# МЕТОД ПРЕДСКАЗАНИЯ В ЗЫКЕ ПЕРВОГО ПОРЯДКА

**Демин[1] А.В., Витяев[2] Е.Е.**

[1]Институт систем информатики имени А. П. Ершова СО РАН г. Новосибирск
[2]Институт математики СО РАН г. Новосибирск, e-mail: vityaev@math.nsc.ru

**Аннотация**

В работе продолжается рассмотрение метода и программной системы «Discovery» обнаружений знаний в данных, реализующие разработанный ранее реляционный подход к обнаружению знаний. Рассматривается метод предсказания, использующий обнаруженные системой «Discovery» закономерности в языке первого порядка с вероятностными оценками. Предлагаемый метод предсказания нетривиален и аналогов не имеет.

## §1. Определение вида гипотез

Данная работа является продолжением работ [1, 2], в которых описывается метод обнаружения закономерностей и его применения. Предлагаемый метод предсказания, опирается на это описание метода и соответствующую ему терминологию, а также использует идеи метода предсказания, изложенного в [3].

Напомним основные понятия многосортной логики первого порядка [5], которые нам понадобятся впоследствии. *Сигнатурой* называется упорядоченная шестерка $\Sigma = \langle \mathbb{S}, \mathbb{P}, \mathbb{F}, \eta, \mu, \kappa \rangle$, где $\mathbb{S}$ – множество сортов, $\mathbb{P}$ – множество предикатных символов, $\mathbb{F}$ – множество функциональных символов, $\eta : \mathbb{P} \cup \mathbb{F} \to \omega$ – отображение местности (арности) символов, $\mu : \bigcup_{a \in \mathbb{P} \cup \mathbb{F}} \{a\} \times \eta(a) \to \mathbb{S}$ – отображение, сопоставляющее каждому аргументу символа его сорт, $\kappa : \mathbb{F} \to \mathbb{S}$ – отображение, сопоставляющее каждому функциональному символу сорт его значения.

Упорядоченная пара $\mathcal{A} = \langle A, \pi \rangle$ называется *многосортной алгебраической системой* сигнатуры $\Sigma = \langle \mathbb{S}, \mathbb{P}, \mathbb{F}, \eta, \mu, \kappa \rangle$, где

- $A = \{A_s \mid s \in \mathbb{S}\}$ – непустое множество, называемое носителем или основным множеством алгебраической системы $\mathcal{A}$;
- $\pi$ – отображение множества $\mathbb{P} \cup \mathbb{F}$ в множество отношений и операций на множестве $A$, называемое интерпретацией сигнатуры $\Sigma$ на $A$;

- если $P \in \mathbb{P}$, то $\pi(P) \subseteq A_{\mu(P,1)} \times ... \times A_{\mu(P,n)}$, где $n = \eta(P)$ – местность символа $P$, и $\pi(P)$ называется многоместным отношением на $A$;
- если $f \in \mathbb{F}$, то $\pi(f)$ – отображение $\pi(f): A_{\mu(f,1)} \times ... \times A_{\mu(f,n)} \to A_{k(f)}$, где $n = \eta(f)$ – местность символа $f$, и $\pi(f)$ – называется многоместной операцией на $A$.

Для определения понятия *терм* будем считать, что с каждым сортом $s$ связано счетное множество символов $V_s$ – переменных сорта $s$. Каждая переменная имеет только один сорт.

*Термом* сорта $s$ называется любая переменная или константа сорта $s$, а также любое конечное выражение вида $f(t_1,...,t_n)$, где $f \in \mathbb{F}$, $\eta(f) = n$, $\kappa(f) = s$, $t_i$ – терм сорта $\mu(f,i)$ для каждого $i = 1,...,n$.

*Атомарной формулой* называется выражение вида $P(t_1,...,t_n)$, где $P \in \mathbb{P}$, $\eta(P) = n$, $t_i$ – терм сорта $\mu(P,i)$ для каждого $i = 1,...,n$.

Значения многосортных термов и формул первого порядка практически не отличается от односортного случая, с единственным ограничением, что переменные сорта $s$ должны интерпретироваться элементами в множестве $A_s$.

Будем предполагать, что исходные данные представлены в виде таблицы значений $D$, строки которой соответствуют объектам, а колонки – признакам объектов. Т.е. $D = \{D(1),...,D(N)\}$, где $D(i)$ – *i*-я строка таблицы (объект с номером *i*), $D(i) = \{D(i,1),...,D(i,m)\}$, $D(i,j)$ – значение таблицы на пересечении *j*-ой колонки и *i*-ой строки (значение *j*-го признака объекта $D(i)$), $D(i,j) \in \mathrm{Re}$, где Re – множество действительных чисел.

Зафиксируем фрагмент многосортного языка логики первого порядка сигнатуры $\Sigma^* = \langle \mathbb{S}, \mathbb{P}, \mathbb{F}, \eta, \mu, \kappa \rangle$, где

- $\mathbb{S} = \{s_{obj}, s_{\mathrm{Re}}\}$, $s_{obj}$ – сорт объектов таблицы $D$, $s_{\mathrm{Re}}$ – сорт действительных чисел;
- $\mathbb{P}$ – множество предикатных символов, таких, что если $P \in \mathbb{P}$, то $\mu(P,i) = s_{\mathrm{Re}}$, где $i = 1,...,\eta(P)$, т.е. все аргументы символов из $\mathbb{P}$ имеют сорт $s_{\mathrm{Re}}$;
- $\mathbb{F} = \mathbb{F}_1 \cup \mathbb{F}_2 \cup \mathbb{F}_3$ – множество функциональных символов, где

  $\mathbb{F}_1$ – множество функциональных символов $f \in \mathbb{F}_1$, $\mu(f,i) \in \{s_{obj}, s_{\mathrm{Re}}\}$, $i = 1,...,\eta(f)$, $\kappa(f) = s_{obj}$, т.е. аргументы символов из $\mathbb{F}_1$ имеют сорт $s_{obj}$ или $s_{\mathrm{Re}}$, а их значения – сорт $s_{obj}$;

$\mathbb{F}_2$ – множество функциональных символов $f \in \mathbb{F}_2$, $\eta(f) = 1$, $\mu(f,1) = s_{obj}$ и $\kappa(f) = s_{\mathrm{Re}}$, т.е. все функциональные символы из $\mathbb{F}_2$ одноместные, их аргументы сорта $s_{obj}$, а значения – сорта $s_{\mathrm{Re}}$.

$\mathbb{F}_3$ – множество функциональных символов $f \in \mathbb{F}_3$, $\mu(f,i) = s_{\mathrm{Re}}$, где $i = 1,...,\eta(f)$, и $\kappa(f) = s_{\mathrm{Re}}$, т.е. все аргументы символов из $\mathbb{F}_3$ и их значения имеют сорт $s_{\mathrm{Re}}$;

Введем многосортную алгебраическую систему $\mathcal{A}^* = \langle \{D, \mathrm{Re}\}, \pi \rangle$ сигнатуры $\Sigma^*$, где $D$ – таблица данных, являющаяся носителем сорта $s_{obj}$, соответственно строки таблицы являются объектами сорта $s_{obj}$, Re – множество действительных чисел, являющееся носителем сорта $s_{\mathrm{Re}}$.

Поясним назначение множеств функциональных символов $\mathbb{F}_1$, $\mathbb{F}_2$ и $\mathbb{F}_3$.

Множество $\mathbb{F}_1$ позволяет описывать функции, организующие доступ к различным строкам таблицы. Это могут быть, к примеру, функции, которые возвращают строки, смещенные на определенное число позиций в таблице относительно заданных строк, или функции, возвращающие строки по их номеру, или функции, организующие поиск строки в таблице.

Множество $\mathbb{F}_2$ позволяет описывать функции, организующие доступ к различным признакам объектов. Это могут быть, к примеру, функции, возвращающие значение определенного признака заданной строки, или функции, возвращающие номер строки.

Множество $\mathbb{F}_3$ позволяет описывать функции, осуществляющие различные преобразования над признаками объектов или значениями констант.

Для дальнейшего описания будем считать, что у нас есть счетное множество символов $V_{obj}$ – переменных сорта $s_{obj}$ и $V_{\mathrm{Re}}$ – переменных сорта $s_{\mathrm{Re}}$.

Введем понятия *шаблона термов* и *шаблона предикатов*.

*Шаблон термов* – это пара $Tf = \langle f, \{\Psi_1,...,\Psi_n\} \rangle$, где $f \in \mathbb{F}_3$, $n = \eta(f)$, $\Psi_i = \{t_1,...,t_{m_i}\}$, $i = 1,...,n$, $t_k$, $k = 1,...,m_i$ – терм сорта $s_{\mathrm{Re}}$, такой, что любая переменная, входящая в терм $t_k$ имеет сорт $s_{obj}$. Шаблон терма $Tf = \langle f, \{\Psi_1,...,\Psi_n\} \rangle$ определяет множество термов, состоящее из всех термов вида $f(t_1,...,t_n)$, где $t_i \in \Psi_i$, $i = 1,...,n$. Будем обозначать через $[Tf]$ – множество термов, определяемое шаблоном $Tf$. Мы можем оп-

ределить множество термов $\Psi_i$ при помощи других шаблонов термов: $\Psi_i = [\boldsymbol{Tf}_1] \cup ... \cup [\boldsymbol{Tf}_k]$, где $\boldsymbol{Tf}_j$, $\boldsymbol{j} = \boldsymbol{1},...,\boldsymbol{k}$ – некоторые шаблоны термов.

Приведем пример шаблона термов.

Пусть *k*-я колонка таблицы данных содержит значения некоторого временного ряда (к примеру, цена закрытия акции).

Пусть $\boldsymbol{r}$ – двухместный функциональный символ, $\boldsymbol{r} \in \mathbb{F}_1$, $\boldsymbol{\mu}(\boldsymbol{r}, \boldsymbol{1}) = \boldsymbol{s}_{obj}$ – первый аргумент $\boldsymbol{r}$ сорта $\boldsymbol{s}_{obj}$, $\boldsymbol{\mu}(\boldsymbol{r}, \boldsymbol{2}) = \boldsymbol{s}_{\mathrm{Re}}$ – второй аргумент $\boldsymbol{r}$ сорта $\boldsymbol{s}_{\mathrm{Re}}$. Пусть интерпретацией символа $\boldsymbol{r}$ на $\mathcal{A}^*$ является отображение $\boldsymbol{\pi}(\boldsymbol{r})(\boldsymbol{D}(\boldsymbol{i}), \boldsymbol{k}) = \boldsymbol{D}(\boldsymbol{i} + \boldsymbol{k})$, которое для заданной *i*-ой строки таблицы возвращает строку с номером $(\boldsymbol{i} + \boldsymbol{k})$ (предполагаем, что $\boldsymbol{r}$ определена только для целочисленных $\boldsymbol{k}$).

Пусть $\boldsymbol{h}_k$ – одноместный функциональный символ, $\boldsymbol{h}_k \in \mathbb{F}_2$, интерпретацией которого на $\mathcal{A}^*$ является отображение $\boldsymbol{\pi}(\boldsymbol{h}_k)(\boldsymbol{D}(\boldsymbol{i})) = \boldsymbol{D}(\boldsymbol{i}, \boldsymbol{k})$, которое для заданного объекта (*i*-ой строки таблицы), возвращает значения его *k*-го признака (значение таблицы на пересечении *k*-ой колонки и *i*-ой строки).

Рассмотрим шаблон термов

$\langle \boldsymbol{t}_1 - \boldsymbol{t}_2, \{\Psi_1, \Psi_2\} \rangle$, где $\Psi_1 = \{\boldsymbol{h}_k(\boldsymbol{r}(\boldsymbol{i}, \boldsymbol{0}))\}$, $\Psi_2 = \{\boldsymbol{h}_k(\boldsymbol{r}(\boldsymbol{i}, -\boldsymbol{1})), \boldsymbol{h}_k(\boldsymbol{r}(\boldsymbol{i}, -\boldsymbol{2})), \boldsymbol{h}_k(\boldsymbol{r}(\boldsymbol{i}, -\boldsymbol{3}))\}$.

Легко видеть, что данный шаблон определяет три терма, интерпретации которых на $\mathcal{A}^*$ представляют собой функции, задающие временные лаги от 1 до 3:

1) $\boldsymbol{D}(\boldsymbol{i}, \boldsymbol{k}) - \boldsymbol{D}(\boldsymbol{i} - \boldsymbol{1}, \boldsymbol{k})$,

2) $\boldsymbol{D}(\boldsymbol{i}, \boldsymbol{k}) - \boldsymbol{D}(\boldsymbol{i} - \boldsymbol{2}, \boldsymbol{k})$,

3) $\boldsymbol{D}(\boldsymbol{i}, \boldsymbol{k}) - \boldsymbol{D}(\boldsymbol{i} - \boldsymbol{3}, \boldsymbol{k})$.

*Шаблон предикатов* – это пара $\boldsymbol{Tp} = \langle \boldsymbol{P}, \{\Theta_1, ..., \Theta_n\} \rangle$, где $\boldsymbol{P} \in \mathbb{P}$, $\boldsymbol{n} = \boldsymbol{\nu}(\boldsymbol{P})$, $\Theta_i = \{\boldsymbol{Tf}_1, ..., \boldsymbol{Tf}_{m_i}\}$, $\boldsymbol{i} = \boldsymbol{1},...,\boldsymbol{n}$, $\boldsymbol{Tf}_k$, $\boldsymbol{k} = \boldsymbol{1},...,\boldsymbol{m}_i$ – шаблоны термов. Шаблон предиката $\boldsymbol{Tp} = \langle \boldsymbol{P}, \{\Theta_1, ..., \Theta_n\} \rangle$ определяет множество атомарных формул, состоящее из всех атомарных формул вида $\boldsymbol{P}(\boldsymbol{t}_1, ..., \boldsymbol{t}_n)$, где $\boldsymbol{t}_i \in [\boldsymbol{Tf}_1] \cup ... \cup [\boldsymbol{Tf}_{m_i}]$, $\boldsymbol{i} = \boldsymbol{1},...,\boldsymbol{n}$, $\boldsymbol{Tf}_k \in \Theta_i$, $\boldsymbol{k} = \boldsymbol{1},...,\boldsymbol{m}_i$. Будем обозначать через $[\boldsymbol{Tp}]$ – множество атомарных формул, определяемых шаблоном $\boldsymbol{Tp}$.

Приведем пример шаблона предикатов.

Предположим, что *k*-я колонка таблицы данных представляет временной ряд со значениями дневной цены закрытия какой-нибудь ценной бумаги. Предположим, что мы

хотим определить множество предикатов, сравнивающих друг с другом цены закрытия последних пяти дней. Это можно сделать, определив следующий шаблон предикатов:

$\boldsymbol{Tp} = \langle \boldsymbol{t_1 < t_2}, \{\Theta_1, \Theta_2\} \rangle$, где $\Theta_1 = \boldsymbol{Tf}$ , $\Theta_2 = \boldsymbol{Tf}$ ,

$\boldsymbol{Tf} = \langle \boldsymbol{x}, \{\Psi\} \rangle$, $\Psi = \{\boldsymbol{h_k(r(i,0)), h_k(r(i,-1)), h_k(r(i,-2)), h_k(r(i,-3)), h_k(r(i,-4))}\}$ ,

функциональные символы $\boldsymbol{r}$ и $\boldsymbol{h_k}$ , и их интерпретации на $\mathcal{A}^*$ определены в предыдущем примере.

В данном примере шаблон термов $\boldsymbol{Tf}$ задает пять термов, интерпретациями которых на $\mathcal{A}^*$ являются следующие функции: $\boldsymbol{x_1(i) = D(i,k)}$, $\boldsymbol{x_2(i) = D(i-1,k)}$,…, $\boldsymbol{x_5(i) = D(i-4,k)}$. Таким образом, шаблон предикатов $\boldsymbol{Tp}$ задает множество атомарных формул, интерпретации которых на $\mathcal{A}^*$ представляют собой предикаты вида $\boldsymbol{D(i-b_1,k) < D(i-b_2,k)}$, где $\boldsymbol{b_1, b_2 \in \{0,-1,...,-4\}}$ .

Теперь, используя понятие шаблона предикатов, мы можем определить понятие *класса гипотез*.

*Класс гипотез* – это пара $\boldsymbol{Th} = \langle \{\boldsymbol{Tp_1,...,Tp_m}\}, \boldsymbol{P_0^{\varepsilon}} \rangle$, где $\boldsymbol{Tp_i}$ – шаблоны предикатов, $\boldsymbol{P_0^{\varepsilon}}$ – целевая литера, $\boldsymbol{P_0}$ – атомарная формула, $\boldsymbol{\varepsilon \in \{0,1\}}$ – обозначает наличие отрицания формулы.

Класс гипотез $\boldsymbol{Th} = \langle \{\boldsymbol{Tp_1,...,Tp_m}\}, \boldsymbol{P_0^{\varepsilon}} \rangle$ определяет множество формул вида

$$\forall \boldsymbol{i_1,...,i_k}(\boldsymbol{P_1^{\varepsilon}} \,\&\, ... \,\&\, \boldsymbol{P_n^{\varepsilon}} \rightarrow \boldsymbol{P_0^{\varepsilon}}),\ \boldsymbol{k \geq 0} \qquad (1)$$

где $\boldsymbol{i_1,...,i_k}$ – переменные сорта $\boldsymbol{s_{obj}}$, $\boldsymbol{P_i \in [Tp_1] \cup ... \cup [Tp_m]}$, $\boldsymbol{i = 1,...,n}$. Формулы вида (1) будем называть *правилами*.

В дальнейшем помимо указанной записи мы также будем использовать более удобную форму записи формул (1) через индивидные константы. Вместо кванторов всеобщности и связанных ими переменных введем константы $\boldsymbol{z_1,...,z_k}$. Тогда формулы вида (1) преобразуются в формулы вида

$$\boldsymbol{P_1^{\varepsilon}} \,\&\, ... \,\&\, \boldsymbol{P_n^{\varepsilon}} \rightarrow \boldsymbol{P_0^{\varepsilon}}. \qquad (2)$$

Смысл формулы (2) состоит в том, что при любой фиксированной замене индивидных констант на объекты из таблицы $\boldsymbol{D}$, из истинности посылки должна следовать истинность заключения.

Данное выше определение понятия класса гипотез позволяет разработать интерактивных способ задания классов гипотез, проверяемых на данных. Для этого достаточно разработать интерактивную систему конструирования термов, атомарных формул, шабло-

нов термов и шаблонов предикатов из заданного набора функциональных и предикатных символов. Подобная интерактивная система задания классов гипотез была реализована в программной системе «Discovery [1, 2, 4].

Предложенный способ задания классов гипотез также указывает способ вычисления различных гипотез для заданного класса. Для этого достаточно реализовать вычислительные процедуры для базового набора функциональных и предикатных символов. Вычисление различных формул вида (1) может быть сведено к вычислению соответствующих базовых символов, из которых построены формулы. Данный способ вычисления гипотез используется в реализованной версии системы «Discovery».

## §2. Общая формулировка метода предсказания

Пусть $\boldsymbol{Reg}(\boldsymbol{Th})$ – множество закономерностей, полученное методом обнаружения закономерностей [1, 2], по заданному классу гипотез $\boldsymbol{Th} = \left\langle \{\boldsymbol{Tp}_1,...,\boldsymbol{Tp}_m\}, \boldsymbol{P}_0^{\varepsilon} \right\rangle$ на обучающем множестве, $\mathbb{D} \subset \mathbb{A}$ случайно выбранном из генеральной совокупности объектов $\mathbb{A}$.

В данном параграфе приводится общая формулировка метода предсказания, использующего множество обнаруженных закономерностей $\boldsymbol{Reg}(\boldsymbol{Th})$.

Пусть $\boldsymbol{P}(\boldsymbol{Th}) = \{\boldsymbol{P}_1,...,\boldsymbol{P}_n\}$ – множество всех атомарных формул, которые мы можем получить с помощью шаблонов предикатов [2] $\{\boldsymbol{Tp}_1,...,\boldsymbol{Tp}_m\}$, входящих в класс гипотез $\boldsymbol{Th}$ [2]. Пусть из генеральной совокупности объектов $\mathbb{A}$ случайно выбран некоторый новый объект $\boldsymbol{b}$. В задачах предсказания считается, что истинностные значения некоторой части атомарных формул $\boldsymbol{P}^{\boldsymbol{И}} \subset \boldsymbol{P}(\boldsymbol{Th})$ на объектах $\mathbb{D}$ и $\boldsymbol{b}$ нам известны. Требуется, используя знания закономерностей из $\boldsymbol{Reg}(\boldsymbol{Th})$, по известным значениям истинности атомарных формул $\boldsymbol{P}^{\boldsymbol{И}}$ на объектах $\mathbb{D} \cup \boldsymbol{b}$ предсказать неизвестные значения истинности остальных атомарных формул $\boldsymbol{P}^{\boldsymbol{П}} = \boldsymbol{P}(\boldsymbol{Th}) \setminus \boldsymbol{P}^{\boldsymbol{И}}$ на этих же объектах $\mathbb{D} \cup \boldsymbol{b}$. Таким образом, задача предсказания состоит в том, чтобы по модели $\boldsymbol{pr}_0 = \left\langle \mathbb{D}, \boldsymbol{P}(\boldsymbol{Th}) \right\rangle$, на которой проводилось обучение, и модели $\boldsymbol{pr}^{\boldsymbol{И}} = \left\langle \mathbb{D} \cup \boldsymbol{b}, \boldsymbol{P}^{\boldsymbol{И}} \right\rangle$ восстановить модель $\boldsymbol{pr} = \left\langle \mathbb{D} \cup \boldsymbol{b}, \boldsymbol{P}(\boldsymbol{Th}) \right\rangle$, используя закономерности из $\boldsymbol{Reg}(\boldsymbol{Th})$.

Обозначим через $\boldsymbol{PS}$ множество всех возможных моделей $\boldsymbol{pr} = \left\langle \mathbb{D} \cup \boldsymbol{b}, \boldsymbol{P}(\boldsymbol{Th}) \right\rangle$, являющихся восстановлениями моделей $\boldsymbol{pr}_0$, $\boldsymbol{pr}^{\boldsymbol{И}}$. Так как множество $\boldsymbol{Reg}(\boldsymbol{Th})$ содержит статистические закономерности, то разные восстановления могут иметь разную вероятность.

Таким образом, метод предсказания должен состоять в том, чтобы по закономерностям $\boldsymbol{Reg(Th)}$ и моделям $\boldsymbol{pr_0}$, $\boldsymbol{pr^И}$ вычислить для каждой модели $\boldsymbol{pr \in PS}$ некоторую оценку её вероятности $\boldsymbol{\nu(pr)}$. Для некоторых моделей $\boldsymbol{pr \in PS}$ оценка вероятности $\boldsymbol{\nu(pr)}$ может быть не определена, так как может, например, оказаться, что для неё нет применимых к ней закономерностей.

**Определение 1.** *Методом предсказания* будем называть алгоритм $\boldsymbol{AP : \langle Reg(Th), pr_0, pr^И \rangle \to \nu}$, преобразующий тройки $\boldsymbol{\langle Reg(Th), pr_0, pr^И \rangle}$ в частично определенное отображение $\boldsymbol{\nu : PS \to [0, 1]}$.

## §3. Метод предсказания

Уточним в чем состоит смысл распространения моделей $\boldsymbol{pr_0}$, $\boldsymbol{pr^И}$ до моделей $\boldsymbol{pr \in PS}$. Если известны все вероятности, то для любой модели $\boldsymbol{pr \in PS}$ можно подсчитать вероятность $\wp(\boldsymbol{pr})$ того, что при случайном выборе объекта $\boldsymbol{b}$ из $\mathbb{A}$ мы в результате эксперимента над $\mathbb{D} \cup \boldsymbol{b}$ получим модель изоморфную $\boldsymbol{pr}$. Поэтому для получения наиболее точного предсказания алгоритм $\boldsymbol{AP}$ должен стремиться получить оценки вероятности $\boldsymbol{\nu(pr)}$ наиболее близкие к вероятности $\wp(\boldsymbol{pr})$.

Для восстановления модели $\boldsymbol{pr}$ надо определить значения истинности всех атомарных формул из $\boldsymbol{P^П}$, на всех наборах объектов, включающих хотя бы одно вхождение объекта $\boldsymbol{b}$. Для этой цели могут быть использованы те закономерности из $\boldsymbol{Reg(Th)}$ в заключение которых стоит атомарная формула из $\boldsymbol{P^П}$. Разобьем это множество закономерностей $\boldsymbol{Reg(Th)}$ на три группы:

$\boldsymbol{Reg_1}$ – множество закономерностей, включающее закономерности, содержащие только одноместные атомарные формулы, содержащие одну индивидную постоянную.

$\boldsymbol{Reg_2}$ – множество закономерностей, заключение которых содержит только одну индивидную постоянную, а посылка содержит, по крайней мере две различные индивидные постоянные.

$\boldsymbol{Reg_3}$ – множество закономерностей, у которых в заключении есть хотя бы две различные индивидные постоянные.

Для произвольного правила $\boldsymbol{R \in Reg(Th)}$, $\boldsymbol{R = P_1^{\varepsilon_1} \& ... \& P_n^{\varepsilon_n} \to P_0^{\varepsilon_0}}$ будем обозначать через $\boldsymbol{D_П = P_1^{\varepsilon_1} \& ... \& P_n^{\varepsilon_n}}$ конъюнкцию литер посылки, а через $\boldsymbol{D_C}$ – заключение пра-

вила $P_0^{\varepsilon_0}$. Будем также обозначать через $z(D)$ множество индивидных констант, входящих в формулу $D$.

Для осуществления предсказания необходимо в первую очередь определить для каждой закономерности $R \in Reg$ множество моделей $PS(R) \subset PS$, которое будет являться прогнозом для данной закономерности.

Если $R = (D_П \to D_C) \in Reg_1(Th)$, то проверим истинность формулы $D_П$ при подстановке в нее объекта $b$. Если $D_П$ истинна, то данная закономерность может быть использована для предсказания. Прогнозом закономерности $R$ будет являться множество $PS(R) = PS(D_C)$ тех моделей, на которых формула $D_C$ истинна.

Закономерности множеств $Reg_2$, $Reg_3$ принципиально отличаются от закономерностей $Reg_1$ тем, что в них есть несколько индивидных постоянных. Поэтому, подставляя объект $b$ вместо одной индивидной постоянной, мы должны подставить некоторые объекты и вместо других индивидных постоянных. Закономерность в этом случае говорит об определенной связи объекта $b$ с другими объектами. Поэтому закономерности множеств $Reg_2$, $Reg_3$ могут быть различным способом использованы для предсказания.

Если закономерность $R = (D_П \to D_C)$ принадлежит $Reg_2(Th)$, то разобьем случайным образом обучающее множество $\mathbb{D}$ на $l$ наборов объектов по $k$ объектов в каждом, где $l = \frac{m}{k}$, $m = \overline{\overline{\mathbb{D}}}$, $k$ – количество индивидных постоянных во множестве $z(D_П) \setminus z(D_C)$. Будем последовательно подставлять эти наборы вместо индивидных постоянных и определять значения истинности формулы $D_П$. Если формула $D_П$ истинна хотя бы на одном наборе объектов, то данную закономерность можно использовать для предсказания. В противном случае по этой закономерности предсказание сделать нельзя. Прогнозом данной закономерности, как и в предыдущем случае, будет множество $PS(R) = PS(D_C)$.

Закономерности из $Reg_3$ принципиально отличаются от закономерностей из $Reg_1$ и $Reg_2$ тем, что в них предсказывается не истинность некоторого отношения, зависящего от одной индивидной постоянной, а предсказывается определенное отношение между одной индивидной постоянной и некоторыми другими индивидными постоянными.

Пусть $R = (D_П \to D_C) \in Reg_3$. Обозначим через $П = \{z_1, ..., z_p\}$, $П \subset z(D_C)$ множество индивидных постоянных из $D_C$, которые входят в $D_П$, но не входят в атомарные формулы из $P^П$. Предсказываемый объект $b$ можно подставлять вместо любой индивидной постоянной из $П$.

Подставив объект $b$ вместо индивидной постоянной $z \in \Pi$, мы получим закономерность $R_b^z = ((D_\Pi)_b^z \to (D_C)_b^z)$. Пусть $z_1,...,z_k$ – остальные индивидные постоянные, входящие в закономерность $R_b^z$. Разобьем случайным образом всё множество $\mathbb{D}$ на $l$ наборов объектов по $k$ объектов в каждом, где $l = \frac{m}{k}$, $m = \overline{\overline{\mathbb{D}}}$. Пусть $\{\langle a_1 \rangle,...,\langle a_l \rangle\}$ – полученное множество наборов $\langle a_i \rangle = \{a_1^i,...,a_k^i\}$, $a_j^i \in \mathbb{D}$, $i = 1,...,l$, $j = 1,...,k$. Подставим эти наборы вместо соответствующих индивидных постоянных и определим значения истинности формул $(D_\Pi)_b^z$, $(D_C)_b^z$. Если формула $(D_\Pi)_b^z$ истинна хотя бы на одном наборе объектов, то закономерность $R_b^z$ можно использовать для предсказания. Множество моделей, являющихся прогнозом закономерности $R_b^z$ определим следующим образом: $PS(R_b^z) = \bigcup_{i=1}^{l} PS((D_C)_b^z(\langle a_i \rangle))$, где $PS((D_C)_b^z(\langle a_i \rangle))$ – множество моделей из $PS$, в которых формула $(D_C)_b^z$ истинна на наборе объектов $\langle a \rangle$.

Подставляя последовательно объект $b$ вместо каждой индивидной константы из $\Pi$ мы получим закономерности $R_b^{z_1},\ldots,R_b^{z_p}$. Если хотя бы одна из этих закономерностей будет применима для предсказания, то закономерность $R$ можно использовать для предсказания. В качестве прогноза закономерности $R$ определим множество моделей $PS(R) = \bigcup_{i=1}^{p} PS(R_b^z)$.

Чтобы оценить точность прогноза $\nu(pr)$ некоторой модели $pr \in PS$, метод предсказания должен основываться на прогнозах всех закономерностей из $Reg(Th)$, способных предсказать модель $pr$.

Обозначим через $Reg(pr)$ множество закономерностей из $Reg(Th)$ таких, что для любого $R \in Reg(pr)$, $pr \in PS(R)$.

Для получения итоговой оценки точности прогноза $\nu(pr)$ для модели $pr$ необходимо задать частично определенную функцию $\lambda : \{\langle pr, Reg(pr) \rangle\} \to [0,1]$, которая для каждой модели $pr \in PS$ на основании множества закономерностей $Reg(pr)$, предсказывающих данную модель, вычисляет оценку точности прогноза $\nu(pr)$. Тогда $\nu(pr) = \lambda(\langle pr, Reg(pr) \rangle)$.

В зависимости от специфики решаемой задачи функция $\lambda$ может быть определена по-разному. Приведем несколько примеров задания функции $\lambda$. Обозначим через

$Reg^+(pr) \subset Reg(pr)$ множество закономерностей применимых для осуществления прогноза.

1. По максимальной вероятности:

$$\lambda(\langle pr, Reg(pr)\rangle) = \max_{R \in Reg^+(pr)} \{\wp(R)\},$$

где $\wp(R)$ – условная вероятность правила $R$.

2. По относительному количеству сработавших правил:

$$\lambda(\langle pr, Reg(pr)\rangle) = \frac{n(Reg^+(pr))}{n(Reg(pr))}, \text{ если } n(Reg(pr)) \neq 0,$$

$$\lambda(\langle pr, Reg(pr)\rangle) = 0, \text{ если } n(Reg(pr)) = 0,$$

где $n(Reg(pr))$ и $n(Reg^+(pr))$ – количество правил во множествах $Reg(pr)$ и $Reg^+(pr)$.

3. По средней вероятности:

$$\lambda(\langle pr, Reg(pr)\rangle) = \frac{\sum_{R \in Reg^+(pr)} \wp(R)}{n(Reg^+(pr))}, \text{ если } n(Reg^+(pr)) \neq 0,$$

$$\lambda(\langle pr, Reg(pr)\rangle) = 0, \text{ если } n(Reg^+(pr)) = 0.$$

4. По средневзвешенной вероятности:

$$\lambda(\langle pr, Reg(pr)\rangle) = \frac{\sum_{R \in Reg^+(pr)} \wp(R)}{\sum_{R \in Reg(pr)} \wp(R)}, \text{ если } \sum_{R \in Reg(pr)} \wp(R) \neq 0,$$

$$\lambda(\langle pr, Reg(pr)\rangle) = 0, \text{ если } \sum_{R \in Reg(pr)} \wp(R) = 0.$$

## §4. Метод предсказания, основанный на оценке максимальной вероятности [2].

**4.1 Вероятностные оценки закономерностей**. Сначала для всех трех видов закономерностей $Reg_1$, $Reg_2$ и $Reg_3$ на обучающем материале подсчитаем некоторые вероятностные оценки, необходимые для получения предсказаний.

Для закономерностей из $Reg_1$ подсчитаем нижнюю доверительную границу для условной вероятности $\wp(D_C \mid D_П)$. При подстановке вместо единственной индивидной постоянной объектов из $\mathbb{D}$, формулы $D_П$, $D_C$ будут принимать определенные значения истинности. Подсчитаем на объектах $\mathbb{D}$ частоту $h(D_C \mid D_П)$ условного события $D_C \mid D_П$. Используем один из известных методов построения доверительных интервалов для условной вероятности. Для фиксированного доверительного уровня $\beta$ по частоте можно опреде-

лить нижнюю доверительную границу $\underline{h}^{\beta}(D_C \mid D_П)$ для условной вероятности, обладающую свойством

$$\wp(\wp(D_C \mid D_П) \geq \underline{h}^{\beta}(D_C \mid D_П)) \geq 1-\beta. \quad (3)$$

Для закономерностей из $Reg_2$, также как и для закономерностей из $Reg_1$, вычислим нижнюю доверительную границу для условной вероятности $p(D_C \mid D_П)$. Для этого, как и в предыдущем случае, достаточно вычислить частоту $h(D_C \mid D_П)$. Для двух и более индивидных постоянных процедура вычисления частоты отличается от предыдущего случая. Разобьем случайным образом множество объектов $\mathbb{D}$ на два множества, которые обозначим соответственно через $A$ и $C$, $\mathbb{D} = A \cup C$, $A \cap C = \varnothing$. Объекты, имитирующие объект $b$, будем брать из множества $C$, а остальные объекты, подставляемые в закономерность, будем брать из множества $A$.

Пусть $z(D_C) = z$ и $z(D_П) = \{z, z_1, ..., z_k\}$. Подсчитает частоту $h(D_C \mid D_П)$. Будем подставлять последовательно все объекты из $C$ вместо индивидной постоянной $z$. Для каждого подставленного объекта случайным образом выберем набор объектов $a_1, ..., a_k$ из множества $A$. Подставим его вместо индивидных постоянных $z_1, ..., z_k$. Формулы $D_C$ и $D_П$ примут определенные значения истинности. Подставив последовательно все объекты из С вместо индивидной постоянной $z$, можно подсчитать частоту $h(D_C \mid D_П)$. Интерпретация условного события $D_C \mid D_П$ определяется как вероятность того, что формула $D_C$ будет истинна при подстановке в нее объекта $b$ вместо индивидной постоянной $z$ и при подстановке случайно выбранных из множества $A$ объектов вместо остальных индивидных постоянных. При заданном доверительном уровне $\beta$ по частоте $h(D_C \mid D_П)$ можно вычислить нижнюю доверительную границу условной вероятности, удовлетворяющую неравенству (3).

Возьмем произвольную закономерность из $Reg_3$. Объекты, с которыми предсказываемый объект $b$ должен находиться в некотором отношении, будем, также как и в предыдущем случае, брать из множества $C$. Пусть $D_C = P^{\varepsilon_0}(z_1, ..., z_{m_0})$. Обозначим через $П \subset \{z_1, ..., z_{m_0}\}$ множество индивидных постоянных из $D_C$, которые в $D_П$ не входят в атомарные формулы из $P^{П}$. По определению множества $Reg_3$, множество $П$ не пусто. Предсказываемый объект $b$ можно подставлять вместо любой индивидной постоянной из $П$.

Для данной закономерности и индивидной постоянной $z \in \Pi$ подсчитаем следующую оценку. Подставим объект $b \in C$ вместо индивидной постоянной $z$. Пусть $z_1,...,z_k$ – остальные индивидные постоянные, входящие в закономерность. Вместо этих индивидных постоянных будем подставлять объекты из множества $A$. Для этого разобьем случайным образом все множество $A$ на $l$ наборов объектов по $k$ объектов в каждом, где $l = \frac{m_1}{k}$, $m_1$ – количество элементов во множестве $A$. Подставив любой из этих наборов вместо соответствующих индивидных постоянных, мы можем определить значения истинности формул $D_\Pi$ и $D_C$. Подставляя последовательно все наборы в закономерность для данного объекта $b$ можно подсчитать частоту $h_z^b(D_C \mid D_\Pi)$. При случайной подстановке наборов объектов множества $A$ вместо индивидных постоянных $z_1,...,z_k$, частота $h_z^b(D_C \mid D_\Pi)$ будет одномерной случайной величиной. Эта случайная величина имеет неизвестное нам распределение $H_z^b$. Будем подставлять вместо объекта $b$ объекты из множества $C$. Получим $m_2$ выборочных значений частоты $h_z^{c_1},...,h_z^{c_{m_2}}$, $C = \{c_1,...,c_{m_2}\}$. Найти доверительные границы для частоты по данным выборочным значениям при неизвестной функции распределения можно при помощи порядковых статистик. Так как частоты $h_z^{c_i}$ равны отношению целых чисел, то случайная величина $h_z^c$ дискретна.

Приведем необходимые результаты из порядковых статистик [6, с. 128-130]. Пусть $h_{(1)} < h_{(2)} < ... < h_{(m)}$ – порядковые статистики в выборке объема $m$ из генеральной совокупности с неизвестной непрерывной функцией распределения $H$. Вероятность того, что новое значение случайной величины будет находиться в пределах $h_{(r)} < h < h_{(s)}$, равна $H(h_{(s)}) - H(h_{(r)})$. Эта вероятность как функция случайных величин сама будет случайной величиной. Используя порядковые статистики $h_{(r)}$ и $h_{(s)}$, можно получить следующие толерантные интервалы для распределения $H$

$$1 - \beta = \wp(H(h_{(s)}) - H(h_{(r)}) \geq \gamma) = \sum_{i=0}^{s-r-1} \binom{m}{i} \gamma^i (1-\gamma)^{m-i}. \qquad (4)$$

Из данного равенства следует, что вероятность того, что с вероятностью большей либо равной $\gamma$ имеет место неравенство $h_{(r)} < h < h_{(s)}$, равна $1 - \beta$. Используя эту формулу, можно для значений вероятностей $\gamma$ и $\beta$, и объема $m$ выборки найти такие порядковые статистики $h_{(r)}$ и $h_{(s)}$, чтобы выполнялось неравенство $\wp(H(h_{(s)}) - H(h_{(r)}) \geq \gamma) \geq 1 - \beta$.

Используя результаты Тьюки [7] можно распространить это неравенство и на дискретное распределение $H$. Если обозначить через $H(h_{(s)}+0)$ и $H(h_{(r)}-0)$ пределы функции распределения $H$ соответственно справа и слева в точках $h_{(s)}$ и $h_{(r)}$, то, согласно [7], должно выполняться неравенство

$$\wp(H(h_{(s)}+0)-H(h_{(r)}-0)\geq\gamma)\geq 1-\beta. \qquad (5)$$

В неравенстве (5) номера $s$ и $r$ можно вычислить, используя формулу (4). Расположим выборочные значения частоты $h_z^{c_i}$, $i=1,...,m_2$ в порядке возрастания. Получим $m_2$ порядковых статистик $h_{(1)}<h_{(2)}<...<h_{(m_2)}$. Фиксируем некоторые вероятности $\gamma$ и $\beta$. Для этих значений вероятности и данного числа $m_2$ порядковых статистик можно по формуле (4) вычислить номера $s$ и $r$ порядковых статистик $h_{(s)}$ и $h_{(r)}$, для которых будет выполняться неравенство (5). Эти статистики определяют толерантный интервал $[h_{(r)},h_{(s)}]$, который обозначим через $[\underline{h}^z,\overline{h}^z]$.

Подсчитаем для каждой закономерности из $Reg_3$ и для каждой индивидной постоянной $z\in\Pi$ толерантный интервал $[\underline{h}^z,\overline{h}^z]$. На этом подсчет статистических оценок для закономерностей из $Reg_1$, $Reg_2$, $Reg_3$ закончен.

**4.2 Предсказание.** Рассмотрим сначала случай, когда множество закономерностей $Reg(Th)$ принадлежит либо $Reg_1(Th)$, либо $Reg_2(Th)$.

Пусть формула $D_\Pi$ истинна. Разобьем множество $PS$ на два множества $PS(D_C)$ и $PS(\overline{D}_C)$, включающие соответственно модели, на которых формула $D_C$ истинна, и на которых она ложна. В соответствии с закономерностью, при истинности $D_\Pi$ предсказываются те модели из $PS$, которые принадлежат множеству $PS(D_C)$. Это множество, рассматриваемое как событие при случайном выборе объекта $b$, имеет некоторую вероятность $\wp(PS(D_C))$, оценкой которой является величина $\underline{h}^\beta(D_C\mid D_\Pi)$. Поэтому определим отображение $\nu$ следующим образом: $\nu(pr)=0$, если $pr\in PS(\overline{D}_C)$, и $\nu(pr)=\underline{h}^\beta(D_C\mid D_\Pi)$, если $pr\in PS(D_C)$. Для такого предсказания в силу неравенства (3) будет выполнено соотношение

$$\wp(\wp(PS(D_C))\geq\underline{h}^\beta(D_C\mid D_\Pi))\geq 1-\beta. \qquad (6)$$

Рассмотрим случай, когда множество $Reg_1(Th)\cup Reg_2(Th)$ состоит из закономерностей $D_{\Pi_1}\to D_{C_1},\ldots,D_{\Pi_k}\to D_{C_k}$, предсказывающих одну и ту же литеру

$P_0^{\varepsilon_0}(z_1,...,z_{m_0}) = D_{C_1} = ... = D_{C_k}$. Проведем процедуру предсказания данной литеры по всем закономерностям, как описано выше. Для тех закономерностей, у которых посылка $D_{\Pi_i}$, $i = 1,2,...,k$ будет истинной, получим предсказания одного и того же множества $PS(D_C)$ с оценками вероятности $\underline{h}^{\beta}(D_{C_1} \mid D_{\Pi_1}),\ldots,\ \underline{h}^{\beta}(D_{C_k} \mid D_{\Pi_k})$. Способ получения результирующей оценки должен определяться дополнительными предположениями, которые можно сделать относительно связи этих закономерностей.

### 4.3 Дополнительные предположения о независимости закономерностей.

**Определение 2.** Две закономерности $D_{\Pi}^1 \to P_0^{\varepsilon}$ и $D_{\Pi}^2 \to P_0^{\varepsilon}$ будем называть *независимыми*, если выполнены следующие условия

1. $\wp(D_{\Pi}^1 \,\&\, D_{\Pi}^2) = \wp(D_{\Pi}^1)\,p(D_{\Pi}^2)$,
2. $\wp((D_{\Pi}^1 \,\&\, \bar{P}_0^{\varepsilon}) \,\&\, (D_{\Pi}^2 \,\&\, \bar{P}_0^{\varepsilon})) = \wp((D_{\Pi}^1 \,\&\, \bar{P}_0^{\varepsilon}))\wp((D_{\Pi}^2 \,\&\, \bar{P}_0^{\varepsilon}))$,

где $\bar{P}_0^{\varepsilon}$ – отрицание литеры $P_0^{\varepsilon}$.

**Лемма 1.** Если закономерности $D_{\Pi}^1 \to P_0^{\varepsilon}$ и $D_{\Pi}^2 \to P_0^{\varepsilon}$ независимы, то условная вероятность закономерности $D_{\Pi}^1 \,\&\, D_{\Pi}^2 \to P_0^{\varepsilon}$ равна

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) = 1 - (1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1))(1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^2)). \tag{7}$$

**Доказательство**. Разделим условие 2 независимости на $\wp(D_{\Pi}^1 \,\&\, D_{\Pi}^2)$. Получим

$$\frac{\wp((D_{\Pi}^1 \,\&\, \bar{P}_0^{\varepsilon}) \,\&\, (D_{\Pi}^2 \,\&\, \bar{P}_0^{\varepsilon}))}{\wp(D_{\Pi}^1 \,\&\, D_{\Pi}^2)} = \frac{\wp(D_{\Pi}^1 \,\&\, \bar{P}_0^{\varepsilon})\wp(D_{\Pi}^2 \,\&\, \bar{P}_0^{\varepsilon})}{\wp(D_{\Pi}^1 \,\&\, D_{\Pi}^2)}. \tag{8}$$

Левая часть равенства равна $\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2)$. Заменим вероятность $\wp(D_{\Pi}^1 \,\&\, D_{\Pi}^2)$ в правой части равенства на $\wp(D_{\Pi}^1)\wp(D_{\Pi}^2)$, согласно условию 1 независимости. Тогда правая часть равенства (8) равна $\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1)\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^2)$. Поэтому равенство (8) переходит в равенство

$$\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) = \wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1)\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^2).$$

Так как

$$\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) = 1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2),$$
$$\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^1) = 1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1),$$
$$\wp(\bar{P}_0^{\varepsilon} \mid D_{\Pi}^2) = 1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^2),$$

то отсюда получаем утверждение леммы ■.

**Лемма 2.** Если закономерности $D_{\Pi}^1 \to P_0^{\varepsilon}$ и $D_{\Pi}^2 \to P_0^{\varepsilon}$ независимы, и их нижние доверительные границы условных вероятностей $p(P_0^{\varepsilon} \mid D_{\Pi}^1)$ и $p(P_0^{\varepsilon} \mid D_{\Pi}^2)$ равны соответственно $\underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^1)$, $\underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^2)$, то нижняя доверительная граница для условной вероятности $\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2)$ равна $\underline{h}^{2\beta}(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) = 1 - (1 - \underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^1))(1 - \underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^2))$.

**Доказательство**. Из условия и леммы 1 следует, что имеют место следующие отношения:

$$\wp(\wp(P_0^\varepsilon \mid D_\Pi^1) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) \ge 1-\beta,$$

$$\wp(\wp(P_0^\varepsilon \mid D_\Pi^2) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2)) \ge 1-\beta,$$

$$\wp(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2) = 1-(1-\wp(P_0^\varepsilon \mid D_\Pi^1))(1-\wp(P_0^\varepsilon \mid D_\Pi^2)).$$

Из первых двух неравенств получаем следующие неравенства:

$$\wp(\wp(P_0^\varepsilon \mid D_\Pi^1) < \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) < \beta,$$

$$\wp(\wp(P_0^\varepsilon \mid D_\Pi^2) < \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2)) < \beta.$$

Применяя теорему о сложении вероятностей, получим

$$\wp((\wp(P_0^\varepsilon \mid D_\Pi^1) < \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) \vee (\wp(P_0^\varepsilon \mid D_\Pi^2) < \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))) < 2\beta,$$

$$\wp((\wp(P_0^\varepsilon \mid D_\Pi^1) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) \,\&\, (\wp(P_0^\varepsilon \mid D_\Pi^2) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))) \ge 1-2\beta. \qquad (9)$$

Нетрудно видеть, что из условия

$$(\wp(P_0^\varepsilon \mid D_\Pi^1) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) \,\&\, (\wp(P_0^\varepsilon \mid D_\Pi^2) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))$$

вытекает следующее неравенство

$$1-(1-\wp(P_0^\varepsilon \mid D_\Pi^1))(1-\wp(P_0^\varepsilon \mid D_\Pi^2)) \ge 1-(1-\underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1))(1-\underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))$$

Из условия и леммы 1 следует, что левая часть этого неравенства равна $\wp(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2)$. Обозначим через $\underline{h}^{2\beta}(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2)$ величину $1-(1-\underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1))(1-\underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))$ правую часть неравенства, тогда вероятность неравенства

$$\wp(\wp(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2) \ge \underline{h}^{2\beta}(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2))$$

равна вероятности, стоящей в левой части неравенства (9)

$$\wp((\wp(P_0^\varepsilon \mid D_\Pi^1) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^1)) \,\&\, (\wp(P_0^\varepsilon \mid D_\Pi^2) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_\Pi^2))).$$

Отсюда следует, что $\wp(\wp(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2) \ge \underline{h}^{2\beta}(P_0^\varepsilon \mid D_\Pi^1 \,\&\, D_\Pi^2)) \ge 1-2\beta$ ■.

**Определение 3.** Закономерности $D_\Pi^1 \to P_0^\varepsilon$, $D_\Pi^2 \to P_0^\varepsilon$,…, $D_\Pi^n \to P_0^\varepsilon$ будем называть *независимыми*, если закономерности $D_\Pi^1 \to P_0^\varepsilon$ и $D_\Pi^2 \to P_0^\varepsilon$ независимы, закономерности $D_\Pi^1 \,\&\, D_\Pi^2 \to P_0^\varepsilon$ и $D_\Pi^3 \to P_0^\varepsilon$ независимы и так далее, закономерности $D_\Pi^1 \,\&\, \ldots \,\&\, D_\Pi^{n-1} \to P_0^\varepsilon$ и $D_\Pi^n \to P_0^\varepsilon$ независимы.

**Лемма 3.** Для независимых закономерностей $D_\Pi^1 \to P_0^\varepsilon$,…, $D_\Pi^n \to P_0^\varepsilon$ выполняются следующие соотношения, обобщающие результаты предыдущих лемм

1. $\wp(P_0^\varepsilon \mid D_\Pi^1 \,\&\, \ldots \,\&\, D_\Pi^n) = 1-\prod_{i=1}^{n}(1-\wp(P_0^\varepsilon \mid D_\Pi^i))$,

2. $\wp(\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^n) \geq \underline{h}^{n\beta}(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^n)) \geq 1 - n\beta$,

$$\underline{h}^{n\beta}(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^n) = 1 - \prod_{i=1}^{n} (1 - \underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^i)).$$

**Доказательство**.

1) Доказательство первого равенства проведем по индукции. Случай $n = 2$ доказан в лемме 2.

Пусть закономерности $D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^k \to P_0^{\varepsilon}$ и $D_{\Pi}^{k+1} \to P_0^{\varepsilon}$ независимы. Тогда применяя к ним лемму 2 получим

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^{k+1}) = 1 - (1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^k))(1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^{k+1})).$$

По индуктивному предположению

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, \ldots \,\&\, D_{\Pi}^k) = 1 - \prod_{i=1}^{k} (1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^i)).$$

Подставляя это выражение в предыдущую формулу, получим равенство 1 леммы.

2) Доказательство пункта 2 леммы проводится аналогично предыдущему пункту, только надо иметь в виду, что результат леммы 2 не измениться, если нижние доверительные границы $\underline{h}^{\beta}(P_0^{\varepsilon} \mid D_{\Pi}^i)$, $i = 1, 2$ будут иметь различный доверительный уровень $\beta$ ■.

Выясним соотношение свойства независимости двух закономерностей $D_{\Pi}^1 \to P_0^{\varepsilon}$, $D_{\Pi}^2 \to P_0^{\varepsilon}$ и критерия для отбора закономерностей, применяемого в методе обнаружения закономерностей. Покажем, что свойство независимости означает независимость вклада этих закономерностей в предсказание $P_0^{\varepsilon}$.

**Лемма 4.** Если закономерности $D_{\Pi}^1 \to P_0^{\varepsilon}$ и $D_{\Pi}^2 \to P_0^{\varepsilon}$ независимы, то формулы $D_{\Pi}^1$ и $D_{\Pi}^2$ «существенны» в закономерности $D_{\Pi}^1 \,\&\, D_{\Pi}^2 \to P_0^{\varepsilon}$, т.е. для них выполняется приведённое ниже условие вероятностных закономерностей [4].

**Доказательство**. Из независимости следует, что

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) = 1 - (1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1))(1 - \wp(P_0^{\varepsilon} \mid D_{\Pi}^2)) =$$

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1) + \wp(P_0^{\varepsilon} \mid D_{\Pi}^2) - \wp(P_0^{\varepsilon} \mid D_{\Pi}^1)\wp(P_0^{\varepsilon} \mid D_{\Pi}^2).$$

Так как вероятность находится в интервале $[0, 1]$, то

$$\wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) > \wp(P_0^{\varepsilon} \mid D_{\Pi}^1),\ \wp(P_0^{\varepsilon} \mid D_{\Pi}^1 \,\&\, D_{\Pi}^2) > \wp(P_0^{\varepsilon} \mid D_{\Pi}^2),$$

что является условием «существенности» формул $D_{\Pi}^1$ и $D_{\Pi}^2$ в закономерности ■.

Таким образом, если закономерности $D_{\Pi}^1 \to P_0^{\varepsilon}$ и $D_{\Pi}^2 \to P_0^{\varepsilon}$ независимы, то их можно улучшить добавлением в посылку одной закономерности посылки из другой зако-

номерности. Если существенность формулы $D_П^1$ для закономерности $D_П^2 \to P_0^\varepsilon$ и формулы $D_П^2$ для закономерности $D_П^1 \to P_0^\varepsilon$ может быть установлена на обучающем материале, как описано в методе обнаружения закономерностей, то должна будет обнаружена и закономерность $D_П^1 \,\&\, D_П^2 \to P_0^\varepsilon$. Поэтому предположение о независимости закономерностей означает, что они улучшают друг друга, но по обучающему материалу это нельзя обнаружить вследствие его ограниченности.

Предположение о независимости позволяет нам по закономерностям $D_П^1 \to P_0^\varepsilon, \ldots, D_П^n \to P_0^\varepsilon$ определить общую закономерность $D_П^1 \,\&\, \ldots \,\&\, D_П^n \to P_0^\varepsilon$ и вычислить для неё оценку вероятности. Поэтому отображение $\nu$ будет иметь вид, приведенный в лемме 3. Для такого предсказания будет выполнено вероятностное неравенство, приведенное в лемме 3.

Если предположение о независимости сделать нельзя, то предсказание по закономерностям $D_П^1 \to P_0^\varepsilon, \ldots, D_П^n \to P_0^\varepsilon$ можно сделать другим способом. Для этого можно выбрать максимальную оценку условной вероятности. Но сама операция выбора изменит эту оценку. Поэтому надо учесть это изменение.

**Лемма 5.** Если среди оценок $\underline{h}^\beta(P_0^\varepsilon \mid D_П^1), \ldots, \underline{h}^\beta(P_0^\varepsilon \mid D_П^n)$ оценка $\underline{h}^\beta(P_0^\varepsilon \mid D_П^k)$ является максимальной, и мы включим в процедуру нахождения лучшей закономерности выбор закономерности $D_П^k \to P_0^\varepsilon$ среди остальных по критерию максимальности, то оценка вероятности изменится и станет равной $\underline{h}^{n\beta}(P_0^\varepsilon \mid D_П^k)$, т.е.

$$\wp(\wp(P_0^\varepsilon \mid D_П^k) \ge \underline{h}^{n\beta}(P_0^\varepsilon \mid D_П^k)) \ge 1 - n\beta .$$

**Доказательство**. Проводя такие же рассуждения, что и при выводе неравенства (9) можно получить неравенство

$$\wp((\wp(P_0^\varepsilon \mid D_П^1) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_П^1)) \,\&\, \ldots \,\&\, (\wp(P_0^\varepsilon \mid D_П^n) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_П^n))) \ge 1 - n\beta .$$

Из конъюнкции неравенств следует, что такое неравенство должно выполняться отдельно для каждого члена, в том числе и того, у которого оценка $\underline{h}^\beta(P_0^\varepsilon \mid D_П^k)$ максимальна. Поэтому вероятность, стоящая в левой части неравенства, только увеличивается, и само неравенство сохранится, если мы конъюнкцию неравенств заменим одним неравенством $\wp(P_0^\varepsilon \mid D_П^k) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_П^k)$, в котором оценка $\underline{h}^\beta(P_0^\varepsilon \mid D_П^k)$ – максимальна. Тогда получим неравенство $\wp(\wp(P_0^\varepsilon \mid D_П^k) \ge \underline{h}^\beta(P_0^\varepsilon \mid D_П^k)) \ge 1 - n\beta$. Остается только индекс $\beta$ у оценки $\underline{h}^\beta(P_0^\varepsilon \mid D_П^k))$ заменить на индекс $n\beta$, поскольку в силу неравенства доверительный уровень равен $n\beta$ ■.

Действуя таким образом, мы свели все закономерности к одной $D_{П}^{k} \to P_0^{\varepsilon}$ и вычислили для нее оценку. Поэтому отображение $\nu$ определяется также как и в предыдущих случаях.

Рассмотрим случай, когда множества $Reg_1(Th)$ и $Reg_2(Th)$ состоят из закономерностей $D_{П_1}^{1} \to P_0^{\varepsilon}, \ldots, D_{П_1}^{n_1} \to P_0^{\varepsilon}$, предсказывающих литеру $P_0^{\varepsilon}$, и закономерностей $D_{П_2}^{1} \to \overline{P}_0^{\varepsilon}, \ldots, D_{П_2}^{n_2} \to \overline{P}_0^{\varepsilon}$, предсказывающих отрицание этой же литеры. Осуществим, принимая предположение о независимости (лемма 3) или не принимая никаких предположений (лемма 5) предсказание литеры $P_0^{\varepsilon}$ по первым закономерностям и предсказание литеры $\overline{P}_0^{\varepsilon}$ по вторым. Получим соответствующие оценки. На этом можно остановиться, получив предсказание двух различных множеств $PS(P_0^{\varepsilon})$ и $PS(\overline{P}_0^{\varepsilon})$, но можно получить и результирующее предсказание.

Предположение, аналогичное предположению о независимости, в данном случае принять нельзя, так как предсказания литеры $P_0^{\varepsilon}$ противоречат предсказаниям отрицания этой же литеры $\overline{P}_0^{\varepsilon}$. В этом случае можно действовать также как в лемме 5. Для этого достаточно заметить, что в доказательстве леммы 5 могут использоваться не только закономерности, предсказывающие $P_0^{\varepsilon}$, но и закономерности, предсказывающие $\overline{P}_0^{\varepsilon}$. Тогда выберем из всех закономерностей $Reg_1(Th) \cup Reg_2(Th)$ одну с максимальной оценкой условной вероятности. Отображение $\nu$ определяется также как и в предыдущем случае.

### 4.4 Предсказание по закономерностям $Reg_1(Th) \cup Reg_2(Th)$ в общем случае.

Пусть $(P_0^{\varepsilon_0})_1, (\overline{P}_0^{\varepsilon_0})_1, \ldots, (P_0^{\varepsilon_0})_l, (\overline{P}_0^{\varepsilon_0})_l$ – все литеры или/и их отрицания предсказываемые закономерностями $Reg_1(Th) \cup Reg_2(Th)$. Найдем оценки условных вероятностей для этих литер, как указано выше. Пусть $(P_0^{\varepsilon_0})_1, (\overline{P}_0^{\varepsilon_0})_1, \ldots, (P_0^{\varepsilon_0})_l, (\overline{P}_0^{\varepsilon_0})_l$ – все литеры, для которых получены соответствующие оценки. Эти литеры определяют множества $PS((P_0^{\varepsilon_0})_1), PS((\overline{P}_0^{\varepsilon_0})_1), \ldots, PS((P_0^{\varepsilon_0})_l), PS((\overline{P}_0^{\varepsilon_0})_l)$. По этим множествам нужно получить предсказание для пересечения любого числа из этих множеств. На основании каких дополнительных предположений это можно сделать? Наиболее естественным нам представляется следующее предположение о независимости (здесь независимость понимается в другом смысле и относится к другим закономерностям).

**Определение 4.** Закономерности $D_{П}^{1} \to (P_0^{\varepsilon_0})_1, \ldots, D_{П}^{n} \to (P_0^{\varepsilon_0})_n$ будем называть *независимыми*, если имеет место равенство

$$\wp((P_0^{\varepsilon_0})_1 \,\&\, ... \,\&\, (P_0^{\varepsilon_0})_n \mid D_П^1 \,\&\, ... \,\&\, D_П^n) = \prod_{i=1}^{n} \wp((P_0^{\varepsilon_0})_i \mid D_П^i)\,. \qquad (10)$$

**Лемма 6.** Если закономерности $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ независимы согласно определению 4, и для каждой из них выполнено неравенство

$$\wp(\wp((P_0^{\varepsilon_0})_i \mid D_П^i) \ge \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)) \ge 1-\beta_i\,,\ i = 1,2,...,k\,,$$

то имеет место неравенство

$$\wp(\wp((P_0^{\varepsilon_0})_1 \,\&\, ... \,\&\, (P_0^{\varepsilon_0})_k \mid D_П^1 \,\&\, ... \,\&\, D_П^k) \ge \underline{h}^{\gamma}((P_0^{\varepsilon_0})_1 \,\&\, ... \,\&\, (P_0^{\varepsilon_0})_k \mid D_П^1 \,\&\, ... \,\&\, D_П^k)) \ge 1-\gamma\,,$$

$$\gamma = \sum_{i=1}^{k} \beta_i\,,$$

$$\underline{h}^{\gamma}((P_0^{\varepsilon_0})_1 \,\&\, ... \,\&\, (P_0^{\varepsilon_0})_k \mid D_П^1 \,\&\, ... \,\&\, D_П^k) = \prod_{i=1}^{k} \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)\,.$$

**Доказательство**. Из условия следует, что для каждого $i = 1,2,...,k$ выполняется неравенство $\wp(\wp((P_0^{\varepsilon_0})_i \mid D_П^i) < \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)) < \beta_i$. Отсюда следует, что

$$\wp((\wp((P_0^{\varepsilon_0})_1 \mid D_П^1) \ge \underline{h}^{\beta_1}((P_0^{\varepsilon_0})_1 \mid D_П^1)) \,\&\, ... \,\&\, (\wp((P_0^{\varepsilon_0})_k \mid D_П^k) \ge \underline{h}^{\beta_k}((P_0^{\varepsilon_0})_k \mid D_П^k))) \ge 1 - \sum_{i=1}^{k} \beta_i\,.$$

Из истинности конъюнкции, стоящей в левой части неравенства следует, что

$$\prod_{i=1}^{k} \wp((P_0^{\varepsilon_0})_i \mid D_П^i) \ge \prod_{i=1}^{k} \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)\,.$$

Отсюда следует, что

$$\wp(\prod_{i=1}^{k} \wp((P_0^{\varepsilon_0})_i \mid D_П^i) \ge \prod_{i=1}^{k} \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)) \ge 1 - \sum_{i=1}^{k} \beta_i\,.$$

Заменяя первое произведение на условную вероятность, согласно равенству (10), получим утверждение леммы ■.

Пусть у нас есть пересечение множеств, взятых из подмножества множества $\{PS((P_0^{\varepsilon_0})_1), ..., PS((P_0^{\varepsilon_0})_k)\}$, например, $PS((P_0^{\varepsilon_0})_1) \cap ... \cap PS((P_0^{\varepsilon_0})_l)$. Определим отображение $\nu$ следующим образом: $\nu(pr) = 0$, если $pr \notin PS((P_0^{\varepsilon_0})_1) \cap ... \cap PS((P_0^{\varepsilon_0})_l)$, и $\nu(pr) = \prod_{i=1}^{l} \underline{h}^{\beta_i}((P_0^{\varepsilon_0})_i \mid D_П^i)$, если $pr \in PS((P_0^{\varepsilon_0})_1) \cap ... \cap PS((P_0^{\varepsilon_0})_l)$. Такое предсказание, в предположении независимости, имеет оценку, приведенную в лемме 6. Другие возможности использования предсказаний приведены в лемме 10.

### 4.5 Предсказание по закономерностям $Reg_3(Th)$.

Рассмотрим сначала случай, когда $Reg_3(Th)$ состоит из одной закономерности $D_П \to P_0^{\varepsilon}$, а множество $П$ для данной закономерности состоит из индивидной постоян-

ной $z$. Для такой закономерности в был получен толерантный интервал $[\underline{h}^z, \overline{h}^z]$, для которого выполняется неравенство (5). Для каждой модели $pr \in PS$, подставляя вместо индивидной постоянной $z$ объект $b$ (см. вывод формулы (4)), можно подсчитать частоту $h_z^b(pr)$. Определим в качестве предсказания подмножество $PS(D_П \to P_0^\varepsilon) \subset PS$ тех моделей $pr$, для которых $\underline{h}^z \le h_z^b(pr) \le \overline{h}^z$. В силу неравенства (5) для заданных $\beta$ и $\gamma$ будет иметь место неравенство

$$\wp(\wp(pr \in PS(D_П \to P_0^\varepsilon)) \ge \gamma) \ge 1 - \beta. \qquad (11)$$

Отображение $\nu$ в данном случае можно определить следующим образом: $\nu(pr) = 0$, если $pr \notin PS(D_П \to P_0^\varepsilon)$ и $\nu(pr) = \gamma$, если $pr \in PS(D_П \to P_0^\varepsilon)$.

Рассмотрим случай произвольного множества закономерностей из $Reg_3(Th)$. Пусть есть закономерности $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ из $Reg_3(Th)$. Для получения результирующего предсказания можно также, либо ввести предположение о независимости, либо не делать дополнительных предположений.

**Определение 5.** Закономерности $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ будем называть независимыми относительно индивидных постоянных $z_1 \in П_1, \ldots, z_k \in П_k$, если для случайных величин $h_{z_1}^b, \ldots, h_{z_k}^b$, имеет место равенство

$$\wp((\underline{h}^{z_1} \le h_{z_1}^b \le \overline{h}^{z_1}) \& \ldots \& (\underline{h}^{z_k} \le h_{z_k}^b \le \overline{h}^{z_k})) = \prod_{i=1}^{k} \wp(\underline{h}^{z_i} \le h_{z_i}^b \le \overline{h}^{z_i}).$$

**Лемма 7.** Если закономерности $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ независимы относительно индивидных постоянных $z_1, \ldots, z_k$ и для них выполнены неравенства

$$\wp(\wp(\underline{h}^{zi} \le h_{zi}^b \le \overline{h}^{zi}) \ge \gamma_i) \ge \beta_i,\ i = 1, \ldots, k,$$

то имеет место неравенство

$$\wp(\wp((\underline{h}^{z_1} \le h_{z_1}^b \le \overline{h}^{z_1}) \& \ldots \& (\underline{h}^{z_k} \le h_{z_k}^b \le \overline{h}^{z_k})) \ge \gamma) \ge 1 - \beta, \qquad (12)$$

$$\gamma = \prod_{i=1}^{k} \gamma_i,\ \beta = \sum_{i=1}^{k} \beta_i.$$

**Доказательство**. Проводится также, как и доказательство леммы 6.

Каждой закономерности соответствует предсказание подмножества $PS(D_П^i \to (P_0^{\varepsilon_0})_i) \subset PS$, $i = 1, \ldots, k$. Используя лемму 7, можно, в предположении независимости, получить предсказание и соответствующие оценки для пересечения произвольного числа этих множеств. Например, если предположение о независимости можно сде-

лать для всех $k$ закономерностей, то отображение $\nu$ можно определить следующим образом:

$\nu(pr) = 0$, если $pr \notin PS(D_П^1 \to (P_0^{\varepsilon_0})_1) \cap ... \cap PS(D_П^k \to (P_0^{\varepsilon_0})_k)$ и

$\nu(pr) = \gamma = \prod_{i=1}^{k} \gamma_i$, если $pr \in PS(D_П^1 \to (P_0^{\varepsilon_0})_1) \cap ... \cap PS(D_П^k \to (P_0^{\varepsilon_0})_k)$.

Для так определённого предсказания будет выполнено неравенство (12).

Предположим, что мы не можем принять предположение о независимости. Тогда мы можем выбрать одно из предсказаний $PS(D_П^e \to (P_0^{\varepsilon_0})_l)$, являющееся лучшим по какому-нибудь критерию. В предыдущем случае таким критерием был максимум нижней доверительной границы, что вполне естественно для закономерностей $Reg_1(Th)$ и $Reg_2(Th)$. Для закономерностей $Reg_3(Th)$ может быть много различных критериев, зависящих от типов шкал, от соотношения желаемой точности и надежности предсказания и т.д. Поэтому в данном случае необходимо определить некоторую функцию $F$ на множестве $PS$, определяющую критерий качества предсказания.

Пусть есть некоторые закономерности $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ и индивидные постоянные $z_1 \in П_1, \ldots, z_k \in П_k$. По ним можно найти предсказания $PS(D_П^1 \to (P_0^{\varepsilon_0})_1), \ldots, PS(D_П^k \to (P_0^{\varepsilon_0})_k)$ удовлетворяющие неравенству (11). В качестве результирующего предсказания можно взять предсказание, на котором значение критерия качества $F$ максимально. Примеры таких функций приведены ниже, а также в следующем параграфе. Найдем оценку вероятности такого предсказания, которое будет включать в себя процедуру выбора лучшего предсказания в соответствии со значением критерия $F$.

**Лемма 8.** Если для закономерностей $D_П^1 \to (P_0^{\varepsilon_0})_1, \ldots, D_П^k \to (P_0^{\varepsilon_0})_k$ относительно переменных $z_1 \in П_1, \ldots, z_k \in П_k$ выполнены неравенства

$$\wp(\wp(pr \in PS(D_П^i \to (P_0^{\varepsilon_0})_i)) \ge \gamma_i) \ge 1 - \beta_i,\ i = 1, 2, ..., k,$$

То для закономерности с номером l, выбираемой в соответствии со значением критерия $F$, выполнено неравенство

$$\wp(\wp(pr \in PS(D_П^l \to (P_0^{\varepsilon_0})_l)) \ge \gamma_l) \ge 1 - \sum_{i=1}^{k} \beta_i . \qquad (13)$$

**Доказательство**. Проводиться аналогично доказательству леммы 5.

Закономерности из леммы 8 не обязательно все различны, поэтому для одной закономерности можно брать сразу несколько индивидных постоянных $z_1, ..., z_e \in П$. Таким образом, если не делать никаких предположений относительно закономерностей из

$Reg_3(Th)$, то можно получить следующее предсказание: выбрать из всех (из части) закономерностей $Reg_3(Th)$ по всем индивидным постоянным, входящим во множество $\Pi$ этих закономерностей, ту закономерность $D^l_\Pi \to (P^{\varepsilon_0}_0)_l$ и индивидную постоянную $z$, дающую наилучшее предсказание в смысле выбранного критерия $F$. Для полученного предсказания будет выполнено неравенство (13). Отображение $\nu$ тогда определяется следующим образом:

$\nu(pr) = 0$, если $pr \notin PS(D^l_\Pi \to (P^{\varepsilon_0}_0)_l)$ и

$\nu(pr) = \gamma_e$, если $pr \in PS(D^l_\Pi \to (P^{\varepsilon_0}_0)_l)$.

Леммы 7, 8 дают возможность также получить оценку предсказания для случая, когда предположение о независимости можно сделать только относительно части закономерностей из $Reg_3(Th)$. В этом случае для закономерностей и индивидных постоянных, для которых можно сделать предположение о независимости, вычисление оценки осуществляется в соответствии с леммой 7, а после этого результирующее предсказание получается в соответствии с леммой 8.

Приведем наиболее естественные, с нашей точки зрения, критерии качества предсказания $F$ для разных шкал.

Если предсказываемое отношение $P^{\varepsilon}_0$ является отношением порядка $\leq$, то для одних и тех же значений $\gamma$ и $\beta$ то предсказание лучше, которое содержит в себе больше точек материала обучения. Точкам материала обучения $a \in \mathbb{D}$ соответствуют такие предсказания $pr_a \in PS$, для которых $a \leq b$ либо $b \leq a$. Поэтому наилучшим предсказанием с точки зрения этого критерия будет такое предсказание $Q \subset PS$, которое содержит максимальное число моделей из множества $\{pr_{a_1}, ..., pr_{a_m}\}$, $\mathbb{D} = \{a_1, ..., a_m\}$. Значение критерия $F(Q)$ – есть полученное максимальное число.

Если предсказываемая величина измеряется в шкале отношений, то имеет смысл говорить о величине интервала предсказания. Каждой модели $pr \in PS$ можно поставить в соответствие интервал $l_{pr}$ тех значений, соответствующих объекту $b$, которые определяются моделью $pr$. Критерий $F$ тогда можно определить как сумму всех интервалов, входящих в предсказание $Q \subset PS : F(Q) = \sum_{pr \in Q} l_{pr}$. По максимуму этого критерия можно определить наилучшее предсказание.

### 4.6 Общий случай произвольных множеств $Reg_1$, $Reg_2$, $Reg_3$.

Получим предсказание для закономерностей $\boldsymbol{Reg_1}$ и $\boldsymbol{Reg_2}$, как указано в леммах 5, 6 и по закономерностям из $\boldsymbol{Reg_3}$, как указано в леммах 7, 8. Рассмотрим, также, как и предыдущих случаях, два способа получения результирующего предсказания: с использованием некоторого дополнительного предположения о независимости и без него. Рассмотрим первый способ.

**Определение 6.** Пусть есть некоторые закономерности $\boldsymbol{D^1_{П1} \to (P_0^{\varepsilon_0})_1, \ldots, D^n_{П1} \to (P_0^{\varepsilon_0})_n}$ из $\boldsymbol{Reg_1 \cup Reg_2}$ и закономерности $\boldsymbol{D^1_{П2} \to (P_0^{\varepsilon_0})_1, \ldots, D^k_{П2} \to (P_0^{\varepsilon_0})_k}$ из $\boldsymbol{Reg_3}$, рассматриваемые относительно индивидных постоянных $\boldsymbol{z_1 \in П_1, \ldots, z_k \in П_k}$. Событие

$$\boldsymbol{h = ((\underline{h}^{z_1} \le h^b_{z_1} \le \overline{h}^{z_1}) \& \ldots \& (\underline{h}^{z_k} \le h^b_{z_k} \le \overline{h}^{z_k}))}$$

и условное событие

$$\boldsymbol{((P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n \mid D^n_{П1} \& \ldots \& D^n_{П1})}$$

будем называть независимыми, если

$$\boldsymbol{\wp((P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n \& h \mid D^n_{П1} \& \ldots \& D^n_{П1}) = \wp((P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n \mid D^n_{П1} \& \ldots \& D^n_{П1}) \wp(h)}.$$

**Лемма 9.** Если для закономерности $\boldsymbol{D^n_{П1} \& \ldots \& D^n_{П1} \to (P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n}$ справедлива оценка

$$\boldsymbol{\wp(\wp((P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n \mid D^n_{П1} \& \ldots \& D^n_{П1}) \ge \underline{h}^{\beta_1}) \ge 1 - \beta_1},$$

приведенная в леммах 5, 6, а для закономерностей $\boldsymbol{D^1_{П2} \to (P_0^{\varepsilon_0})_1, \ldots, D^k_{П2} \to (P_0^{\varepsilon_0})_k}$ справедлива оценка

$$\boldsymbol{\wp(\wp(pr \in PS(D^1_{П2} \to (P_0^{\varepsilon_0})_1) \cap \ldots \cap PS(D^k_{П2} \to (P_0^{\varepsilon_0})_k)) \ge \gamma) \ge 1 - \beta_2},$$

приведенная в леммах 7, 8, и соответствующие события независимы, в соответствии с определением 8, то справедлива следующая оценка

$$\boldsymbol{\wp(\wp((P_0^{\varepsilon_0})_1 \& \ldots \& (P_0^{\varepsilon_0})_n \& h \mid D^n_{П1} \& \ldots \& D^n_{П1}) \ge \gamma^1) \ge 1 - \beta},\ \boldsymbol{\gamma^1 = \underline{h}^{\beta_1} \gamma},\ \boldsymbol{\beta = \beta_1 + \beta_2}.$$

**Доказательство**. Проводится аналогично доказательству леммы 6.

Если предположение о независимости сделать нельзя, то предсказание можно получить путем выбора наилучшего из них в соответствии со значением критерия качества $\boldsymbol{F}$. Для этого критерий $\boldsymbol{F}$ надо определить для множества $\boldsymbol{PS(D^1_{П2} \to (P_0^{\varepsilon_0})_1) \cap \ldots \cap PS(D^k_{П2} \to (P_0^{\varepsilon_0})_k)}$ и для множества $\boldsymbol{PS((P_0^{\varepsilon_0})_1) \cap \ldots \cap PS((P_0^{\varepsilon_0})_n)}$. Оценка выбранного предсказания находится также, как и в лемме 5.

Отображение $\boldsymbol{\nu}$ определяется также, как и в предыдущих случаях.

Рассмотрим дополнительные возможности получения предсказания. Закономерности из $Reg_1 \cup Reg_2 \cup Reg_3$ определяют различные подмножества: $PS((P_0^{\varepsilon_0})_i)$ или $PS(D_{\Pi}^{j} \to (P_0^{\varepsilon_0})_j)$. Все рассмотренные предсказания состоят либо из этих подмножеств, либо из различных их пересечений. Для того, чтобы расширить возможности получения предсказаний, рассмотрим предсказания в виде объединения каких-либо подмножеств множества $PS$.

**Лемма 10.** Если множества $PS_1,\ldots,PS_n$; $PS_i \in PS$, $i = 1,2,...,n$ попарно не пересекаются, и для каждого из них справедлива оценка $\wp(\wp(PS_i) \ge \gamma_i) \ge 1-\beta_i$, $i = 1,2,...,n$, то для объединения этих множеств справедлива оценка

$$\wp(\wp(\bigcup_{i-1}^{n} PS_i) \ge \sum_{i=1}^{n}\gamma_i) \ge 1-\sum_{i=1}^{n}\beta_i .$$

**Доказательство**. Из условия вытекает следующее неравенство

$$\wp((\wp(PS_1) \ge \gamma_1)\,\&\ldots\&\,(\wp(PS_1) \ge \gamma_n)) \ge 1-\sum_{i=1}^{n}\beta_i .$$

Если высказывание, стоящее в левой части неравенства истинно, то

$$\sum_{i=1}^{n}\wp(PS_i) \ge \sum_{i=1}^{n}\gamma_i .$$

Но так как множества $PS_i$, $i = 1,2,...,n$ попарно не пересекаются, то

$$\sum_{i=1}^{n}\wp(PS_i) = \wp(\bigcup_{i-1}^{n} PS_i) \ \blacksquare.$$

Таким образом, мы рассмотрели различные возможности определения алгоритма $AP$.

## §5. Метод принятия решений

В практических задачах полученные предсказания, как правило, используются для принятия определенных решений. К примеру, в финансовых задачах полученные предсказания о дальнейшем движении цен могут быть использованы для принятия решения о покупки или продажи соответствующей ценной бумаги. Рассмотрим общий механизм принятия решений на основе предсказаний.

Пусть нам известно множество вариантов решений $S = \{s_1,...,s_n\}$. Предположим также, что нам известно, какое одно конкретное решение должно быть выбрано в том случае, если бы нам было известно, какая модель из $PS$ окажется истинной. Это означает, что для каждого решения $s \in S$ можно указать множество моделей $PS(s) \subset PS$, для кото-

рых должно быть выбрано данное решение. Будем называть множества $PS(s_i)$, $i = 1,...,n$ вариантами исхода. Отметим также, что поскольку для каждой модели может быть выбрано только одно решение, то множества $PS(s_i)$, $i = 1,...,n$ попарно не пересекаются.

Поскольку реально нам не известна модель, которая на самом деле окажется истинной, то задачей принятия решения является выбор одного варианта решения, основываясь на прогнозах различных вариантов исходов. Таким образом, метод принятия решений мы можем определить как функцию $Dec : \{\langle \lambda(PS(s_1)),...,\lambda(PS(s_n))\rangle\} \to \{S \cup \varnothing\}$, которая для каждого набора оценок вероятностей исходов $\langle \lambda(PS(s_1)),...,\lambda(PS(s_n))\rangle$ возвращает один выбранный вариант решения $s \in S$ либо $\varnothing$ в том случае, если решение не может быть принято.

Прежде чем переходить к построению метода принятия решений, необходимо задать функцию оценки вероятностей исходов $\lambda : \{PS(s_1),..., PS(s_n)\} \to [0,1]$, которая для любого исхода $PS(s) \subset PS$ вычисляет результирующую оценку вероятности данного исхода, основываясь на оценках $\nu(pr)$ всех моделей $pr \in PS(s)$, входящих в этот исход $PS(s)$.

Способы определения функции $\lambda$ могут быть различны и зависят от специфики решаемой задачи. Приведем два примера определения функции $\lambda$.

1. По максимальной оценке: $\lambda(PS(s)) = \max\limits_{pr \in PS(s)} \{\nu(pr)\}$.
2. По средней оценке: $\lambda(PS(s)) = \dfrac{\sum\limits_{pr \in PS(s)} \nu(pr)}{n(PS(s))}$, если $n(PS(s)) \neq 0$, и $\lambda(PS(s)) = 0$, если $n(PS(s)) = 0$. Здесь $n(PS(s))$ – количество моделей во множестве $PS(s)$.

Для определения метода принятия решения в различных задачах также могут быть использованы различные способы задания функции $Dec$. В качестве одного достаточно универсального способа определения функции $Dec$ можно предложить следующий вариант, учитывающий при принятии решения согласованность прогнозов различных вариантов исходов.

Для каждого варианта исхода $PS(s_i)$, $i = 1,...,n$ рассчитывается показатель согласованности его прогноза по формуле $Ctr_i = \kappa(PS(s_i)) - \max\limits_{j \neq i} \{\kappa(PS(s_j))\}$, т.е. как разность между оценкой вероятности данного исхода и максимальной оценкой вероятности остальных исходов. В качестве окончательного решения выбирается решение, соответствующее исходу, показатель согласованности которого строго больше заданного порога

$0 < \delta < 1$, т.е. $Dec = s_k$, где $k = \arg\max_{i=1,\ldots,n} \{Ctr_i : Ctr_i > \delta\}$. Порог $\delta$ будем называть *порогом согласованности*. В случае, если не существует исхода, показатель согласованности которого выше указанного порога, то решение не принимается и $Dec = \varnothing$. Таким образом, регулируя порог согласованности $\delta$, можно регулировать степень уверенности при принятии решений.

## БЛАГОДАРНОСТИ

Работа поддержана грантом РФФИ 08-07-00272-а; интеграционными проектами СО РАН №№ 1, 115, а также работа выполнена при финансовой поддержке Государственного контракта 2007-4-1.4-00-04 и Совета по грантам Президента РФ и государственной поддержке ведущих научных школ (проект НШ-335.2008.1).

## ЛИТЕРАТУРА

1. Демин А.В., Витяев Е.Е. Разработка универсальной системы извлечения знаний «Discovery» и ее применения. Вестник НГУ, Новосибирск, 2008, (в печати).
2. Демин А.В., Витяев Е.Е. Реализация универсальной системы извлечения знаний «discovery» и ее применение в задачах финансового прогнозирования // Вычислительные системы, Новосибирск, 2008, настоящий сборник.
3. Витяев Е.Е. Анализ данныз в языках эмпирических систем. Диссертация на соискание ученой степени кандидата технических наук, Новосибирск, 1983, р.192.
4. Витяев Е.Е. Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов: Моногр. // Новосибирский гос. ун-т. Новосибирск, 2006. 293 с.
5. Model-theoretic logics. Eds: J. Barwise, S. Feferman, Springer-Verlag, NY, 1985.
6. Дискретная математика и математические вопросы кибернетики / Под ред. Яблонского С.В., Лупанова О.Б., т. 1. – М.: Наука, 1974. – 311 с.
7. Scheffe H., Tukey J.M. Non-parametric estimation 1. Validation of order statistics. – Ann. Math. Stat., v.16, 1945, p. 187-192.