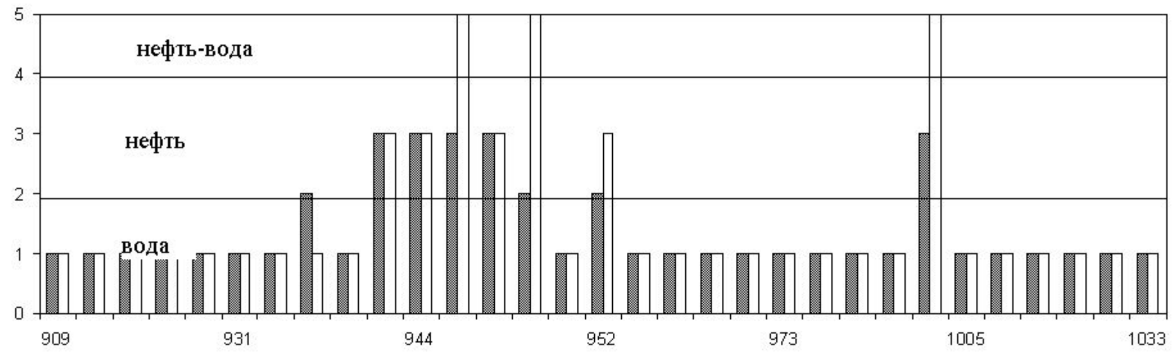
Скважина №13022

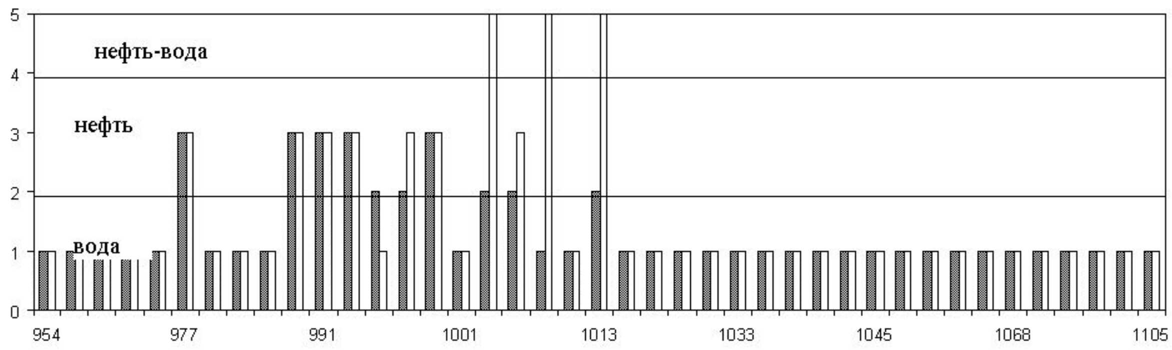


Скважина №13416

Разделение пластов нейросетевым методом.

Результаты прогнозирования продуктивности пластов





Прогнозу соответствуют темные столбики, расшифровке – светлые

От системы данных к системе с поведением

При исследовании объекта ему сопоставляется система, определяемая набором соответствующих свойств объекта. Каждому свойству соответствует определенная переменная. Переменной называется образ свойства объекта, определяемый конкретной процедурой измерения. Операционное представление, используемое для определения различий в наблюдении одного и того же свойства, называется параметром.

Переменные можно разделить на два типа: входные и выходные. Выходные параметры исходной системы рассматриваются как переменные, значения которых при соответствующих значениях параметров определяются внутри системы. Значения входных переменных задаются извне. Факторы, влияющие на определение входных переменных, называются средой системы.

Система, заданная на объекте, называется **системой объекта**. Она представляет множество свойств, с каждым из которых связано множество его проявлений, и параметрическое множество. **Исходная система** представляет связи с реальным миром, проходящие через систему объекта и **канал наблюдения**.

Наблюдение представлено в виде упорядоченной пары, состоящей из значения полного параметра, при котором было сделано наблюдение, и зафиксированного полного состояния переменных. Множество этих упорядоченных пар является функцией d, отображающей полное параметрическое множество в полное множество состояний. Такая функция представляет собой **данные.** Эта функция любому значению **полного параметра** ставит в соответствие одно **полное состояние переменных.**

Исходная система описывает потенциальные состояния переменных, а функция d дает информацию об их действительных состояниях. Соединение исходной системы и функции d образует систему более высокого уровня, называемую **системой данных.**

При системном анализе данных может быть два подхода: **вероятностный**, когда число наблюдений достаточно большое; **возможностный**, когда количество данных ограничено.

Находятся частоты состояний $N(c)$ для всех состояний $c \in C$.
Функция поведения, характеризующая рассматриваемую систему, для **вероятностного** подхода определяется как относительная частота появления состояния переменных $c \in C$ выражением

$$f_B(c) = \frac{N(c)}{\sum_{a \in C} N(a)} . \qquad (1)$$

Для **возможностного** подхода

$$f_B(c) = \frac{N(c)}{\max_{a \in C} N(a)} . \qquad (2)$$

Рассмотрим систему данных $D = (I, d)$ в виде таблицы $\mathbf{V} = [\mathbf{v}_j] = [v_{ij}], i = \overline{1, m}, j = \overline{1, n}$,

где $\mathbf{v}_j$ - вектор переменных системы, определенный на параметрическом множестве $\{j = \overline{1, n}\}$. Данные, полученные в абсолютной или интервальной шкале, переводятся в шкалу наименований, в которой значению переменной соответствует целое число $0, 1, \ldots, L_i - 1$. Преобразование осуществляется по формуле

$$u_{ij} = Int\left[\frac{v_{ij} - v_{i\min}}{\frac{v_{i\max} - v_{i\min}}{L} + \varepsilon}\right], \quad v_{i\min} = \min_j(v_{ij}), v_{i\max} = \max_j(v_{ij}),$$

где $[\cdot]$ - целая часть выражения.

Таблица 1. Исходные данные

v_1	v_2	v_3	u_1	u_2	u_3
0.979819	**0.306882**	**0.581423**	**1**	**0**	**1**
1.022144	**0.574559**	**0.027927**	**1**	**1**	**0**
0.391474	**0.152535**	**0.048503**	**0**	**0**	**0**
0.575403	**0.901264**	**0.56662**	**1**	**1**	**1**
1.371131	**0.707757**	**0.276064**	**1**	**1**	**0**
0.735222	**0.210558**	**0.392142**	**1**	**0**	**0**
0.929947	**0.865052**	**0.936275**	**1**	**1**	**1**
1.739056	**0.740511**	**0.315321**	**2**	**1**	**0**
1.032435	**0.693717**	**0.253766**	**1**	**1**	**0**
0.792377	**0.383503**	**0.388327**	**1**	**0**	**0**
0.795056	**0.429954**	**0.540299**	**1**	**0**	**1**
1.213448	**0.916344**	**0.968882**	**1**	**1**	**1**
1.901828	**0.949548**	**0.778201**	**2**	**1**	**1**
1.741663	**0.977377**	**0.65902**	**2**	**1**	**1**
1.343817	**0.392217**	**0.107393**	**1**	**0**	**0**
0.355132	**0.103262**	**0.194024**	**0**	**0**	**0**
0.2833	**0.075291**	**0.309124**	**0**	**0**	**0**
0.725762	**0.757983**	**0.438573**	**1**	**1**	**0**
0.820559	**0.00599**	**0.674281**	**1**	**0**	**1**
1.051059	**0.747567**	**0.758138**	**1**	**1**	**1**
1.20016	**0.136476**	**0.527675**	**1**	**0**	**1**
0.998835	**0.805844**	**0.498763**	**1**	**1**	**0**
0.936913	**0.070458**	**0.190571**	**1**	**0**	**0**
0.49803	**0.54446**	**0.091961**	**0**	**1**	**0**
0.754745	**0.781109**	**0.679843**	**1**	**1**	**1**
1.363118	**0.585441**	**0.788218**	**1**	**1**	**1**
1.190638	**0.219398**	**0.903279**	**1**	**0**	**1**
1.027162	**0.028367**	**0.124957**	**1**	**0**	**0**
0.618141	**0.958001**	**0.294943**	**1**	**1**	**0**
0.852273	**0.156659**	**0.838129**	**1**	**0**	**1**

Состояние c переменной описывается вектором $\mathbf{u}_k, k=\overline{1,N}, N \le n$. По формуле (1) или (2) вычисляется функция поведения системы. Величина $N(c)$ равна количеству одинаковых векторов $\mathbf{u}_k, k=\overline{1,N}, N \le n$.

В табл.2 представлены состояния $c=(u_1,u_2,u_3)$ и соответствующие частоты этих состояний $N(c)$. По формулам (1), (2) рассчитаны функции поведения для вероятностного и возможностного подходов, соответственно.

Таблица 2. Функции поведения

u_1	u_2	u_3	$N(c)$	Probability $f_B(c)$	Possibility $f_B(c)$
1	0	1	6	0.2	1
1	1	0	6	0.2	1
0	0	0	3	0.1	0.5
1	1	1	6	0.2	1
1	0	0	5	0.16667	0.833333
2	1	0	1	0.03333	0.166667
2	1	1	2	0.06667	0.333333
0	1	0	1	0.03333	0.166667

Целью исследования объектов и систем является получение новых знаний об их поведении в виде **зависимостей, правил, моделей.** Обнаружение в данных неизвестных нетривиальных знаний, необходимо для принятия обоснованных решений при управлении системами.

Следующим этапом после **системы данных** является построение **системы с поведением**. Поведение характеризует общие параметрически инвариантные **ограничения** на переменные представляющей системы. Описание параметрически инвариантного ограничения может быть использовано для **порождения** состояний переменных при данном параметрическом множестве. Системы с такими ограничениями называются **порождающими.**

Соседство на упорядоченном параметрическом множеств называется **маской** и определяется через **переменные, параметрическое множество и набор правил сдвига** на параметрическом множестве. Правило сдвига – это однозначная функция $r_j : W \to W$, которая каждому элементу W ставит в соответствие другой единственный элемент W. На полностью упорядоченном параметрическом множестве $r_j(w) = w + \rho$, где ρ – целая константа. **Выборочными переменными** называются переменные $S = \{s_1, s_2, ...\}$, вводимые с помощью уравнений $s_{k,w} = v_{i,r_j(w)}$. Множество всех выборочных переменных $V \times R$. Отношение $M \subseteq V \times R$ представляет схему соседства на упорядоченном множестве и является **маской**. Для полностью упорядоченных параметрических множеств маска может быть изображена в виде **вырезки** из матрицы $V \times R$.

Маска представляет **точку зрения**, в соответствии с которой представляются **ограничения** на базовые переменные. Функция поведения $f_B(c)$ определяет реально встречающиеся состояния с, но не определяет значение параметра, при котором они имеют место. Следовательно эта функция является **параметрически инвариантной**. Система, характеризующая параметрически инвариантное ограничение на множество переменных через функции поведения, называется **системой с поведением** и определяется тройкой $F_B = (I, M, f_B)$.

Степень детерминированности системы измеряется обобщенной **нечеткостью**, сопутствующей порождению данных и определяется через функции поведения.

Вероятностный подход.

Для вероятностных функций распределения мерой нечеткости является энтропия Шеннона

$$H(f(x)\,|\,x\in X) = -\sum_{x\in X} f(x)\log_2 f(x). \qquad (3)$$

Возможностный подход.

Для любого распределения возможностей и для любого действительного $l\in[0,1]$ функция C_L называется функцией уровня, а множество $c(f,l)=\{i\in N_{|X|}\,|\,\varphi_i\ge l\}$ называется множеством l-го уровня от f. Обозначим через $L_f=\{l_1,...,l_q\}$ уровневое множество для f, где $l_1=0, q=|L_f|; l_i<l_j\,|\,i<j; l_f=\max_i \varphi_i, l_f=l_q\in L_f$.

Функция U -нечеткости имеет вид

$$U(f)=\frac{1}{l_f}\int_0^{l_f}\log_2\left|C_L(f,l)\right|dl$$ или

$$U(f)=\frac{1}{l_f}\sum_{k=1}^{q-1}(l_{k+1}-l_k)\log_2\left|C_L(f,l_{k+1})\right|. \qquad (4)$$

На примере таблицы 2 вычислим меру неопределенности по вероятностному подходу по формуле 3

$$H = -\begin{pmatrix} 3\cdot 0.2\cdot \ln 0.2 + 0.1\cdot \ln 0.1 + 0.16667\cdot \ln 0.16667 + \\ + 0.0667 \ln 0.0667 + 2\cdot 0.0333 \ln 0.0333 \end{pmatrix} / \ln 2 = 2.7437 .$$

Для расчета меры нечеткости по возможностному подходу определим, в соответствии с табл.2, уровневое множество $L_f = \{0;0.16667;0.33333;0.5;0.83333;1\}$ и функцию уровня $C_L(f,l) = \{8;8;6;5;4;3\}$. По формуле (4) рассчитываем меру нечеткости:

$$U = \begin{pmatrix} 0.16667 \ln 8 + 0.16667 \ln 6 + \\ 0.16667 \ln 5 + 0.3333 \ln 4 + 0.16667 \ln 3 \end{pmatrix} / \ln 2 = 2.2483 .$$

От системы данных с упорядоченным параметрическим множеством и наибольшей допустимой маской необходимо перейти к выбору систем с поведением, удовлетворяющим требованиям **детерминированности, согласованности и простоты**. Для подобного описания выборочные переменные нужно разбить на два подмножества:

1) переменные, состояния которых порождаются из ограничения, называются порождаемыми;
2) переменные, состояния которых используются как условия в процессе генерации, называются порождающими.

Для заданной системы с поведением для определения порождаемых и порождающих переменных необходимо определить для данной маски M две подмаски $M_g, M_{\bar{g}}$, где $M_g, M_{\bar{g}} \subset M$, $M_g \cup M_{\bar{g}} = M$, $M_g \cap M_{\bar{g}} = 0$. Подмаски $M_g, M_{\bar{g}}$ называются **порождаемыми и порождающими**, соответственно.

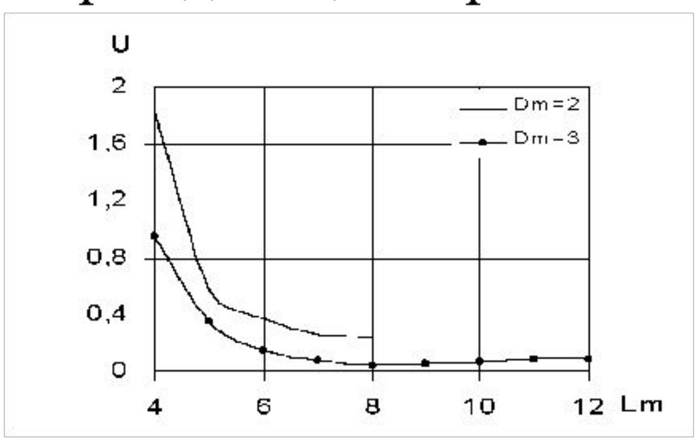
Рассмотрим маску глубиной 2. Наибольшая допустимая маска имеет вид

s_2	s_4	s_6
s_1	s_3	s_5

Из множества подмасок с количеством выборочных переменных 4 минимальную неопределенность $H(G\,|\,\overline{G}) =) = 2.375$ дает подмаска

s_2		
s_1	s_3	s_5

Порождающая переменная s_2 , порождаемые $\{s_1, s_3, s_5\}$.



Зависимость нечеткости от размера маски.

Контрольное задание

2	0	0
0	2	1
0	0	0
0	1	1
0	1	0
0	2	2
0	0	0
2	1	1
2	0	0
2	2	2
2	0	0
0	0	0
1	0	0
2	0	0
1	0	0
2	0	0
2	2	2
1	2	2
2	0	0
0	0	0
1	0	0
1	0	0
2	1	1
2	2	2
1	0	0
1	1	1
2	1	2
2	0	0
1	0	0
0	2	1

Вероятностный подход

$N(c)$	f_B
7	0.2333
2	0.0667
4	0.1333
1	0.0333
1	0.0333
1	0.0333
2	0.0667
3	0.1
6	0.2
1	0.0333
1	0.0333
1	0.0333

$$f_B \ln f_B$$

$$H = \frac{1}{\ln 2} \sum f_B \ln f_B$$

0.340	0.181	0.269	0.113	0.113	0.113	0.181	0.230	0.322	0.113	0.113	0.113	**3.177**

Возможностный подход

$N(c)$	f_B
7	1
2	0.2857
4	0.5714
1	0.1429
1	0.1429
1	0.1429
2	0.2857
3	0.4286
6	0.8571
1	0.1429
1	0.1429
1	0.1429

$N(c)$	f_B
1	0.1429
1	0.1429
1	0.1429
1	0.1429
1	0.1429
1	0.1429
2	0.2857
2	0.2857
3	0.4286
4	0.5714
6	0.8571
7	1

Функция уровня C_L	12	6	4	3	2	1	
Уровневое множество L_f	0.143	0.286	0.429	0.571	0.857	1.000	
$l_{k+1} - l_k$	0.1429	0.1429	0.1429	0.1429	0.2857	0.1429	
$U(f) = \frac{1}{l_f} \sum_{k=1}^{q-1} (l_{k+1} - l_k) \log_2 \lvert C_L(f, l_{k+1}) \rvert$	0.355	0.256	0.198	0.1569	0.198	0	1.679