S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 6, стр. 340–365 (2009) УДК 510.646+.647
MSC 03B48

СИНТЕЗ ЛОГИКИ, ВЕРОЯТНОСТИ И ОБУЧЕНИЯ: ФОРМАЛИЗАЦИЯ ПРЕДСКАЗАНИЯ

С. О. СМЕРДОВ, Е. Е. ВИТЯЕВ

Abstract. Presented paper is devoted to the question of prediction formalized in probabilistic and logical terms. The aim of investigation is to examine different methods such as based on SLD-inferences and alternative semantic approach. Prediction is introduced as a statement of abductive sort attained by inductive schemes. One of the significant problems concerns unregulated decrease of trusting estimations for regularities obtained during the process of inference organized by analogy with syntax logical systems. Suggested semantic approach generalizes the notion of inference and reveals essential advantages in many aspects without assuming rather strong constraints. In particular, a special set of probabilistic laws is synthesized inductively, this collection has an optimal ability to predict (in the context of available data). Semantic definition of prediction leads us to a new paradigm, where deduction is replaced with computability concept: it rises conditional probability during the steps of inference (in contrast to SLD) and also maximally specifies resulted prediction rule. Moreover, we prove that probabilistic estimations obtained by semantic predictions are greater or equal to those by corresponding SLD-analogical systems. In conclusion practical applications are discussed.

Keywords: prediction; explanation; probability, logic & learning synthesis; probabilistic logic programming; relational data mining; scientific discovery.

Smerdov, S.O., Vityaev, E.E., Probability, logic & learning synthesis: formalizing prediction concept.

Работа поддержана РФФИ (грант 08-07-00272-а), интеграционными проектами СО РАН 47, 111, 119, а также Советом по грантам Президента РФ и государственной поддержке ведущих научных школ (проект НШ-335.2008.1).

Поступила 9 июня 2009 г., опубликована 7 ноября 2009 г.

1. Предварительные сведения и обзор

Разрабатываемые методы *искусственного интеллекта* (*artificial intelligence*) в области синтеза логико-вероятностных подходов [21] нередко тяготеют к двум популярным направлениям: *байесовы логические программы* [22] и *логика с независимым выбором* [23]. Опорные положения обоих, однако, подразумевают независимость составных элементов задействованных правил — в то время как одна из важнейших функций, отведённых вероятностной аргументации, заключена в умелом использовании доступной нам информации с намерением *предсказать/объяснить* свойства малоизученных объектов. Таким образом, противоречие наблюдается при самой интуитивной постановке задачи, ибо взаимосвязь упомянутых частей продиктована как формулировкой, так и средствами решения (закономерностями, по которым происходит прогноз): нам нужно именно выявить наиболее естественные и ожидаемые характеристики для интересующего нас объекта через родство с уже данными. Возникающие трудности подчас сводятся к следующему ключевому моменту: обобщая понятие истины (до вероятности), авторы статей всё же стараются полностью сохранить "синтаксичность" вывода, его дедуктивность; с другой стороны, речь о предсказании, т.е. получении нового знания вероятностной / статистической природы — стало быть, об индуктивных и абдуктивных схемах. Классика логического вывода справедлива в пределах одной модели, но в нашем распоряжении целый их класс (по нему строится вероятностная мера на возможных мирах), когда вероятностные закономерности имеют место быть на некоторых подмножествах этого объемлющего класса. Применение, скажем, двух правил подряд в ходе дедуктивных рассуждений означало бы взятие пересечения соответствующих подклассов: организация такого типа сталкивается с резким убыванием вероятностных оценок по мере продвижения вывода; более того, означенную проблему нельзя разрешить выбором, например, лишь закономерностей близкой к единице условной вероятности. Эти, а также иные сопутствующие трудности обуславливают выявленную ещё в трудах Карла Гемпеля [4, 5] невозможность прямолинейных перенесений логической аргументации (оперирующей средствами, по умолчанию полагаемыми абсолютными и заведомо достоверными) на *индуктивно произведённое знание*.

Тем не менее, формализм *логического программирования* [6] представляется достаточно удобным в описании идеи предсказания (хотя стоит всё же чётко понимать, что успешность его вероятностных аналогов действует по сути своей "вслепую", часто требуя немалых, порой даже неподъёмных, вычислительных затрат на реализацию) и, раз логические значения суть частные случаи вероятностных, вполне согласован с целями. По указанной причине мы посвятим §§3–4 определению предсказания в терминах вероятностных разновидностей резолюций логического программирования, разбору сопряжённых ситуаций и некоторым алгоритмам. Отметим, несмотря на то, что вопрос нахождения множества вероятностных моделей по заданному списку ожидаемых ограничений (§4, fix-point аппарат см. в работе [9]) заслуживает должного внимания, как только нами фиксированна какая-либо вероятностная мера — используемый тип правил вместе с организацией вывода в виде SLD-подобных структур оказываются на практике слабоприемлемыми и малоэффективными в решении задачи осуществления предсказания и поиска соответствующих оценок.

Далее сама собой возникает мысль о нахождении не только альтернативного подхода, обобщающего понятие вывода вслед за обобщением истинностных значений до вероятностных (см. §2), но и оптимальных средств (программы) для осуществления предсказания. Таковая программа с наилучшей предсказательной способностью нами строится в §5, являя собой наглядное и притом не избыточное представление исходного знания с помощью закономерностей. С точки зрения ряда работ её разумно было бы назвать *обученной программой* (см. *обучение из доказательств* [13, 19, 20]), однако, заметим: синтез оптимальной программы не требует предварительной процедуры построения SLDF(SLDp)-дерева (а поиск его, опять же, мог оказаться весьма ресурснозатратным). Применение правил синтезированной программы способно обеспечить значительное ускорение поиска успешных SLDF(SLDp)-выводов прежними конструкциями, заменить дедуктивный вывод — *вычислением* (точнее, пересчётом вероятностных оценок параллельно с истинностью на классе моделей) и, в результате, переосмыслить саму концепцию предсказания с привлечением т.н. *семантического μ-вывода*, сводя итоговую парадигму к обнаружению *максимально специфичных μ-законов* — изложенный план действий есть расширенная формализация идей, заявленных в [1, 2, 3]. Предлагаемый механизм *семантического μ-предсказания* не делает излишних шагов при уточнении правил в ходе реализации, причём вероятностная оценка только возрастает (при желании, процесс можно остановить на определённом этапе — закономерность будет удовлетворять данным); он не может потерпеть неудачу и завершиться неуспехом, коль скоро был успешно начат (в отличие от вероятностных SLD-предсказаний, разобранных в §§3–4); наконец, его результирующие оценки аппроксимируют (сверху) аналоги, основывающиеся на вероятностного сорта системах логического программирования.

Шестой параграф даёт читателю представление о принципах индуктивного синтеза закономерностей, механизме применения предсказания к статистически значимым (относительно вероятностной меры) наборам данных. В заключении обсуждаются приложения на практике (по материалу §7).

2. Вероятность на множестве основных предложений

В первую очередь зафиксируем язык первого порядка $\mathfrak{L}$ конечной сигнатуры

$$\Sigma = \{\mathrm{P}_1, \mathrm{P}_2, ..., \mathrm{P}_{n_1}; f_1, f_2, ..., f_{n_2}; \mathrm{c}_1, \mathrm{c}_2, ..., \mathrm{c}_{n_3}\},$$

где $n_1 \geqslant 1$, $n_2 \geqslant 0$, $n_3 \geqslant 1$; с операциями $\wedge$, $\vee$ и $\neg$. Пусть X — счётное множество переменных. Обозначим: $\mathbf{T}_\mathrm{L}$ — множество всех термов; $\mathbf{T}^0_\mathrm{L}$ — множество основных термов; $\mathbf{A}_\mathrm{L}$ — множество всех атомов; $\mathbf{A}^0_\mathrm{L}$ — множество основных атомов; $\mathbf{F}_\mathrm{L}$ — множество всех формул; $\mathbf{S}_\mathrm{L}$ — множество предложений; $\mathbf{S}^0_\mathrm{L}$ — множество основных предложений (бескванторных формул без свободных переменных). Отметим, что $\mathbf{S}^0_\mathrm{L}$ совпадает с замыканием $\mathbf{A}^0_\mathrm{L}$ относительно рассматриваемых логических операций.

Определение. *Подстановкой* назовём отображение $\theta : \mathrm{X} \mapsto \mathbf{T}_\mathrm{L}$ ($x\theta \leftrightharpoons \theta(x)$ для $x \in \mathrm{X}$). Действие каждой θ гомоморфным образом распространяется на произвольные бескванторные выражения, по индукции:

1. $\mathrm{c}\theta \leftrightharpoons \mathrm{c}$, где c — символ константы Σ;

2. $f(t_1, ..., t_n)\theta \leftrightharpoons f(t_1\theta, ..., t_n\theta)$, где $t_1, ..., t_n \in \mathbf{T}_\mathrm{L}$, f — символ n-местной функции Σ;

3. $\mathrm{P}(t_1,...,t_n)\,\theta \leftrightharpoons \mathrm{P}(t_1\theta,...,t_n\theta)$, где $t_1,...,t_n \in \mathbf{T}_\mathrm{L}$, P — символ n-местного предиката Σ;

4. $(\mathrm{A}\wedge\mathrm{B})\,\theta \leftrightharpoons (\mathrm{A}\theta\wedge\mathrm{B}\theta)$, где $\mathrm{A},\mathrm{B}\in\mathbf{F}_\mathrm{L}$;

5. $(\mathrm{A}\vee\mathrm{B})\,\theta \leftrightharpoons (\mathrm{A}\theta\vee\mathrm{B}\theta)$, где $\mathrm{A},\mathrm{B}\in\mathbf{F}_\mathrm{L}$;

6. $(\neg\mathrm{A})\,\theta \leftrightharpoons (\neg\mathrm{A}\theta)$, где $\mathrm{A}\in\mathbf{F}_\mathrm{L}$.

Далее Θ — совокупность всевозможных подстановок; $\Theta^0 \leftrightharpoons \left\{\theta|\theta : \mathrm{X}\to\mathbf{T}^0_\mathrm{L}\right\}$ — подстановки основных термов.

Опираясь анализ, проведённый в работе [10], сформулируем

Определение. *Вероятность* на $F\subseteq\mathbf{S}^0_\mathrm{L}$, замкнутом относительно $\wedge$, $\vee$ и $\neg$, есть функция $\mu : F\mapsto[0,1]$, обладающая следующими свойствами:

i. если $\vdash$ [1]ϕ («ϕ — тавтология»), то $\mu(\phi)=1$;

ii. если $\vdash\neg(\phi\wedge\psi)$, то $\mu(\phi\vee\psi)=\mu(\phi)+\mu(\psi)$.

Кроме того, легко установить

Следствие.

1. $\mu(\phi)=1-\mu(\neg\phi)$;
в частности, для $\vdash\neg\phi$ имеем $\mu(\phi)=0$;

2. если $\vdash\neg\phi\vee\psi$ (классическая импликация), то $\mu(\psi)\geqslant\mu(\phi)$;
значит, при $\vdash\psi\equiv\phi$ (формулы эквивалентны) получим $\mu(\phi)=\mu(\psi)$;

3. принцип включения и исключения ($\wedge$ ассоциируется с $\cap$, а $\vee$ — с $\cup$):
$\mu(\phi\vee\psi)=\mu(\phi)+\mu(\psi)-\mu(\phi\wedge\psi)$;

4. верны двойственные неравенства:
$\mu(\phi\vee\psi)\geqslant\max\{\mu(\phi),\mu(\psi)\}$ и $\mu(\phi\wedge\psi)\leqslant\min\{\mu(\phi),\mu(\psi)\}$.

Из пп.1–4 немедленно вытекает равносильность введённого нами определения понятию *вероятностной функции на языке логики первого порядка* (см. [11]). Требуется, однако, пояснить, как именно введённое нами определение соотносится с вероятностью классической. Укажем здесь две возможные трактовки:

$\boxed{1}$ На фактор-множестве $\mathbf{B}_\mathrm{L}\leftrightharpoons\{\phi/_\equiv|\phi\in\mathbf{F}_\mathrm{L}\}$ зададим операции $\wedge$, $\vee$ и $\neg$:

a. $(\phi/_\equiv)\wedge(\varphi/_\equiv)\leftrightharpoons(\phi\wedge\varphi)/_\equiv$;

b. $(\phi/_\equiv)\vee(\varphi/_\equiv)\leftrightharpoons(\phi\vee\varphi)/_\equiv$;

c. $\neg(\phi/_\equiv)\leftrightharpoons(\neg\phi)/_\equiv$.

Отдельно выделим в множестве $\mathbf{B}_\mathrm{L}$ наименьший $\mathbf{0}_\mathrm{L}\leftrightharpoons(\phi\wedge\neg\phi)/_\equiv$ и наибольший $\mathbf{1}_\mathrm{L}\leftrightharpoons(\phi\vee\neg\phi)/_\equiv$ элементы (разумеется, $\neg\mathbf{0}_\mathrm{L}=\mathbf{1}_\mathrm{L}$). Построенная система $\mathfrak{B}_\mathrm{L}\leftrightharpoons\langle\mathbf{B}_\mathrm{L};\wedge,\vee,\neg,\mathbf{0}_\mathrm{L},\mathbf{1}_\mathrm{L}\rangle$ является хорошо известной (булевой) алгеброй Линденбаума–Тарского. Аналогичные рассуждения, конечно же, проходят и для множества $\mathbf{B}^0_\mathrm{L}\leftrightharpoons\left\{\phi/_\equiv|\phi\in\mathbf{S}^0_\mathrm{L}\right\}$, то есть при исключении переменных. Нетрудно видеть, что μ определяется (притом, в силу п.2 следствия, корректно) на подалгебре $\mathfrak{B}^0_\mathrm{L}$, играя роль конечно аддитивной меры (напоминаем, в силу теоремы Стоуна всякая булева алгебра изоморфна некоторой алгебре подмножеств). Последнее согласуется с колмогоровским определением вероятности.

$\boxed{2}$ Ранее приведённая нами дефиниция вероятности (на F) носит аксиоматический характер. Конкретизируем конструкцию. С этой целью обратимся к $2^{\mathbf{A}^0_\mathrm{L}}\leftrightharpoons\left\{\upsilon|\upsilon:\mathbf{A}^0_\mathrm{L}\mapsto\{0,1\}\right\}$; любая $\upsilon\in2^{\mathbf{A}^0_\mathrm{L}}$ устанавливает истинность на $\mathbf{S}^0_\mathrm{L}$, в одном из «возможных миров».[2] Пусть P — произвольная вероятностная мера

[1] $\vdash$ — обычное отношение выводимости в классическом исчислении предикатов.

[2] Более того, сама υ суть пример из простейших вероятности на F — *модели Эрбрана.*

на $2^{\mathbf{A}^0_{\mathrm{L}}}$ (в качестве пространства элементарных исходов), удовлетворяющая аксиомам Колмогорова; $[\![\phi]\!] \leftrightharpoons \left\{\omega \in 2^{\mathrm{A}^0_{\mathrm{L}}} | \omega(\phi) = 1\right\}$. Полагая $\mu(\phi) \leftrightharpoons \mathsf{P}([\![\phi]\!])$ (мера миров, в которых ϕ истинна) для всех $\phi \in \mathbf{S}^0_{\mathrm{L}}$, убеждаемся в справедливости свойств 1-2, ведь связкам $\wedge$, $\vee$ явно сопоставлены $\cap$ и $\cup$ событий. При таком подходе формулы из $\mathbf{S}^0_{\mathrm{L}}$ становятся дискретными случайными величинами на $2^{\mathbf{A}^0_{\mathrm{L}}}$ со значениями $0, 1$ (положить $\phi(\omega) \leftrightharpoons \omega(\phi)$).

Мера μ обобщает наличествующее знание о возможных мирах (интерпретациях), а также об истинности формул на них. Важно осознавать согласованность вероятностных значений формул с истинностно-логическими; тем не менее, именно первые (будучи обобщением вторых) обуславливают семантику, на базе которой происходит большинство действий предсказательного аппарата (§§3–5). Классическая двузначная истинность продолжает учитываться и присутствует внутри устройства самой μ; мы вновь возвращаемся к ней на заключительном этапе осуществления предсказания (§6).

3. SLDF-вывод/-предсказание

Из соображений наглядности и во избежании оговорок будем считать, что в текущем и последующих за ним разделах вероятность μ определена нами на всём множестве $\mathbf{S}^0_{\mathrm{L}}$; иными словами, $F = \mathbf{S}^0_{\mathrm{L}}$.

Определение. Пусть $\mathrm{A}_i \in \mathbf{A}_{\mathrm{L}}$ ($i = \overline{0, m}$ для $m \geqslant 1$). Рассмотрим выражения трёх видов[3]:

- ***a.*** $\mathrm{A}_0 \Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m$;
- ***b.*** $\mathrm{A}_0 \Leftarrow$;
- ***c.*** $\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m$.

Представителей сорта (a) будем называть *правилами* (образуемое ими множество есть $\mathtt{Rule}_{\mathrm{L}}$), ($b$) — *фактами* ($\mathtt{Fact}_{\mathrm{L}}$), ($c$) — *запросами* ($\mathtt{Query}_{\mathrm{L}}$). Иногда вместо «*запрос*» пользуются термином «*цель*». При отсутствии свободных переменных (когда все $\mathrm{A}_i \in \mathbf{A}^0_{\mathrm{L}}$) подмножества, аналогичные введённым выше, именуются соответственно $\mathtt{Rule}^0_{\mathrm{L}}$, $\mathtt{Fact}^0_{\mathrm{L}}$ и $\mathtt{Query}^0_{\mathrm{L}}$. Результат действия перестановки множества X на правило, факт или же запрос называется *вариантом* подвергшегося преобразованию объекта.[4]

Определение. *Программа* представляет из себя конечный набор правил и фактов $\mathtt{Prog} \subset \mathtt{Rule}_{\mathrm{L}} \cup \mathtt{Fact}_{\mathrm{L}}$. Иначе: $\mathtt{Prog} = \mathtt{Rule}_{\mathrm{L}}[\mathtt{Prog}] \cup \mathtt{Fact}_{\mathrm{L}}[\mathtt{Prog}]$, где $\mathtt{Rule}_{\mathrm{L}}[\mathtt{Prog}] \leftrightharpoons \mathtt{Rule}_{\mathrm{L}} \cap \mathtt{Prog}$, $\mathtt{Fact}_{\mathrm{L}}[\mathtt{Prog}] \leftrightharpoons \mathtt{Fact}_{\mathrm{L}} \cap \mathtt{Prog}$. Коль скоро μ задаётся на $\mathbf{S}^0_{\mathrm{L}}$, нас будут интересовать основные частные случаи правил и фактов программы: $\mathtt{Rule}^0_{\mathrm{L}}[\mathtt{Prog}] \leftrightharpoons \mathtt{Rule}^0_{\mathrm{L}} \cap \mathtt{Prog}\Theta^0$, $\mathtt{Fact}^0_{\mathrm{L}}[\mathtt{Prog}] \leftrightharpoons \mathtt{Fact}^0_{\mathrm{L}} \cap \mathtt{Prog}\Theta^0$, где $\mathtt{Prog}\Theta^0$ состоит из результатов подстановок $\theta \in \Theta^0$ в элементы $\mathtt{Prog}$.

Чтобы оперировать вероятностями на $\mathtt{Rule}^0_{\mathrm{L}}$, $\mathtt{Fact}^0_{\mathrm{L}}$ и $\mathtt{Query}^0_{\mathrm{L}}$, требуется определить, что понимать под вероятностной мерой основного правила, факта, либо запроса. Для последних двух просто полагаем

[3]При обобщённом определении правила/факта/запроса взамен атомов могут браться конъюнкции/дизъюнкции атомов или даже произвольные бескванторные предложения; рассуждения при этом, однако, изменятся не значительно и не по существу.

[4]Договорённость: при возникновении в запросах или условных частях правил одинаковых атомов происходит их “склейка”.

$$\mu\,(\mathrm{A}_0 \Leftarrow) \leftrightharpoons \mu\,(\mathrm{A}_0) \text{ и } \mu\,(\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) \leftrightharpoons \mu\,(\mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m),$$
$$\text{где } (\mathrm{A}_0 \Leftarrow) \in \mathtt{Fact}^0_{\mathrm{L}},\ (\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) \in \mathtt{Query}^0_{\mathrm{L}}.$$

С основными правилами чуть сложнее — здесь речь идёт уже об условной вероятности (причём мера посылки должна быть ненулевой):

$$\mu\,(\mathrm{A}_0 \Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) \leftrightharpoons \mu\,(\mathrm{A}_0|\mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) = \tfrac{\mu(\mathrm{A}_0\wedge\mathrm{A}_1\wedge...\wedge\mathrm{A}_m)}{\mu(\mathrm{A}_1\wedge...\wedge\mathrm{A}_m)},$$
$$\text{где } (\mathrm{A}_0 \Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) \in \mathtt{Rule}^0_{\mathrm{L}},\ \mu\,(\mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m) \neq 0.$$

$\mathtt{Rule}^{0,\mu}_{\mathrm{L}}$, $\mathtt{Fact}^{0,\mu}_{\mathrm{L}}$ и $\mathtt{Query}^{0,\mu}_{\mathrm{L}}$ — соответствующие подмножества $\mathtt{Rule}^0_{\mathrm{L}}$, $\mathtt{Fact}^0_{\mathrm{L}}$ и $\mathtt{Query}^0_{\mathrm{L}}$, на которых μ определена.[5] $\mathtt{Rule}^{0,\mu}_{\mathrm{L}}\,[\mathtt{Prog}] = \mathtt{Rule}^{0,\mu}_{\mathrm{L}} \cap \mathtt{Rule}^0_{\mathrm{L}}\,[\mathtt{Prog}]$, $\mathtt{Fact}^{0,\mu}_{\mathrm{L}}\,[\mathtt{Prog}] = \mathtt{Fact}^0_{\mathrm{L}}\,[\mathtt{Prog}]$, $\mathtt{Query}^{0,\mu}_{\mathrm{L}}\,[\mathtt{Prog}] = \mathtt{Query}^0_{\mathrm{L}}\,[\mathtt{Prog}]$.

Зафиксируем $\mathtt{Prog}$ — некоторую программу. $\mathrm{Q} = (\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{A}_m)$ — запрос, в котором возможно наличие *выделенных*[6] атомов, чьё назначение будет прояснено несколько позже. Далее, выбирается какой-нибудь невыделенный атом A_i (здесь $1 \leqslant i \leqslant m$) запроса и такой вариант правила $\mathrm{C} = (\mathrm{B}_0 \Leftarrow \mathrm{B}_1 \wedge ... \wedge \mathrm{B}_k)$ программы, что все входящие в него переменные отличны от переменных Q. Затем в ситуации, если нашёлся унификатор $\theta \in \Theta$ атомов A_i и B_0 ($\mathrm{A}_i\theta = \mathrm{B}_0\theta$), запрос

$$\mathrm{Q}' = (\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathrm{B}_1 \wedge ... \wedge \mathrm{B}_k \wedge ... \wedge \mathrm{A}_m)\,\theta$$

называется *выводимым* из Q по правилу C при помощи подстановки θ.[7] Если в приведённых рассуждениях заменить правило на факт $\mathrm{C} = (\mathrm{B}_0 \Leftarrow)$ программы, то выводим запрос

$$\mathrm{Q}' = (\Leftarrow \mathrm{A}_1 \wedge ... \wedge \mathbf{A}_i \wedge ... \wedge \mathrm{A}_m)\,\theta$$
(A_i становится выделенным).

Замечание. Главным образом, представленная процедура отличается от классической [6] тем, что нами с необходимостью хранится информация обо всех использованных в процессе вывода фактах (происходит выделение, а не удаление), столь важная для оценки вероятности реализации запроса при условии наличествующих данных. Из определений видно: при добавлении некоторого атома к посылке в условной вероятности последняя может как убывать, так и возрастать.

Определение. *SLDF-вывод*[8] запроса Q средствами $\mathtt{Prog}$ — максимальная последовательность $\mathrm{Q}_0, \mathrm{Q}_1, ...$ (причём все $\mathrm{Q}_i \in \mathtt{Query}_{\mathrm{L}}$, а в $\mathrm{Q}_0 = \mathrm{Q}$ нет выделенных атомов), обладающая свойством: существуют $\mathrm{C}_0, \mathrm{C}_1, ...$ и $\theta_0, \theta_1, ...$ такие, что всякий Q_{i+1} выводим из Q_i по варианту правила $\mathrm{C}_i \in \mathtt{Prog}$ при помощи

---

[5] Если для основных фактов и запросов принадлежность $\mathtt{Fact}^{0,\mu}_{\mathrm{L}}$ и $\mathtt{Query}^{0,\mu}_{\mathrm{L}}$ при μ, заданной на всём $\mathbf{S}^0_{\mathrm{L}}$, означает лишь указание актуальной вероятностной меры, то множество правил $\mathtt{Rule}^{0,\mu}_{\mathrm{L}}$ совсем не обязано совпадать с $\mathtt{Rule}^0_{\mathrm{L}}$ даже в такой ситуации.

[6] Или, что то же, *подчёркнутых*; на письме отмечаются жирным шрифтом. Возвращаясь к “склеиванию” идентичных объектов: склейка выделенного атома с невыделенным — всегда даёт выделенный.

[7] Без ограничения общности, переменные запросов Q и Q′ считаем непересекающимися — порой это удобно.

[8] Linear resolution with Selection rule for Definite clauses and underlined Facts. В ходе автоматизированного поиска вывода нам потребуется стратегия R, выбирающая невыделенный атом в Q; алгоритм, определяющий приоритетное $\mathrm{C} \in \mathtt{Prog}$; функция перечисления всевозможных стратегий (какие-то из них, не исключено, приводят к неудачам); не стоит забывать и о модуле унификации двух атомов.

$\theta_i \in \Theta$. *Успешным* называется SLDF-вывод, оканчивающийся запросом-конъюнкцией выделенных атомов[9]; *тупиковым* — безуспешный конечный SLDF-вывод, в котором ни один невыделенный атом не унифицируем ни с каким из заключениий правил и фактов `Prog`.

Примечание. Нетрудно определить подстановки так, чтобы никакие два запроса в SLDF-выводе не содержали бы общих переменных. Символически: $\mathrm{X}(\mathrm{Q}_i) \cap \mathrm{X}(\mathrm{Q}_j) = \emptyset$ для любых $i \neq j$, где $\mathrm{X}(\mathrm{Q}_i)$ и $\mathrm{X}(\mathrm{Q}_j)$ — обозначения для множеств переменных Q_i и Q_j.

Отношение выводимости программой одних запросов из других позволяет говорить о *пространстве вычислений* для `Prog` на $\mathtt{Query}_{\mathrm{L}}$. Максимальный путь с началом Q в таком пространстве — и есть SLDF-вывод запроса Q средствами `Prog`; совокупность упомянутых путей принято изображать в виде дерева (т.н. *SLDF-дерева вычислений*); *успешное* SLDF-дерево обязано включать хотя бы один успешный вывод.

Успешному SLDF-выводу[10] $\mathrm{Q}_0, \mathrm{Q}_1, ..., \mathrm{Q}_{n-1}, \mathrm{Q}_n$ запроса $\mathrm{Q} = \mathrm{Q}_0$, по определению, сопоставлены $\mathrm{C}_0, ..., \mathrm{C}_{n-1} \in \mathtt{Prog}$ и $\theta_0, ..., \theta_{n-1} \in \Theta$. Тогда

$$(\mathrm{Q}_0\theta_0\theta_1...\theta_{n-1}), (\mathrm{Q}_1\theta_1...\theta_{n-1}), ..., \mathrm{Q}_{n-1}\theta_{n-1}, \mathrm{Q}_n$$

— снова успешный SLDF-вывод, но теперь для запроса $(\mathrm{Q}_0\theta_0\theta_1...\theta_{n-1})$. Эта последовательность совпадает с $\mathrm{Q}_0\theta, \mathrm{Q}_1\theta, ..., \mathrm{Q}_{n-1}\theta, \mathrm{Q}_n\theta$, где $\theta \in \Theta$ строится по $\{\theta_i\}$ так, чтобы $\theta|_{\mathrm{X}(\mathrm{Q}_i)} = \theta_i...\theta_{n-1}$ для всех $1 \leqslant i \leqslant n-1$, а на $\mathrm{X}(\mathrm{Q}_n)$ действие θ тождественно.[11] В силу примечания, θ легко задать инструкцией

$$\theta(x) \rightleftharpoons \begin{cases} \theta_i(x), & \textit{если } x \in \mathrm{X}(\mathrm{Q}_i) \ \textit{при каждом } i \in \{1, ..., n-1\} \\ x, & \textit{если } x \in \mathrm{X}(\mathrm{Q}_n) \end{cases}$$

(мы будем писать $\theta \rightleftharpoons \theta_0...\theta_{n-1}$, хоть это и не совсем композиция[12]).

Для подстановки $\theta' \in \Theta$ последовательность $\mathrm{Q}_0\theta\theta', ..., \mathrm{Q}_n\theta\theta'$ оказывается успешным SLDF-выводом $\mathrm{Q}_0\theta\theta'$ программой `Prog`. Рассмотрим основные $\theta^0 \in \Theta^0$ и $\mathrm{Q}_i^0 \rightleftharpoons \mathrm{Q}_i\theta\theta^0$; тогда SLDF-вывод $\mathrm{Q}_0^0, ..., \mathrm{Q}_n^0 \in \mathtt{Query}_{\mathrm{L}}^0$ успешно осуществляется цепочкой

$$\mathrm{C}_0^0, ..., \mathrm{C}_{n-1}^0 \in \mathtt{Prog}\Theta^0 \ (\mathrm{C}_i^0 \rightleftharpoons \mathrm{C}_i\theta\theta^0),$$

причём до запроса Q_r^0 используются только правила $\mathrm{C}_0^0, ..., \mathrm{C}_{r-1}^0 \in \mathtt{Rule}_{\mathrm{L}}^0[\mathtt{Prog}]$, после чего наступает очередь одних лишь фактов $\mathrm{C}_r^0, ..., \mathrm{C}_{n-1}^0 \in \mathtt{Fact}_{\mathrm{L}}^0[\mathtt{Prog}]$. Формально здесь нужна не составляющая труда перенумерация нижних индексов: применение фактов среди $\mathrm{C}_0^0, ..., \mathrm{C}_{n-1}^0$ пропускается, откладывается до заключительной фазы (это никак не помешает успешному исходу вывода).

Определение. Вышеописанный успешный SLDF-вывод $\mathrm{Q}_0^0, ..., \mathrm{Q}_n^0 \in \mathtt{Query}_{\mathrm{L}}^0$ с закреплёнными за ним цепочками $\mathrm{C}_0^0, ..., \mathrm{C}_{r-1}^0 \in \mathtt{Rule}_{\mathrm{L}}^0[\mathtt{Prog}]$, $\mathrm{C}_r^0, ..., \mathrm{C}_{n-1}^0 \in$

---

[9] К завершающему запросу успешного SLDF-вывода разрешено добавлять произвольные выделенные атомы.

[10] Пока в поле нашего зрения попадает всего одна программа, позволим себе время от времени опускать “средствами `Prog`” (когда это не приводит к путанице и разночтениям).

[11] Переменные полученных запросов, конечно, могут пересекаться. Нам важна именно такая запись. Однако, если указанный эффект потребуется устранить, это поправимо: в любом SLDF-выводе переобозначение переменных какого-либо промежуточного запроса $\tilde{\mathrm{Q}}_i$ подразумевает несложную корректировку двух соседних с ним унификаторов $\tilde{\theta}_{i-1}$ и $\tilde{\theta}_i$, поэтому множества переменных запросов легко развести, не теряя успешности.

[12] Впрочем, она становится таковой, если предполагать действие любого θ_i вне $\mathrm{X}(\mathrm{Q}_i)$ — тождественным.

$\in$ $\mathtt{Fact}^0_{\mathrm{L}}[\mathtt{Prog}]$ и памятью о происхождении (от Q: $\mathrm{Q}^0_0 = \mathrm{Q}\theta\theta^0$) назовём *нормализованным SLDF-выводом* Q средствами Prog.[13] $\mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog})$ — семейство всех нормализованных SLDF-выводов запроса Q программой Prog.

Замечание. Без сомнений, $\mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog}) \supseteq \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q}\theta';\mathtt{Prog})$ для всякой $\theta' \in \Theta$; вместе с тем $\mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog}) = \bigcup_{\theta^0\in\Theta^0}\mathtt{SLDF}^0_{\mathrm{L}}\left(\mathrm{Q}\theta^0;\mathtt{Prog}\right)$. Казалось бы, находить $\mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog})$ или все $\mathtt{SLDF}^0_{\mathrm{L}}\left(\mathrm{Q}\theta^0;\mathtt{Prog}\right)$ — задачи равновеликие. Меж тем, их эквивалентность на практике может и не иметь места: одно дело — совершать вывод со свободными переменными, намереваясь затем подставлять термы $\mathbf{T}^0_{\mathrm{L}}$ в него; другое — искать эти основные экземпляры-выводы по отдельности.

Пусть $\bar{\mathrm{Q}}^0 \rightleftharpoons \left(\mathrm{Q}^0_0,...,\mathrm{Q}^0_n\right)$ — нормализованный SLDF-вывод с присущими ему, по определению, атрибутами. Величина условной вероятности $\mu\left(\mathrm{Q}^0_0|\mathrm{Q}^0_n\right)$ отражает специфику зависимости основного частного случая $\mathrm{Q} \in \mathtt{Query}_{\mathrm{L}}$ от фактов $\mathtt{Prog}\Theta^0$ (напоминаем, заключительный запрос Q^0_n включает в себя все необходимые для проведения рассматриваемого вывода элементы $\mathtt{Fact}^0_{\mathrm{L}}[\mathtt{Prog}]$ и только их; мы будем позволять себе запись $\mathrm{Q}^0_n = \mathrm{C}^0_r \wedge ... \wedge \mathrm{C}^0_{n-1}$). Знание о μ подразумевается неполным[14], оно частично и относится к имеющейся в нашем распоряжении программе. Значит, разумно оценить $\mu\left(\mathrm{Q}^0_0|\mathrm{Q}^0_n\right)$ через доступную информацию относительно Prog.

Предполагаем известными:

а. вероятности самих правил $\mathtt{Rule}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}]$, а также их правых частей;

б. вероятности конъюнкций фактов из $\mathtt{Fact}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}]$.

Примем соглашение: $\mathrm{C}^0_i = \left(\mathrm{A}^0_i \Leftarrow \mathrm{Body}\left(\mathrm{C}^0_i\right)\right)$, $\mathrm{i} = \overline{1, r-1}$.

Лемма 1 (об оценке, [1]). *Если* $\underbrace{\mathrm{C}^0_0,...,\mathrm{C}^0_{r-1} \in \mathtt{Rule}^{0,\mu}_{\mathrm{L}} \text{ и величина } \mu\left(\mathrm{Q}^0_n\right) \neq 0}_{\circledast}$, *то справедливо неравенство:*

$$\mu\left(\mathrm{Q}^0_0|\mathrm{Q}^0_n\right) \geqslant \underbrace{\max\left\{0, 1 - \sum_{i=1}^{r-1}\frac{\left(1-\mu\left(\mathrm{C}^0_i\right)\right)\mu\left(\mathrm{Body}\left(\mathrm{C}^0_i\right)\right)}{\mu\left(\mathrm{C}^0_r \wedge ... \wedge \mathrm{C}^0_{n-1}\right)}\right\}}_{\divideontimes}.$$

Определение. *Оценкой нормализованного SLDF-вывода* $\bar{\mathrm{Q}}^0 \in \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog})$ будем называть

$$\nu\left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right) \rightleftharpoons \begin{cases} \divideontimes, & \textit{если } \circledast \\ 0, & \textit{иначе} \end{cases}$$

Лемма 2. *По любому* $\bar{\mathrm{Q}}^0 \in \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog})$ *строится* $\bar{\mathrm{U}}^0 \in \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog})$:

i. $\nu\left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right) \leqslant \nu\left(\bar{\mathrm{U}}^0;\mathtt{Prog}\right)$;

ii. *закреплённые за* $\bar{\mathrm{Q}}^0$ *правила и факты участвуют в процессе вывода* $\bar{\mathrm{U}}^0$ *не более раза каждое, прочие элементы множества* $\mathtt{Rule}^0_{\mathrm{L}}[\mathtt{Prog}] \cup \mathtt{Fact}^0_{\mathrm{L}}[\mathtt{Prog}]$ *— не используются.*[15]

---

[13] Обращаем внимание, именно исходного Q; правда, тот же вывод подойдёт и какому угодно $\tilde{\mathrm{Q}} \in \mathtt{Query}_{\mathrm{L}}$ при выполнении $\tilde{\mathrm{Q}}\tilde{\theta} = \mathrm{Q}^0_0$, где $\tilde{\theta} \in \Theta^0$.

[14] Мы ещё затронем эту тему в тексте §§4–5.

[15] В соответствующих $\bar{\mathrm{U}}^0$ последовательностях правил и фактов нет повторений.

Доказательство. Пусть отображение

$$\kappa : \left(\mathtt{Rule}^0_{\mathrm{L}}\left[\mathtt{Prog}\right] \cup \mathtt{Fact}^0_{\mathrm{L}}\left[\mathtt{Prog}\right]\right) \times \mathtt{SLDF}^0_{\mathrm{L}}\left(\mathrm{Q};\mathtt{Prog}\right) \mapsto \mathbb{N}$$

выдаёт значение кратности правила либо факта в последовательности, закреплённой за нормализованным выводом. Положим, нам удалось переделать $\bar{\mathrm{Q}}^0$ в некий нормализованный вывод $\bar{\mathrm{R}}^0$ (которому, в свою очередь, сопоставлены $\mathrm{D}^0_0, ..., \mathrm{D}^0_{s-1} \in \mathtt{Rule}^0_{\mathrm{L}}\left[\mathtt{Prog}\right]$ и $\mathrm{D}^0_s, ..., \mathrm{D}^0_{m-1} \in \mathtt{Fact}^0_{\mathrm{L}}\left[\mathtt{Prog}\right]$) с учётом неравенства

$$\kappa\left(\mathrm{D}^0_j; \bar{\mathrm{R}}^0\right) \leqslant \kappa\left(\mathrm{D}^0_j; \bar{\mathrm{Q}}^0\right),\ j = \overline{1, m-1}.$$

Выходит: $m \leqslant n$, $s \leqslant r$; $\left\{\mathrm{D}^0_s, ..., \mathrm{D}^0_{m-1}\right\} \subseteq \left\{\mathrm{C}^0_r, ..., \mathrm{C}^0_{n-1}\right\}$ и $\left\{\mathrm{D}^0_0, ..., \mathrm{D}^0_{s-1}\right\} \subseteq \subseteq \left\{\mathrm{C}^0_0, ..., \mathrm{C}^0_{r-1}\right\}$. Первое из включений даёт $\mu\left(\mathrm{D}^0_s \wedge ... \wedge \mathrm{D}^0_m\right) \geqslant \mu\left(\mathrm{C}^0_r \wedge ... \wedge \mathrm{C}^0_n\right)$, ибо $\mathrm{D}^0_s \wedge ... \wedge \mathrm{D}^0_m$ – обеднение $\mathrm{C}^0_r \wedge ... \wedge \mathrm{C}^0_n$ (см. п.4 следствия определения μ). Второе же гарантирует

$$\sum_{i=1}^{r-1}\left(1 - \mu\left(\mathrm{C}^0_i\right)\right)\mu\left(\mathrm{Body}\left(\mathrm{C}^0_i\right)\right) \geqslant \sum_{j=1}^{s-1}\left(1 - \mu\left(\mathrm{D}^0_j\right)\right)\mu\left(\mathrm{Body}\left(\mathrm{D}^0_j\right)\right).$$

Итогом служит, что $\circledast$ не убывает при переходе от $\bar{\mathrm{Q}}^0$ к $\bar{\mathrm{R}}^0$, другими словами

$$\nu\left(\bar{\mathrm{Q}}^0; \mathtt{Prog}\right) \leqslant \nu\left(\bar{\mathrm{R}}^0; \mathtt{Prog}\right).$$

Стало быть, (ii) автоматически даст и (i).

В нормализованном SLDF-выводе $\bar{\mathrm{Q}}^0$ каждый невыделенный атом A^0_i запроса $\mathrm{Q}^0_i \in \mathtt{Query}^0_{\mathrm{L}}$ (не обязательно начального) проделывает свой собственный путь вплоть до превращения в подмножество фактов $\mathtt{Fact}^0_{\mathrm{L}}\left[\mathtt{Prog}\right]$. Вернее: среди $\mathrm{C}^0_i, ..., \mathrm{C}^0_{n-1}$ выбираем те и только те, что относились к A^0_i (чьи заголовки совпадают с A^0_i) или его потомкам (атомам, в процессе вывода порождённым из A^0_i при помощи $\mathrm{C}^0_i, ..., \mathrm{C}^0_{n-1}$). Сформированная родословная — индуцированное $\bar{\mathrm{Q}}^0$ разветвление с корнем A^0_i — само нормализованный SLDF-вывод запроса $\Leftarrow \mathrm{A}^0_i$ (будем говорить “по линии Q^0_i”).

Рассмотрим ситуацию, когда $\mathrm{C}^0_i = \left(\mathrm{A}^0_i \Leftarrow \mathrm{Body}\left(\mathrm{C}^0_i\right)\right) \in \mathtt{Rule}^0_{\mathrm{L}}\left[\mathtt{Prog}\right]$ применялось дважды: к отличным друг от друга запросам Q^0_i и Q^0_j, $i < j$; т.о. $\mathrm{C}^0_i = \mathrm{C}^0_j$. Тогда оба раза дело касалось одного и того же атома $\mathrm{A}^0_i \in \mathbf{A}_{\mathrm{L}}$, совпадающего с A^0_j и в единственном числе присутствующего как в Q^0_i, так и в Q^0_j. Родословную A^0_i по Q^0_j мы вынуждены (в общем случае) оставлять, а вот побеги по линии Q^0_i, не лежащие на ветвях от Q^0_j, лишь мешают нам раньше завершить нормализованный вывод $\bar{\mathrm{Q}}^0$, ведь атому $\mathrm{A}^0_i = \mathrm{A}^0_j$ волей-неволей суждено появиться вновь. Следовательно, “пропуская” воздействие-разветвление A^0_i средствами C^0_i на этапе Q^0_i и придерживаясь прежней стратегии в остальном, мы попадаем в ранее разобранную ситуацию преобразования нормализованного вывода. Здесь остановимся подробнее на изложении алгоритма, перестраивающего $\bar{\mathrm{Q}}^0$.

Итак, требуется построить новую последовательность запросов и правил. Вначале наши действия повторяются, потому до $(i+1)$-ого шага определяем $\mathrm{R}^0_s \rightleftharpoons \mathrm{Q}^0_s$ и $\mathrm{D}^0_s \rightleftharpoons \mathrm{C}^0_s$, где $s = 1, ..., i$. Далее не трогаем A^0_i до поры C^0_j, точнее: пусть нами обозревается очередное (после C^0_i) правило/факт C^0_k для $k \geqslant i+1$; если заголовок C^0_k присутствует в текущем запросе (исключая случай A^0_i при $k < j$), то счётчик s наращивается на единицу и осуществляется шаг вывода с помощью $\mathrm{D}^0_s \rightleftharpoons \mathrm{C}^0_k$ к запросу R^0_{s+1}; иначе C^0_k исчезает из закреплённого за выводом кортежа[16], а мы переходим к следующему правилу/факту последовательности $\left\{\mathrm{C}_k^0\right\}_{i=1}^{n-1}$. Отметим, правило C_j^0 при этом не могло потерять своей области применения, ибо мы об этом позаботились заранее: в нужный момент среди атомов запроса есть специально заготовленный (сохранённый со времени своего возникновения в Q_i^0) для данной цели $\mathrm{A}_j^0 = \mathrm{A}_i^0$. Нетрудно понять, результирующий $\bar{\mathrm{R}}^0$ окажется успешным и полностью удовлетворяющим описанию в первой половине доказательства; ему будет отвечать подмножество $\left\{\mathrm{D}_1^0, ..., \mathrm{D}_{m-1}^0\right\} \subseteq \left\{\mathrm{C}_1^0, ..., \mathrm{C}_{n-1}^0\right\}$, причём длинна соответствующей $\bar{\mathrm{R}}^0$ цепочки правил и фактов строго уменьшилась (не менее чем на единицу) по сравнению с исходным $\bar{\mathrm{Q}}^0$ $(m \leqslant n-1)$. По индукции получим нужный вывод. □

Определение. Предположим $\mathtt{SLDF}_{\mathrm{L}}^0\left(\mathrm{Q};\mathtt{Prog}\right) \neq \emptyset$. *Оценкой предсказания* для запроса Q и программы Prog называется

$$\eta\left(\mathrm{Q};\mathtt{Prog}\right) \rightleftharpoons \sup\left\{\nu\left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right) \mid \bar{\mathrm{Q}}^0 \in \mathtt{SLDF}_{\mathrm{L}}^0\left(\mathrm{Q};\mathtt{Prog}\right)\right\}.$$

Следствие. $\eta\left(\mathrm{Q}\theta';\mathtt{Prog}\right) \leqslant \eta\left(\mathrm{Q};\mathtt{Prog}\right)$ для $\theta' \in \Theta$. □

Выбор операции sup не регламентируется чисто интуитивными соображениями: авторы руководствуются тем, что максимизация оценки теснейшим образом сопряжена с непротиворечивостью / недвусмысленностью предсказаний — весьма важным аспектом, обсуждение которого, однако, выходит за рамки непосредственно данной статьи.

Определение. Если $\left|\left\{\mu\left(\phi\right) \mid \phi \in F\right\}\right| < \infty$, то вероятность μ на множестве формул $F \subseteq \mathbf{S}_{\mathrm{L}}^0$, замкнутом относительно логических операций, называется *финитной.*

Теорема 1. *Пусть* μ *— финитна, а SLDF-дерево вычислений запроса* Q *средствами* Prog *— успешно. Тогда найдётся вывод* $\bar{\mathrm{U}}^0 \in \mathtt{SLDF}_{\mathrm{L}}^0\left(\mathrm{Q};\mathtt{Prog}\right)$ *такой, что* $\eta\left(\mathrm{Q};\mathtt{Prog}\right) = \nu\left(\bar{\mathrm{U}}^0;\mathtt{Prog}\right)$.

Доказательство. Пусть $conj\left(\mathtt{Fact}_{\mathrm{L}}^0\left[\mathtt{Prog}\right]\right)$ суть множество всех конъюнкций атомов-фактов $\mathtt{Fact}_{\mathrm{L}}^0\left[\mathtt{Prog}\right]$. Слагаемые в формуле ⋇ из определения оценки нормализованного SLDF-вывода имеют вид $\delta\left(\mathrm{C};\mathrm{D}\right) \rightleftharpoons \frac{\mu(\mathrm{Body(C)})-\mu(\mathrm{A}\wedge\mathrm{Body(C)})}{\mu(\mathrm{D})}$,

$$\mathrm{C} = \left(\mathrm{A} \Leftarrow \mathrm{Body}\left(\mathrm{C}\right)\right) \in \mathtt{Rule}_{\mathrm{L}}^{0,\mu}\left[\mathtt{Prog}\right],\ \mathrm{D} \in conj\left(\mathtt{Fact}_{\mathrm{L}}^0\left[\mathtt{Prog}\right]\right),\ \mu\left(\mathrm{D}\right) \neq 0. \quad (\star)$$

Видно, что $\delta\left(\mathrm{C};\mathrm{D}\right) \geqslant 0$, а финитность μ обеспечит $\left|\left\{\delta\left(\mathrm{C},\mathrm{D}\right)|\star\right\}\right| < \infty$. Упорядочим положительные элементы $\left\{\delta\left(\mathrm{C},\mathrm{D}\right)|\star\right\}$ по старшинству: $0 < \delta_1 < ... < \delta_K$. Нулевой супремум заведомо достижим, рассмотрим $\sup\left\{\nu\left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right)\right\} \neq 0$. Ежели искомое $\bar{\mathrm{U}}^0 \in \mathtt{SLDF}_{\mathrm{L}}^0\left(\mathrm{Q};\mathtt{Prog}\right)$ отсутствует, то с необходимостью существует бесконечно много нормализованных SLDF-выводов со строго возрастающими оценками вида ⋇: $0 < \nu_1 < \nu_2 < \nu_3 < ...$. Количество нефиктивных (больше нуля) слагаемых под эгидами индексов «$\mathrm{r}\left(\nu_i\right) - 1$» в выражениях для ν_i также может быть сколь угодно велико (иначе супремум вычислялся бы перебором конечного числа комбинаций $\delta > 0$). Допустим, функция $\pi : \mathbb{N} \mapsto \mathbb{N}$ — предъявляет номер $\pi\left(j\right) \in \mathbb{N}$, что выражение для $\nu_{\pi(j)}$ (вида ⋇) содержит сумму из не менее чем $j \in \mathbb{N}$ весомых слагаемых; π всюду определена. Но при

[16] Неверно, что удаляются абсолютно все правила/факты, относящиеся к возникавшим до момента Q_j^0 потомкам A_i^0 по линии Q_i^0, так как некоторые из указанных потомков могли быть “склеены” с побегами от других атомов.

$N \rightleftharpoons \lceil 1/\delta_1 \rceil + 1 \in \mathbb{N}$ получим: $\nu_{\pi(N)} \leqslant 1 - \sum_{i=1}^{N} \delta_1 \leqslant 1 - N \cdot \delta_1 < 1 - 1 = 0$ — противоречие. $\square$

$\mathbf{T}_{\mathrm{L},\leqslant \mathrm{d}}$ — термы $\mathbf{T}_{\mathrm{L}}$, в записи которых встречается не более $\mathrm{d} \in \mathbb{N}$ вхождений функциональных символов Σ. Далее полагаем $\mathbf{A}_{\mathrm{L},\leqslant \mathrm{d}}$ — атомы на базе $\mathbf{T}_{\mathrm{L},\leqslant \mathrm{d}}$; $\mathbf{A}^0_{\mathrm{L},\leqslant \mathrm{d}} \rightleftharpoons \mathbf{A}_{\mathrm{L},\leqslant \mathrm{d}} \cap \mathbf{A}^0_{\mathrm{L}}$. Запрещая участие термов не из $\mathbf{T}_{\mathrm{L},\leqslant \mathrm{d}}$, получим ограниченной глубины подмножества: $\mathtt{Query}_{\mathrm{L},\leqslant \mathrm{d}} \subseteq \mathtt{Query}_{\mathrm{L}}$ и т.д.

$$\mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog}) \rightleftharpoons \left\{ \bar{\mathrm{Q}}^0 \in \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog}) \,|\, \mathrm{Q}^0_0, ..., \mathrm{Q}^0_n \in \mathtt{Query}^0_{\mathrm{L},\leqslant \mathrm{d}} \right\}$$

— для таких выводов пишем «$\nu \restriction_{\mathrm{d}}$», акцентируя внимание на сужении области определения функции (как нетрудно заметить, $\mathrm{Q}^0_0, ..., \mathrm{Q}^0_n \in \mathtt{Query}^0_{\mathrm{L},\leqslant \mathrm{d}}$ влечёт $\mathrm{C}^0_0, ..., \mathrm{C}^0_{r-1} \in \mathtt{Rule}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}]$ и $\mathrm{C}^0_r, ..., \mathrm{C}^0_{n-1} \in \mathtt{Fact}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}]$). В условии непустоты $\mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog})$ *оценкой предсказания* (*запроса* Q) *глубины* d называется

$$\eta \restriction_{\mathrm{d}} (\mathrm{Q};\mathtt{Prog}) \rightleftharpoons \sup \left\{ \nu \restriction_{\mathrm{d}} \left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right) | \bar{\mathrm{Q}}^0 \in \mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog}) \right\}.$$

Теорема 2. *Пусть* d*-ограничение SLDF-дерева вычислений* Q *программой* $\mathtt{Prog}$ *— успешно. Тогда оценка предсказания той же глубины достигается на конкретном выводе: существует* $\bar{\mathrm{U}}^0 \in \mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog})$*, для которого* $\nu \restriction_{\mathrm{d}} \left(\bar{\mathrm{U}}^0;\mathtt{Prog}\right) = \eta \restriction_{\mathrm{d}} (\mathrm{Q};\mathtt{Prog})$.

Доказательство. Можно рассуждать в духе теоремы 1, поскольку $\mathbf{A}^0_{\mathrm{L},\leqslant \mathrm{d}}$ — конечно (откуда $\left| \mathtt{Rule}^{0,\mu}_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}] /_{\leftrightarrow} \right| < \infty$ и $\left| conj \left(\mathtt{Fact}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}] \right) /_{\leftrightarrow} \right| < \infty$). Однако (вместо неявного решения) лучше прямо укажем алгоритм построения такого конечного подмножества $\mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog})$, что супремум оценок $\nu \restriction_{\mathrm{d}}$ по нему совпадёт с $\eta \restriction_{\mathrm{d}} (\mathrm{Q};\mathtt{Prog})$ (само $\mathrm{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathrm{Prog})$, возможно, и бесконечно). Согласно лемме 1, любой из выводов $\mathtt{SLDF}^0_{\mathrm{L},\leqslant \mathrm{d}}(\mathrm{Q};\mathtt{Prog})$ укорачивается до менее $N \rightleftharpoons \left| \mathtt{Rule}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}] \cup \mathtt{Fact}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}] \right|$ шагов, притом без потери в величине $\nu \restriction_{\mathrm{d}}$. Значит, если за $N-1$ шаг нами не получен запрос-конъюнкция фактов $\mathtt{Fact}^0_{\mathrm{L},\leqslant \mathrm{d}}[\mathtt{Prog}]$ — цепочка более не представляет интерес; иными словами, достаточно перебрать все нормализованные $\mathtt{SLDF} \restriction_{\mathrm{d}}$-выводы длинны не более N запросов (а их конечное число). $\square$

Замечание. Изложенная нами при доказательстве процедура позволяет эффективно находить $\eta \restriction_{\mathrm{d}} (\mathrm{Q};\mathtt{Prog})$, а также $\mathtt{SLDF} \restriction_{\mathrm{d}}$-вывод, на котором эта оценка реализуется.

Определение. В условиях теоремы 1: *предсказание* запроса Q программой $\mathtt{Prog}$ — это нормализованный SLDF-вывод с оценкой $\eta(\mathrm{Q};\mathtt{Prog})$. Аналогично (исходя уже из теоремы 2) вводится понятие *предсказания глубины* d.

4. SLDp-вывод/-предсказание

Пусть в Σ нет функциональных символов (условия работы [9]), то есть $n_2 = 0$. Стоит отметить, тогда $\left|\mathbf{A}^0_{\mathrm{L}}\right| < \infty$; произвольная вероятность на $\mathbf{S}^0_{\mathrm{L}}$ — финитна; предсказание же есть предсказание некоторой глубины $\mathrm{d} \in \mathbb{N}$ (какой именно — неважно). Полагаем $r \rightleftharpoons \left| 2^{\mathbf{A}^0_{\mathrm{L}}} \right| < \infty$ и $\mathbb{W} \rightleftharpoons 2^{\mathbf{A}^0_{\mathrm{L}}} = \{\mathrm{K}_1, ..., \mathrm{K}_r\}$.

Определение [9, №14,17]. *Вероятностная ядерная интерпретация* — это такая функция $KI : \mathbb{W} \mapsto [0,1]$, что $\sum_{j=1}^{r} KI(K_j) = 1$. Имеющаяся KI может быть расширена до *вероятностной интерпретации*, которая (в свою очередь) являет собой отображение $I : \mathbf{S}^0_{\mathrm{L}} \mapsto [0,1]$, действующее следующим образом

$$I(\phi) = \sum_{K_j \in \mathbb{W} \text{ } and \text{ } K_j(\phi)=1} KI(K_j), \text{ где } \phi \in \mathbf{S}^0_{\mathrm{L}}.$$

Легко видеть, KI — просто (классическая) дискретная вероятность на $\mathbb{W}$; а I — соответствующая ей μ на $\mathbf{S}^0_{\mathrm{L}}$ (проверяется непосредственно, согласно определению из §2). Подобные вышеприведённому понятия фигурируют в работах [10, 11, 14, 15, 18, 23] (иногда употребляется термин *“степени доверия”*).

Так называемые *р-правила/-факты/-запросы* (см. [9]) имеют отличительную особенность: их составляющим (атомам — в данном контексте; а при обобщении понятий правила/факта/запроса речь идёт о конъюнкциях/дизъюнкциях атомов или даже бескванторных формулах) приписаны *аннотации* — подотрезки типа $[\alpha, \beta] \subseteq [0,1]$, чьё семантическое назначение описано чуть ниже. Конечная совокупность р-правил и р-фактов образует уже *р-программу* (ещё говорят *“вероятностную программу”*); для последней нами используется обозначение $\ulcorner$`Prog`$\urcorner$.[17] Программа `Prog` получается из $\ulcorner$`Prog`$\urcorner$ при отказе от аннотаций.

Определение [9, №18]. Пусть I — вероятностная интерпретация (расширение ядерной KI); $\phi_0, ..., \phi_m$ — основные предложения; $\psi_1, ..., \psi_m$ — бескванторные формулы; $\xi_0, ..., \xi_m$ — аннотации. Отношение $\vDash$ даётся индукцией по строению интересующего нас материала:

1. $I \vDash \phi_0 : \xi_0$ $\underline{\text{ттт}}$[18] $I(\phi_0) \in \xi_0$;
2. $I \vDash \phi_1 : \xi_1 \wedge ... \wedge \phi_m : \xi_m$ $\underline{\text{ттт}}$ $I(\phi_i) \in \xi_i$ для $i = \overline{1,m}$;
3. $I \vDash \phi_0 : \xi_0 \Leftarrow \phi_1 : \xi_1 \wedge ... \wedge \phi_m : \xi_m$ $\underline{\text{ттт}}$ $I \vDash \phi_0 : \xi_0$ или $I \nvDash \phi_1 : \xi_1 \wedge ... \wedge \phi_m : \xi_m$;
4. $I \vDash \exists \underbrace{x_1, ..., x_n}_{\overline{x}} (\psi_1(x_1, ..., x_n) : \xi_1 \wedge ... \wedge \psi_m(x_1, ..., x_n) : \xi_m)$ $\underline{\text{ттт}}$

$I \vDash \psi_1(\overline{x}/\overline{t}) : \xi_1 \wedge ... \wedge \psi_m(\overline{x}/\overline{t}) : \xi_m$ для какого-то набора термов $\overline{t} = (t_1, ..., t_n)$, где $t_j \in \Theta^0$ $(1 \leqslant j \leqslant n)$;

5. $I \vDash \forall \overline{x} (\psi_1(\overline{x}) : \xi_1 \wedge ... \wedge \psi_m(\overline{x}) : \xi_m)$ $\underline{\text{ттт}}$ $I \vDash \psi_1(\overline{x}/\overline{t}) : \xi_1 \wedge ... \wedge \psi_m(\overline{x}/\overline{t}) : \xi_m$ для всякого набора основных термов $\overline{t}$.

Замечание. Мы забываем о знаке «$\Leftarrow$» в р-факте как формуле (т.о. допускается $m = 0$ в пункте 3 определения). Что до р-запросов, то на них разумно навесить блок кванторов существования по всем переменным: $(\Leftarrow \psi_1 : \xi_1 \wedge ... \wedge \psi_m : \xi_m)$ преобразуется в $\exists \overline{x} (\psi_1 : \xi_1 \wedge ... \wedge \psi_m : \xi_m)$. Выражение $I \vDash \ulcorner\phi\urcorner$ (где $\ulcorner\phi\urcorner$ — аннотированное предложение) по традиции читается *“I удовлетворяет $\ulcorner\phi\urcorner$”*.

Разные по природе источников и происхождению статистики могут провоцировать появление различных аннотаций одного и того же атома, участвующего в р-правилах/-фактах $\ulcorner$`Prog`$\urcorner$. С другой стороны, устройство самой вероятности служит поводом для зависимости между элементами $\mathbf{S}^0_{\mathrm{L}}$. Общая аннотация р-программы отнюдь не исключает возникновения ситуации отсутствия каких-либо семантических релизаций (то есть удовлетворяющих программу интерпретаций). Итак, $\ulcorner$`Prog`$\urcorner$ представляет наше знание (в том числе о возможной

[17] В работе [9] должным образом не разделяются структуры р-факта и р-правила; добиваясь согласованности, мы здесь восполняем пробел самостоятельно. Для удобства, на протяжении параграфа запись “C = A $\Leftarrow$ Body” означает возможность и пустой посылки.

[18] Сокр. “тогда и только тогда”.

вероятностной мере на $\mathbf{S}^0_{\mathrm{L}}$, конкретные значения которой заранее не известны), а её непротиворечивость (см. определения далее) необходима для корректности подхода, рассматриваемого в [9].

Определение [9]. *Вероятностная модель p-программы* $\ulcorner\mathtt{Prog}\urcorner$ — интерпретация I, удовлетворяющая любому основному экземпляру (результату подстановки основных термов) p-правила и p-факта в $\ulcorner\mathtt{Prog}\urcorner$.

Определение [9, №19]. $\ulcorner\mathtt{Prog}\urcorner$ называется *противоречивой* в том и только том случае, когда для неё не существует вероятностной модели.

Считается, что объективно существующая (но изначально нам не предъявленная) реальная вероятность на множестве $\mathbf{S}^0_{\mathrm{L}}$ совпадает с одной из вероятностных моделей данной p-программы $\ulcorner\mathtt{Prog}\urcorner$. Здесь нет смысла вдаваться в определение fix-point семантики (согласно ей в [9] устанавливается критерий непротиворечивости p-программ); отметим, однако, способ отыскания непустого множества вероятностных моделей $\ulcorner\mathtt{Prog}\urcorner$ таким методом сталкивается с проблемой подчас неподъёмной вычислительной сложности (попытки снизить её обсуждаются в [17]). На деле (среди работ по синтезу логики и вероятности) нередко звучит постановка задачи о нахождении всех допустимых вероятностных моделей p-программы; нам для осуществления предсказания хватит лишь пп.(*а-б*) описания §3 для каждой вероятностной модели $\ulcorner\mathtt{Prog}\urcorner$, в коей нам бы хотелось осуществлять предсказание (например, это могут быть некие "экстремальные" модели той или иной ситуации). Пишем $\ulcorner\mathtt{Prog}\urcorner \models \ulcorner\phi\urcorner$ если и только если $I \models \ulcorner\phi\urcorner$ сразу для всех $I \in \mathrm{p}\mathbf{M}\left(\ulcorner\mathtt{Prog}\urcorner\right)$, где $\mathrm{p}\mathbf{M}\left(\ulcorner\mathtt{Prog}\urcorner\right)$ — множество вероятностных моделей $\ulcorner\mathtt{Prog}\urcorner$.

Определение [9, №25]. Пара различных p-правил/-фактов (или p-правила с p-фактом) $\mathrm{C}_1 = \mathrm{A}_1 : \xi_1 \Leftarrow \mathrm{Body}_1$ и $\mathrm{C}_2 = \mathrm{A}_2 : \xi_2 \Leftarrow \mathrm{Body}_2$ таких, что заголовки A_1 и A_2 унифицируемы (и, как следствие, существует наиболее общий унификатор θ), порождает новое p-правило $\mathrm{R}_{\mathrm{C}_1,\mathrm{C}_2} = \mathrm{A}_1\theta : (\xi_1 \cap \xi_2) \Leftarrow \mathrm{Body}_1 \wedge \mathrm{Body}_2$. *Замыкание p-программы* $\ulcorner\mathtt{Prog}\urcorner$, символически записываемое $\mathrm{CL}\left(\ulcorner\mathtt{Prog}\urcorner\right)$ — p-программа, полученная повторяющимся добавлением к $\ulcorner\mathtt{Prog}\urcorner$ всех возможных $\mathrm{R}_{\mathrm{C}_1,\mathrm{C}_2}$ для $\{\mathrm{C}_1, \mathrm{C}_2\} \subseteq \ulcorner\mathtt{Prog}\urcorner$.

Замечание. Замыкание не зависит от того, будем ли мы брать пары правил исходной $\ulcorner\mathtt{Prog}\urcorner$ или же пары из пополнений p-программы, образуемых пошагово в результате добавления какого-либо $\mathrm{R}_{\mathrm{C}_1,\mathrm{C}_2}$.

Определение [9, №7]. Пусть $[\alpha_1, \beta_1], [\alpha_2, \beta_2] \subseteq [0, 1]$.
1. $[\alpha_1, \beta_1] \otimes [\alpha_2, \beta_2] \rightleftharpoons [\max\{0, \alpha_1 + \alpha_2 - 1\}, \min\{\beta_1, \beta_2\}]$;
2. $[\alpha_1, \beta_1] \oplus [\alpha_2, \beta_2] \rightleftharpoons [\max\{\alpha_1, \alpha_2\}, \min\{1, \beta_1 + \beta_2\}]$.

Из комментария к определению вероятностной интерпретации становится прозрачна суть операций: они — прямое переложение известных двойных неравенств для пересечения и объединения событий с произвольной (классической) вероятностной мерой. Непосредственно $\mathsf{P}(\mathrm{A} \cup \mathrm{B})$ (ровно как $\mathsf{P}(\mathrm{A} \cap \mathrm{B})$) не выражается через $\mathsf{P}(\mathrm{A})$ и $\mathsf{P}(\mathrm{B})$. Тем не менее, границы достижимы на конкретных мерах (в чём нетрудно убедиться); таким образом, уже преобразованные интервалы точны для семейства всех вероятностных интерпретаций в нашем $\mathfrak{L}$ (без добавочных условий). Заключаем: формально $\otimes$ и $\oplus$ используют только самую общую информацию (об устройстве вероятностной интерпретации), но игнорируют сведение о том, что пригодна лишь часть из них (а точнее —

p$\mathbf{M}$ ($\ulcorner\mathtt{Prog}\urcorner$)); как результат, заложенное p-программой знание ими в некой степени размывается.

Определение [9, №26]. Выберем варианты p-правил/-фактов $\ulcorner\mathtt{Prog}\urcorner$ так, чтобы переменные для каждого были свои и не пересекались; $m = |\ulcorner\mathtt{Prog}\urcorner|$. Обозначим NF ($\ulcorner\mathtt{Prog}\urcorner$) — *нормальную форму p-программы* $\ulcorner\mathtt{Prog}\urcorner$, построенную согласно схеме[19]:

1. $\mathrm{CF}_1(\ulcorner\mathtt{Prog}\urcorner) = \mathrm{DF}_1(\ulcorner\mathtt{Prog}\urcorner) \rightleftharpoons \mathrm{CL}(\mathrm{REDUN}(\ulcorner\mathtt{Prog}\urcorner))$,

где $\mathrm{REDUNN}(\ulcorner\mathtt{Prog}\urcorner) \rightleftharpoons (\ulcorner\mathtt{Prog}\urcorner) \cup \big\{(\mathrm{A} : [0,1] \Leftarrow) \,|\, \mathrm{A} \in \mathbf{A}^0_{\mathrm{L}}\big\}$;

2. для всех $2 \leqslant i \leqslant m$ полагаем

$$\mathrm{CF}_i(\ulcorner\mathtt{Prog}\urcorner) \rightleftharpoons \Big\{(\mathrm{A}_1 \wedge ... \wedge \mathrm{A}_i) : \xi \Leftarrow \textstyle\bigwedge_{j=1}^{i} \mathrm{Body}_j \mid \Delta \;\; \text{\textit{и}} \;\; \xi = \xi_1 \otimes ... \otimes \xi_{\mathrm{i}}\Big\},$$
$$\mathrm{DF}_i(\ulcorner\mathtt{Prog}\urcorner) \rightleftharpoons \Big\{(\mathrm{A}_1 \vee ... \vee \mathrm{A}_i) : \xi \Leftarrow \textstyle\bigwedge_{j=1}^{i} \mathrm{Body}_j \mid \Delta \;\; \text{\textit{и}} \;\; \xi = \xi_1 \oplus ... \oplus \xi_{\mathrm{i}}\Big\},$$

$$\text{где } \Delta = \begin{cases} \mathrm{A}_j : \xi_j \Leftarrow \mathrm{Body}_j \in \mathrm{CL}(\mathrm{REDUN}(\ulcorner\mathtt{Prog}\urcorner)) \\ \mathrm{A}_k \neq \mathrm{A}_s \;\; \text{\textit{как только}} \;\; k \neq s \\ 1 \leqslant j,k,s \leqslant i \end{cases}$$

3. наконец, $\mathrm{NF}(\ulcorner\mathtt{Prog}\urcorner) \rightleftharpoons \bigcup\limits_{i=1}^{m} (\mathrm{CF}_i(\ulcorner\mathtt{Prog}\urcorner) \cup \mathrm{DF}_i(\ulcorner\mathtt{Prog}\urcorner))$.

SLDp-вывод [9] внешне аналогичен SLDF-выводу, но с дополнительными оговорками по поводу участвующих аннотаций:

i. интервал-аннотация левой части применяемого p-правила должен содержаться в интервале-аннотации атома, избранного для унификации в текущем p-запросе;

ii. вместо самой p-программы используется её вышеописанное следствие NF ($\ulcorner\mathtt{Prog}\urcorner$); акцентируя внимание и дабы не возникало путаницы, говорим *“SLDp-вывод нормальной формой...”*;

iii. атом удаляется из запроса при унификации с p-фактом; эта операция эквивалентна подчёркиванию, ибо однажды выделенный атом далее не участвует в выводе запросов;

iv. успешный SLDp-вывод — вывод, оканчивающийся пустым запросом; так как для оценивания важно “помнить” список понадобившихся фактов, то будем придерживаться представления о выделении (с учётом (*iii*)).

Зафиксируем $\mu \in \mathrm{p}\mathbf{M}(\ulcorner\mathtt{Prog}\urcorner)$ — по ней мы желаем вести предсказание. Пункты (*а*–*б*) из §3 относительно $\mathtt{Rule}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}] \cup \mathtt{Fact}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}]$ предполагаем известными (предсказанию предшествует извлечение таковых). Собственно как и для родственного SLDF, разумно подвергнуть анализу возможности p-программы $\ulcorner\mathtt{Prog}^{\mu}\urcorner \rightleftharpoons \mathtt{Rule}^{0,\mu}_{\mathrm{L}}[\ulcorner\mathtt{Prog}\urcorner] \cup \mathtt{Fact}^{0,\mu}_{\mathrm{L}}[\ulcorner\mathtt{Prog}\urcorner]$. Знание о ненулевой вероятности посылок p-правил следует (по существу) “зашить” внутрь $\ulcorner\mathtt{Prog}^{\mu}\urcorner$: считаем, что аннотации, приписанные атомам условных частей p-правил из $\ulcorner\mathtt{Prog}^{\mu}\urcorner$, отделены от нуля; иными словами, существует $\varepsilon > 0$ и все левые границы указанных интервалов больше либо равны ему (очевидно, если вероятность какого-нибудь атома нулевая, то и конъюнкции, его включающей — тоже). Естественно, среди основных экземпляров аннотированного запроса $\ulcorner\mathrm{Q}\urcorner$

[19]Здесь разрешается брать конъюнкции и дизъюнкции атомов в качестве заголовков правил и фактов.

нам интересны те, которые имеют ненулевую вероятность появления; формально — потребуем, чтобы интервалы в аннотациях $\ulcorner\mathrm{Q}\urcorner$ не содержали ноль.[20] Итак, все последующие утверждения раздела проводятся в предположении, что нижние границы аннотаций строго больше нуля (отделены от него).

Лемма 3. *Допустим, для запроса $\ulcorner\mathrm{Q}\urcorner$ с аннотацией, отделённой от нуля, существует успешный SLDp-вывод нормальной формой p-программы $\ulcorner\mathtt{Prog}^{\mu}\urcorner$. Тогда отыщется и нормализованный SLDF-вывод запроса* Q *(с оценкой по μ) программой $\mathtt{Prog}^{\mu} \rightleftharpoons \mathtt{Rule}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}] \cup \mathtt{Fact}^{0,\mu}_{\mathrm{L}}[\mathtt{Prog}]$.*

Доказательство. Коль скоро в аннотации $\ulcorner\mathrm{Q}\urcorner$ левые границы всех интервалов строго больше нуля, сами по себе элементы $\{(\mathrm{A} : [0,1] \Leftarrow) \,|\, \mathrm{A} \in \mathbf{A}^0_{\mathrm{L}}\} \subseteq$ $\subseteq \mathrm{REDUN}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$ на первом шаге применены быть не могут (см. первое отличие SLDp от SLDF). На шагах последующих атомы выделяются, либо заменяются правыми частями p-правил $\mathrm{NF}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$. А из определения NF ясно: ежели аннотации условных частей p-правил $\ulcorner\mathtt{Prog}^{\mu}\urcorner$ отделены от нуля, то данное свойство присуще и p-правилам $\mathrm{NF}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$, ибо посылки для последних составляются из посылок $\ulcorner\mathtt{Prog}^{\mu}\urcorner$. Итак, в чистом виде $\{(\mathrm{A} : [0,1] \Leftarrow) \,|\, \mathrm{A} \in \mathbf{A}^0_{\mathrm{L}}\}$, ровно как и его замыкание, не используются (ведь $\otimes$ и $\oplus$ сохраняют ноль в нижней границе для интервалов типа $[0,\beta_1]$ и $[0,\beta_2]$). Значит, любое $\ulcorner\mathrm{C}_k\urcorner$ из числа $\mathrm{NF}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$, задействованное в нашем успешном SLDp-выводе, могло получиться исключительно при участии некоторого p-правила/-факта $\ulcorner\mathtt{Prog}^{\mu}\urcorner$ — чьим неаннотированным аналогом C из $\mathtt{Prog}^{\mu}$ (произвольного $\ulcorner\mathrm{C}\urcorner \in \ulcorner\mathtt{Prog}^{\mu}\urcorner$ среди внёсших лепту в создание $\ulcorner\mathrm{C}_k\urcorner$) мы правомерно заменяем соответствующий элемент $\mathrm{NF}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$ при переходе к нормализованному SLDF-выводу средствами $\mathtt{Prog}^{\mu} \subseteq \mathtt{Prog}$. В обоснованности легко (с учётом определения нормальной формы и замечаний к SLDp) убедиться, поскольку:

а. в построении замыканий заключения p-правил/-фактов (или: p-правила и p-факта) $\ulcorner\mathtt{Prog}^{\mu}\urcorner$ предварительно унифицируются (совпадают — для основных экземпляров);

б. унификация атома и дизъюнкции (конъюнкции) атомов влечёт унифицируемость того же атома с каждым из входящих в упомянутую дизъюнкцию (конъюнкцию);

в. посылка $\mathrm{C} \in \mathtt{Prog}^{\mu}$ всегда полностью содержится в посылке образованного с его участием p-правила/-факта $\mathrm{NF}(\ulcorner\mathtt{Prog}^{\mu}\urcorner)$; вывод упрощается, т.к. в получающихся запросах атомов становится ещё меньше.[21] $\square$

Теорема 3. *Пусть $\ulcorner\mathtt{Prog}^{\mu}\urcorner$ — непротиворечивая p-программа, $\ulcorner\mathrm{Q}\urcorner$ — запрос с отделённой от нуля аннотацией. Если $\ulcorner\mathtt{Prog}^{\mu}\urcorner \models \ulcorner\mathrm{Q}\urcorner$, то предсказание* Q *средствами* $\mathtt{Prog}$ *определено (для μ в качестве вероятностной меры на $\mathbf{S}^0_{\mathrm{L}}$).*

Доказательство. По теореме №8 [9] существует успешный SLDp-вывод $\ulcorner\mathrm{Q}\urcorner$ из нормальной формы p-программы $\ulcorner\mathtt{Prog}^{\mu}\urcorner$. Комбинируя с нашей леммой 3, заключаем $\mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog}) = \mathtt{SLDF}^0_{\mathrm{L}}(\mathrm{Q};\mathtt{Prog}^{\mu}) \neq \emptyset$. $\square$

[20] Отметим, выполнение $\ulcorner\mathtt{Prog}^{\mu}\urcorner \models \ulcorner\mathrm{Q}\urcorner$ влечёт $\ulcorner\mathtt{Prog}^{\mu}\urcorner \models \ulcorner\mathrm{Q}\theta\urcorner$ для некоторого $\theta \in \Theta^0$ (см. лемму №18 в [9]). Атомам запроса Q (при возникающей необходимости) припишем $[\varepsilon, 1]$; за ε позволительно взять «δ_1» из доказательства теоремы 1.

[21] Некоторые из правил/фактов в результате могут потерять свою область приложения — происходит удаление таковых из закреплённого за выводом кортежа (действуя по аналогии с доказательством леммы 2).

Замечание. В доказательстве леммы 3 нами описан способ эффективного и естественного образования конечной совокупности нормализованных SLDF-выводов Q программой $\mathtt{Prog}^{\mu}$ по некоему успешному SLDp-выводу $\ulcorner \mathrm{Q} \urcorner$ p-программой NF ($\ulcorner \mathtt{Prog}^{\mu} \urcorner$). Значит, если в запасе теперь конечное множество SLDp-выводов $\ulcorner \mathrm{Q} \urcorner$ — организуем объединение совокупностей SLDF-выводов, причём каждый вывод (при надобности) сокращается с привлечением второй леммы (см. алгоритм уже в её доказательстве) до, как минимум, не менее плодотворного. На базе означенного объединения мы можем дать *частичную оценку предсказания* (*частичное предсказание*) аннотированного запроса $\ulcorner \mathrm{Q} \urcorner$ p-программой $\ulcorner \mathtt{Prog} \urcorner$ для вероятностной меры $\mu \in \mathrm{p}\mathbf{M}(\ulcorner \mathtt{Prog} \urcorner)$. Отталкиваясь от всех успешных SLDp-выводов $\ulcorner \mathrm{Q} \urcorner$ средствами $\ulcorner \mathtt{Prog} \urcorner$, определим (*полное*) *предсказание* $\ulcorner \mathrm{Q} \urcorner$ *для* μ и *его оценку* (аналогично).

5. Семантические μ-вывод/-предсказание

С этого момента мы вновь возвращаемся к сигнатуре Σ при $n_2 \geqslant 0$, т.е. возможно наличие функциональных символов (как и ранее для параграфов 2–3). В данном разделе мы займёмся переопределением предсказания, отталкиваясь от альтернативных SLD-подходу парадигм *семантики* и *вычислимости.*

Определение. Пусть $\mathrm{C}_1 = \left(\mathrm{A}_0^1 \Leftarrow \mathrm{A}_1^1 \wedge ... \wedge \mathrm{A}_{m_1}^1\right)$, $\mathrm{C}_2 = \left(\mathrm{A}_0^2 \Leftarrow \mathrm{A}_1^2 \wedge ... \wedge \mathrm{A}_{m_2}^2\right)$ — правила (фиксируются варианты с непересекающимися переменными). Отношение $\mathrm{C}_1 \succ \mathrm{C}_2$ (*«быть более общим»*) имеет место тогда и только тогда, когда можно подобрать такую $\theta \in \Theta$, что $\mathrm{A}_0^1\theta = \mathrm{A}_0^2$ и $\left\{\mathrm{A}_1^1\theta, ..., \mathrm{A}_{m_1}^1\theta\right\} \subseteq \left\{\mathrm{A}_1^2, ..., \mathrm{A}_{m_2}^2\right\}$, причём $m_1 \leqslant m_2$ и C_1 не является вариантом C_2.

$$\mathtt{Rule}_{\mathrm{L}}^{\mu} \rightleftharpoons \left\{\mathrm{C} \in \mathtt{Rule}_{\mathrm{L}} | \mathrm{C}\theta \in \mathtt{Rule}_{\mathrm{L}}^{0,\mu} \text{ \textit{для какого-нибудь} } \theta \in \Theta^0\right\};$$

$$\underline{\mu}(\mathrm{C}) \rightleftharpoons \inf\left\{\mu(\mathrm{C}\theta) | \theta \in \Theta^0 \text{ \textit{и} } \mathrm{C}\theta \in \mathtt{Rule}_{\mathrm{L}}^{0,\mu}\right\}, \text{ где } \mathrm{C} \in \mathtt{Rule}_{\mathrm{L}}^{\mu}.^{22}$$

Определение. Отношение $\mathrm{C}_1 \sqsubset \mathrm{C}_2$ (*«вероятностная выводимость правила* C_2 *из правила* C_1*»*) для $\mathrm{C}_1, \mathrm{C}_2 \in \mathtt{Rule}_{\mathrm{L}}^{\mu}$ эквивалентно двум условиям: $\mathrm{C}_1 \succ \mathrm{C}_2$ вместе с $\underline{\mu}(\mathrm{C}_1) < \underline{\mu}(\mathrm{C}_2)$.

Примечание. Оба $\succ$ и $\sqsubset$ суть строгие частичные порядки (на $\mathtt{Rule}_{\mathrm{L}}$ и $\mathtt{Rule}_{\mathrm{L}}^{\mu}$).

Определение. Назовём *μ-законом* всякое $\mathrm{C} \in \mathtt{Rule}_{\mathrm{L}}^{\mu}$, для которого $\mathrm{C}' \succ \mathrm{C}$ влечёт $\mathrm{C}' \sqsubset \mathrm{C}$ при любом $\mathrm{C}' \in \mathtt{Rule}_{\mathrm{L}}^{\mu}$ (μ-закон нельзя обобщить, не уменьшив его условной вероятности); $\mathtt{GLaw}_{\mathrm{L}}^{\mu}$ — совокупность всех μ-законов.

Определение. *Семантическим μ-выводом* атома $\mathrm{A} \in \mathbf{A}_{\mathrm{L}}$ называется максимальная (или бесконечная) цепочка $\mathrm{C}_1 \sqsubset \mathrm{C}_2 \sqsubset ...$ правил $\mathrm{C}_i = (\mathrm{A}_i \Leftarrow \mathrm{Body}(\mathrm{C}_i))$ в $\mathtt{GLaw}_{\mathrm{L}}^{\mu}$ $(i = 1, 2, ...)$, где все атомы A_i унифицируемы с A. Семантический μ-вывод *минимален* в случае, когда мы не прибегаем к избыточным уточнениям: нельзя подобрать такого $\mathrm{C}_{i+1/2}$ из $\mathtt{GLaw}_{\mathrm{L}}^{\mu}$ (отличного от вариантов C_i и C_{i+1}), что окажется $\mathrm{C}_i \sqsubset \mathrm{C}_{i+1/2} \sqsubset \mathrm{C}_{i+1}$.

Помимо прочего, обновлённое понятие предсказания должно вбирать в себя преимущества семантического μ-вывода, вместе с тем учитывая специфику и роль тех основных фактов, что допускают непосредственную проверку в изучаемых нами моделях и сами не нуждаются в предсказании, хотя весьма полезны

[22] В условиях теорем 1–2 инфимум можно заменить минимумом.

при оценке достоверности остальных свойств. Множество таких фактов обозначим $\mathtt{Fact}_{\mathrm{o}} \subseteq \mathtt{Fact}_{\mathrm{L}}^{0,\mu}$. Для вычисления оценок и предсказывающих закономерностей (подобно знанию о программе в §3) нам понадобятся:

а. элементы $\mathtt{GLaw}_{\mathrm{L}}^{\mu}$ и соответствующие им условные вероятности;
б. элементы $\mathtt{Fact}_{\mathrm{o}}$ и вероятности их конъюнкций.

Лемма 4. *Допустим,* A *— произвольный атом,* $\mathrm{C}_1 \succ \mathrm{C}_2 \succ \ldots$ *— некоторая последовательность правил* $\mathrm{C}_i = \left(\mathrm{A}_0^i \Leftarrow \mathrm{A}_1^i \wedge \ldots \wedge \mathrm{A}_{n_i}^i\right)$*. Если какое-нибудь* C_j *обладает свойством*

«найдётся $\theta \in \Theta^0$: $\mathrm{A}_0^j\theta = \mathrm{A}\theta$, $\left\{\mathrm{A}_1^j\theta \Leftarrow, \ldots, \mathrm{A}_{n_j}^j\theta \Leftarrow\right\} \subseteq \mathtt{Fact}_{\mathrm{o}} \subseteq \mathtt{Fact}_{\mathrm{L}}^{0,\mu}$*»,*

то и каждому правилу-предшественнику с номером $k \leqslant j$ *в последовательности это свойство присуще.*

Доказательство. Индукционная гипотеза: пусть для C_k $(k \leqslant j)$ нужное нам свойство уже установлено. Проведём шаг индукции. $\mathrm{C}_{k-1} \succ \mathrm{C}_k$, поэтому для $\theta_{k-1} \in \Theta$ имеем $\mathrm{A}_0^{k-1}\theta_{k-1} = \mathrm{A}_0^k$ и $\left\{\mathrm{A}_1^{k-1}\theta_{k-1}, \ldots, \mathrm{A}_{n_{k-1}}^{k-1}\theta_{k-1}\right\} \subseteq \left\{\mathrm{A}_1^k, \ldots, \mathrm{A}_{n_k}^k\right\}$. По предположению отыщется θ — унификатор A_0^k и A ($\mathrm{A}_0^k\theta = \mathrm{A}\theta$) такой, чтобы $\left\{\mathrm{A}_1^k\theta \Leftarrow, \ldots, \mathrm{A}_{n_k}^k\theta \Leftarrow\right\} \subseteq \mathtt{Fact}_{\mathrm{o}}$. Следовательно, $\mathrm{A}_0^{k-1}\theta_{k-1}\theta = \mathrm{A}\theta$, а

$$\left\{\mathrm{A}_1^{k-1}\theta_{k-1}\theta \Leftarrow, \ldots, \mathrm{A}_{n_{k-1}}^{k-1}\theta_{k-1}\theta \Leftarrow\right\} \subseteq \left\{\mathrm{A}_1^k\theta \Leftarrow, \ldots, \mathrm{A}_{n_k}^k\theta \Leftarrow\right\} \subseteq \mathtt{Fact}_{\mathrm{o}}.$$

Не ограничивая общности, множества переменных $\mathrm{X}(\mathrm{C}_{k-1})$, $\mathrm{X}(\mathrm{C}_k)$ и $\mathrm{X}(\mathrm{A})$ попарно не пересекаются (разумеется, это верно для подходящих вариантов правил C_{k-1} и C_k): $\theta_{k-1}\theta$ действует на $\mathrm{X}(\mathrm{C}_{k-1})$, когда θ — на $\mathrm{X}(\mathrm{C}_k) \cup \mathrm{X}(\mathrm{A})$; вдобавок $\mathrm{X}(\mathrm{C}_{k-1}) \cap (\mathrm{X}(\mathrm{C}_k) \cup \mathrm{X}(\mathrm{A})) = \emptyset$. Определяя подстановку

$$\tilde{\theta}(x) = \begin{cases} \theta_{k-1}\theta(x), & \textit{если } x \in \mathrm{X}(\mathrm{C}_{k-1}) \\ \theta(x), & \textit{если } x \in \mathrm{X}(\mathrm{C}_k) \cup \mathrm{X}(\mathrm{A}) \end{cases},$$

получаем требуемые

равенство $\mathrm{A}_0^{k-1}\tilde{\theta} = \mathrm{A}\tilde{\theta}$ и включение $\left\{\mathrm{A}_1^{k-1}\tilde{\theta} \Leftarrow, \ldots, \mathrm{A}_{n_{k-1}}^{k-1}\tilde{\theta} \Leftarrow\right\} \subseteq \mathtt{Fact}_{\mathrm{o}}$. □

Замечание. Вследствие контрапозиции леммы 4: ежели правило с номером i не удовлетворяет указанному свойству — то же будет касаться и любого правила-последователя с номером $j \geqslant i$.

Определение. *Семантическое μ-предсказание* $\mathrm{A} \in \mathbf{A}_{\mathrm{L}}$ средствами $\mathtt{GPro}^{\mu} \rightleftharpoons \mathtt{GLaw}_{\mathrm{L}}^{\mu} \cup \mathtt{Fact}_{\mathrm{o}}$ есть начальный сегмент $\mathrm{C}_1 \sqsubset \mathrm{C}_2 \sqsubset \ldots \sqsubset \mathrm{C}_k$ минимального семантического μ-вывода атома A, обладающий характеристиками:

i. для $\mathrm{C}_k = (\mathrm{A}_k \Leftarrow \mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n) \in \mathtt{GLaw}_{\mathrm{L}}^{\mu}$ найдётся подстановка $\theta \in \Theta^0$ такая, что $\mathrm{A}_k\theta = \mathrm{A}\theta$, $\{\mathrm{B}_1\theta \Leftarrow, \ldots, \mathrm{B}_n\theta \Leftarrow\} \subseteq \mathtt{Fact}_{\mathrm{o}}$, $\mu((\mathrm{B}_1 \wedge \ldots \wedge \mathrm{B}_n)\theta) \neq 0$ и $\mu(\mathrm{A}_k\theta) < \underline{\mu}(\mathrm{C}_k)$;

ii. не существует $\mathrm{C}_{k+1} \in \mathtt{GLaw}_{\mathrm{L}}^{\mu}$, одновременно удовлетворяющего (i) и уточняющего C_k со строгим увеличением условной вероятности ($\mathrm{C}_k \sqsubset \mathrm{C}_{k+1}$);[23]

iii. на C_k достигается максимум $\underline{\mu}(\cdot)$ среди μ-законов, наделённых свойствами (i-ii); все эти μ-законы, очевидно, можно считать содержащимися в каких-то своих последовательностях семантических μ-выводов A;

[23]Никакую цепь с C_k нельзя продлить по $\sqsubset$ при сохранении (i).

iv. *результатом* семантического μ-предсказания будем называть θ (из пункта первого), а *оценкой* — величину $\gamma\,(\mathrm{A};\mathtt{GPro}^{\mu}) \rightleftharpoons \underline{\mu}\,(\mathrm{C}_k)$.

Замечание. Прежде (во втором параграфе) мы уже касались вопроса ограничения количества вхождений функциональных символов (в термы сигнатуры Σ) числом $\mathrm{d} \in \mathbb{N}$. Само собой, здесь также можно рассмотреть множество $\mathtt{Rule}^{\mu}_{\mathrm{L},\leqslant\mathrm{d}}$ и тому подобные. Итогом послужит определение *семантического μ-предсказания глубины* d *средствами* $\mathtt{GPro}^{\mu}_{\leqslant\mathrm{d}} \leftrightharpoons \mathtt{GLaw}^{\mu}_{\mathrm{L},\leqslant\mathrm{d}} \cup \mathtt{Fact}_{\mathrm{o}}$, где $\mathtt{Fact}_{\mathrm{o}} \subseteq$ $\subseteq \mathtt{Fact}^{0,\mu}_{\mathrm{L},\leqslant\mathrm{d}}$ (или *семантического μ-*d*-предсказания* желаемого запроса).

Следует отметить, роль базового набора $\mathtt{Fact}_{\mathrm{o}}$ в предсказании (см. пункт (i)) может выполнять некоторое актуальное для данной ситуации подмножество фактов, вовсе не обязательно совпадающее со всем $\mathtt{Fact}^{0,\mu}_{\mathrm{L}}$. В §6 мы также ответим на вопросы:

• каким образом строится мера μ (а значит и множество $\mathtt{GLaw}^{\mu}_{\mathrm{L}}$) по выборке из генеральной совокупности;

• в чём заключён смысл индуктивного синтеза закономерностей и обучения, рассмотрим механизм применения предсказания к значимым относительно μ наборам основных данных.

На практике процедура семантического μ-предсказания представляет из себя поиск по дереву вывода, упорядоченному согласно присущей правилам вообще и μ-законам (в частности) *двупараметричной иерархии*, а именно — по числу атомов в посылке s и глубине вхождений функциональных символов d; точнее: k-й уровень образован правилами с параметрами из $\{(s,d)\,|\,s+d=k\}$. Само собой, мощность каждого уровня конечна. Дерево ориентированно и размечено, начиная от предсказываемого атома (метка корня) и заканчивая максимально специфичными на данных закономерностями (листья), путь до которых лежит через последовательности μ-законов (с учётом равномерного, без излишних прыжков, движения вдоль упомянутой иерархии). Подробнее остановимся на описании указанного процесса. Итак, в силу леммы 4 поиск стоит начинать с максимально общих правил, поэтому в качестве отправной совокупности выступает 1-й уровень, а среди его элементов уже выбираются удовлетворяющие п.1 определения μ-предсказания. Обозначим полученное множество S. Далее просматриваем 2-ой уровень и те его μ-законы, которые вероятностно выводятся из правил S при сохранении п.1, добавляем к S. Аналогично поступаем со следующим уровнем. Отметим, что мы не пропустим нужного μ-закона. Действительно, если некоторый элемент i-ого уровня C удовлетворяет п.1 и не находится в отношении $\sqsubset$ с правилами текущего S (среди них обязательно есть более общие, ибо свойство принадлежности к $\mathtt{Rule}^{\mu}_{\mathrm{L}}$ наследуется по $\succ$ вверх), то возможна лишь ситуация: найдётся $\mathrm{C} \in S$, причём $\mathrm{C}' \succ \mathrm{C}$ и $\underline{\mu}\,(\mathrm{C}') \geqslant \underline{\mu}\,(\mathrm{C})$ — это противоречит условию $\mathrm{C} \in \mathtt{GLaw}^{\mu}_{\mathrm{L}}$.

Утверждение 1. *Правила $\mathtt{GLaw}^{\mu}_{\mathrm{L}}$, для которых выполнены оба пункта (i-ii) определения семантического μ-предсказания (какого-нибудь атома* A*), будут не сравнимы по $\succ$.*

Доказательство. Предположим, C_1, C_2 — нужные нам правила и (например) $\mathrm{C}_1 \succ \mathrm{C}_2$. Коль скоро $\mathrm{C}_2 \in \mathtt{GLaw}^{\mu}_{\mathrm{L}}$, то (по определению μ-закона) $\underline{\mu}\,(\mathrm{C}_1) < \underline{\mu}\,(\mathrm{C}_2)$. Совмещая, $\mathrm{C}_1 \sqsubset \mathrm{C}_2$ — противоречие с пунктом (ii) уже для C_1. $\square$

В последующих утверждениях для рассматриваемых нами программ всегда подразумевается верным включение $\mathtt{Fact}_{\mathrm{L}}^{0,\mu}[\mathtt{Prog}] \subseteq \mathtt{Fact}_{\mathrm{o}}$.

Теорема 4. *Пусть μ — финитна, а запрос $(\Leftarrow \mathrm{A})$ предсказывается некоторой программой* $\mathtt{Prog}$ *с оценкой $\eta((\Leftarrow \mathrm{A});\mathtt{Prog}^{\mu}) > \mu(\mathrm{A}\theta)$, где θ – соответствующая предсказанию подстановка. Тогда* A *семантически μ-предсказывается средствами $\mathtt{GPro}^{\mu}$ с оценкой $\gamma(\mathrm{A};\mathtt{GPro}^{\mu}) \geqslant \eta(\Leftarrow \mathrm{A};\mathtt{Prog})$.*

Доказательство. По теореме 1 найдётся нормализованный SLDF-вывод $\bar{\mathrm{Q}}^0$, для которого оценка $\nu\left(\bar{\mathrm{Q}}^0;\mathtt{Prog}\right) = \eta((\Leftarrow \mathrm{A});\mathtt{Prog})$. Ясно, что будут выполнены $\mu\left(\mathrm{Q}_0^0|\mathrm{Q}_n^0\right) \geqslant \eta((\Leftarrow \mathrm{A});\mathtt{Prog}) > \mu(\mathrm{A}\theta)$, $\mathrm{Q}_0^0 = (\Leftarrow \mathrm{A})\,\theta$ и $\mathrm{Q}_n^0 \in conj(\mathtt{Fact}_{\mathrm{o}})$. Выписанные соотношения практически равносильны пункту (i) для правила $\mathrm{C} \rightleftharpoons \left(\mathrm{Q}_0^0 \Leftarrow \mathrm{Q}_n^0\right)$ и атома A, но с единственной оговоркой: не хватает принадлежности $\mathtt{GLaw}^{\mu}$ (хотя $\mathrm{C} \in \mathtt{Rule}_{\mathrm{L}}^{0,\mu} \subseteq \mathtt{Rule}_{\mathrm{L}}^{\mu}$). Допустим, C можно обобщить до $\mathrm{C}' \in \mathtt{Rule}_{\mathrm{L}}^{\mu}$ без потери в условной вероятности, то есть $\underline{\mu}(\mathrm{C}) \leqslant \underline{\mu}(\mathrm{C}')$. Либо $\mathrm{C}' \in \mathtt{GLaw}^{\mu}$, либо повторяем рассуждение теперь для $\mathrm{C}' \in \mathtt{Rule}_{\mathrm{L}}^{\mu}$ (вместо C). И так далее. В конце концов (для фиксированного правила, разумеется, имеется только ограниченное число правил, более общих относительно строгого порядка $\succ$) получаем $\tilde{\mathrm{C}} \in \mathtt{GLaw}^{\mu}$. Как вытекает из доказательства леммы 4, (i) для $\tilde{\mathrm{C}}$ сохранится. Вместе с тем, семантическое μ-предсказание определено. Действительно, нами уже найдено $\tilde{\mathrm{C}}$, а потому множество минимальных семантических μ-выводов $\mathrm{A} \in \mathbf{A}_{\mathrm{L}}$, в которых встречаются удовлетворяющие пункту (i) правила, не пусто; максимум границы условных вероятностей (см. $(ii$-$iii)$) достигается, ибо μ принимает лишь конечное число значений (значит, границ условных вероятностей не может быть бесконечно много). Наконец, пусть C_k — правило, обладающее всеми характеристиками $(i$-$iii)$. Из пункта (iii) и принимая во внимание механизм получения $\tilde{\mathrm{C}}$ (неубывающая по $\underline{\mu}(\cdot)$ цепочка правил с началом C) вытекает:

$$\gamma(\mathrm{A};\mathtt{GPro}^{\mu}) \rightleftharpoons \underline{\mu}(\mathrm{C}_k) \geqslant \underline{\mu}(\tilde{\mathrm{C}}) \geqslant ... \geqslant \underline{\mu}(\mathrm{C}') \geqslant \underline{\mu}(\mathrm{C}) =$$

$$= \mu\left(\mathrm{Q}_0^0|\mathrm{Q}_n^0\right) \geqslant \eta((\Leftarrow \mathrm{A});\mathtt{Prog}). \ \square$$

Теорема 5. *Допустим, $(\Leftarrow \mathrm{A})$ d-предсказывается средствами некоторой* $\mathtt{Prog}$ *с оценкой $\eta((\Leftarrow \mathrm{A});\mathtt{Prog}^{\mu}) > \mu(\mathrm{A}\theta)$, где θ — соответствующая предсказанию подстановка. Тогда* A *семантически μ-d-предсказывается (глубина та же) средствами $\mathtt{GPro}^{\mu}$ с оценкой $\gamma\upharpoonright_{\mathrm{d}}(\mathrm{A};\mathtt{GPro}^{\mu}) \geqslant \eta\upharpoonright_{\mathrm{d}}(\Leftarrow \mathrm{A};\mathtt{Prog})$.* $\square$

Аналогичные результаты справедливы, как нетрудно убедиться, и в предсказании (возможно, частичном) p-запросов, основанном на родственном SLDp.

Следствие. В условиях теоремы 4 (5) имеем $\eta(\Leftarrow \mathrm{A};\mathtt{GPro}^{\mu}) \geqslant \eta(\Leftarrow \mathrm{A};\mathtt{Prog}^{\mu})$ (соответственно $\eta\upharpoonright_{\mathrm{d}}(\Leftarrow \mathrm{A};\mathtt{GPro}^{\mu}) \geqslant \eta\upharpoonright_{\mathrm{d}}(\Leftarrow \mathrm{A};\mathtt{Prog}^{\mu})$). $\square$

6. Применение предсказания

Пусть $\mathbb{U}$ — предметная область (*универсум*); $M \subset \mathbb{U}$ — представительная совокупность объектов, признанных достаточно исследованными как по индивидным свойствам, так и характеру поведения в рамках систем (*база данных*). Фундаментом для синтетической вероятностной модели $\mathfrak{M} = \langle M;\mu\rangle$ служит $\mathfrak{G} = \{\langle M;\mathcal{J}_k\rangle\}_{k\in\mathrm{K}}$ $(\mathrm{K} \neq \emptyset)$ — непустой класс логических моделей языка 1-ого порядка $\mathfrak{L}$ единого носителя M, $\mathcal{J}_k$ — соответствующие интерпретации Σ; можно сказать, что в $\mathfrak{G}$ находит отражение наше частичное знание относительно большего класса $\mathfrak{G}^*$. Интуитивно: модели из $\mathfrak{G}^*$ — это вероятные траектории, сценарии возникающих взаимодействий, планы развития событий. Каждому миру $\upsilon \in 2^{\mathbf{A}^0_{\mathrm{L}}}$ сопоставляется свой подкласс

$$\mathfrak{G}^*_\upsilon \rightleftharpoons \left\{\mathfrak{A} \in \mathfrak{G} \mid \forall\phi \in \mathbf{A}^0_{\mathrm{L}} \left(\mathfrak{A} \vDash \phi \Leftrightarrow \upsilon(\phi) = 1\right)\right\} \subseteq \mathfrak{G}^*.$$

С очевидностью заключаем: $\upsilon_1 \neq \upsilon_2$ влечёт $\mathfrak{G}^*_{\upsilon_1} \cap \mathfrak{G}^*_{\upsilon_2} = \emptyset$ (ведь истинность на модели определена однозначно); ясно также, что каждая $\mathfrak{A} \in \mathfrak{G}^*$ попадает в какой-то из подклассов, а $\bigcup_\upsilon \mathfrak{G}^*_\upsilon$ — разбиение $\mathfrak{G}^*$. Имея вероятностную меру P на $\mathfrak{G}^*$, определяем $\mu(\phi) \leftrightharpoons \mathsf{P}(\{\mathfrak{A} \mid \mathfrak{A} \vDash \phi\})$. Это описание μ укладывается в схему $\boxed{2}$ первого параграфа с учётом $\mathsf{P}'(\{\omega\}) \leftrightharpoons \mathsf{P}(\{\mathfrak{A} \mid \mathfrak{A} \in \mathfrak{G}^*_\omega\})$.

Всякая интерпретация $\mathcal{J}$ символов Σ порождает истинность на $\mathbf{S}^0_{\mathrm{L}}$. Своего рода обратное суждение (скорее относящееся к разряду математического фольклора) позволяет до известной степени отождествить манипуляции с $\mathbf{S}^0_{\mathrm{L}}$ и логику высказываний:

Утверждение 2. *Для произвольного* $\upsilon \in 2^{\mathbf{A}^0_{\mathrm{L}}}$ *найдётся модель* $\mathfrak{A}$ *сигнатуры* Σ *(мощностью не более счётной) такая, что*

$$\forall\phi \in \mathbf{S}^0_{\mathrm{L}} \left(\mathfrak{A} \vDash \phi \Leftrightarrow \upsilon(\phi) = 1\right).$$

Доказательство. Множеству $\Gamma = \left\{\mathrm{A} \in \mathbf{A}^0_{\mathrm{L}} \mid \upsilon(\mathrm{A}) = 1\right\}$ взаимно-однозначно соответствует (эрбранова) интерпретация $\mathcal{J}_\Sigma$:

1. значением произвольного $t \in \mathbf{T}^0_{\mathrm{L}}$ является он сам;

2. для n-местного предикатного символа $\mathrm{P} \in \Sigma$ и $t_1, ..., t_n \in \mathbf{T}^0_{\mathrm{L}}$ выполняется

$$\mathrm{P}(t_1, ..., t_n) = \textbf{и} \;\Leftrightarrow\; \mathrm{P}(t_1, ..., t_n) \in \Gamma$$

Теперь пару $\langle \mathbf{T}^0_{\mathrm{L}}; \mathcal{J}_\Sigma \rangle$ мы можем взять в качестве искомой модели Γ. □

Непосредственно осуществлению предсказания предшествует ряд этапов:

- построение вероятностной модели μ по репрезентативной выборке $\mathfrak{G}$;
- индуктивный синтез предсказывающих закономерностей, а вместе с тем и подбор статистически значимых данных для изучаемой системы — в соответствии с ними будут принимаются/отклоняются гипотезы.

Первый этап сопровождается применением аппарата математической статистики [33], в результате чего имеется распределение для μ, аппроксимирующее “настоящие” значения *эмпирической вероятности* μ^* с точностью до нужной величины $\varepsilon > 0$. Естественно считать μ^* действующей на основе расширенного класса моделей 1-ого порядка $\mathfrak{G}^* \supset \mathfrak{G}$ сигнатуры Σ (как и прежде, носители — подмножества $\mathbb{U}$); в информационном плане указанный класс уже полон, но нам не доступен. Качество *репрезентативности* выборки обеспечивает нашу способность судить о мере в целом, заменяя μ^* на μ с учётом погрешности ε. Далее ставится задача предсказания свойств системы $\mathfrak{B} = \langle S; \mathcal{J}_S \rangle \in \mathfrak{G}^*$.

Исходный материал (для семантического подхода):

i. $\mathtt{GLaw}^\mu_{\mathrm{L}} \subseteq \mathtt{Rule}^\mu_{\mathrm{L}}$ — актуальные правила, чьё использование допустимо в процессе предсказания при любых $\mathfrak{B}' \in \mathfrak{G}^*$;

ii. $\mathtt{Fact}_{\mathrm{o}} \subseteq \mathtt{Fact}^{0,\mu}_{\mathrm{L}}$ — те из формул (опуская $\Rightarrow$), что позволяют непосредственную проверку в реально возникающей $\mathfrak{B}' \in \mathfrak{G}^*$.

Примечание. В случае SLDF-структур и базирующегося на них предсказания наборы (при фиксации $\mathtt{Prog}$) следует заменить на $\mathtt{Rule}_{\mathrm{L}}^{0,\mu}[\mathtt{Prog}]$ и $\mathtt{Fact}_{\mathrm{L}}^{0,\mu}[\mathtt{Prog}]$ соответственно. А SLDp-описание сводится к SLDF — см. §4.

Пункт (*ii*) постулирует наличие отображения $\zeta_{\mathfrak{B}} : \mathtt{Fact}_{\mathrm{o}} \mapsto \{0,1\}$; обозначим

$$\mathtt{Data}\,(\mathfrak{B}) \rightleftharpoons \{\mathrm{C} \mid \mathrm{C} \in \mathtt{Fact}_{\mathrm{o}} \text{ и } \zeta_{\mathfrak{B}}\,(\mathrm{C}) = 1\}\,.$$

Иногда предсказание делается по $\mathtt{D}\,(\mathfrak{N}) \subseteq \left\{\mathrm{A} \Leftarrow \mid \mathrm{A} \in \mathbf{A}_{\mathrm{L}}^{0} \text{ и } \mathfrak{N} \vDash \mathrm{A}\right\}$ — данным какой-либо конкретной модели $\mathfrak{N}$ из $\mathfrak{G}^*$. Нетрудно заметить, когда $\mathtt{Data}\,(\mathfrak{B})$ совместно на значимом относительно меры μ подклассе $\mathfrak{G}^*$, т.е.

$$\mathfrak{G}' = \{\mathfrak{A} \in \mathfrak{G}^* \,|\, \mathfrak{A} \vDash \mathrm{A},\ \text{где}\ \ (\mathrm{A} \Leftarrow) \in \mathtt{Data}\,(\mathfrak{B})\} \ \text{ и } \mathsf{P}\,(\mathfrak{G}') > 0,$$

его разумно считать данными всякой $\mathfrak{A} \in \mathfrak{G}'$. Обратно: принимая $\mathsf{P}\,(\{\mathfrak{N}\}) \neq 0$, для произвольных данных модели $\mathfrak{N}$ получим

$$\mathsf{P}\,(\{\mathfrak{A} \in \mathfrak{G}^* \mid \mathfrak{A} \vDash \mathtt{Data}\,(\mathfrak{N})\}) \geqslant \mathsf{P}\,(\{\mathfrak{N}\}) > 0.$$

Т.о. обе трактовки дают эквивалентное понимание *статистической значимости* по μ, необходимой для проведения предсказания.

Пусть нас интересует выполнимость А на $\mathfrak{B}$. В предположениях теоремы 4 вычисляем семантическое μ-предсказание $\mathrm{C}_1 \sqsubset \mathrm{C}_2 \sqsubset \ldots \sqsubset \mathrm{C}_k$ атома А с оценкой $\gamma\,(\mathrm{A}; \mathtt{GPro}^{\mu}) \rightleftharpoons \underline{\mu}\,(\mathrm{C}_k)$; сопоставленная ему (*индуктивно выведенная*) *закономерность* суть $\mathrm{C}_k\theta$, где θ — результат предсказания. Симметрично для теоремы 1: получаем предсказание $\bar{\mathrm{U}}^0 \rightleftharpoons \left(\mathrm{U}_0^0, \ldots, \mathrm{U}_n^0\right)$ запроса $\mathrm{Q} \rightleftharpoons (\Leftarrow \mathrm{A})$ с оценкой $\eta \rightleftharpoons \eta\,(\mathrm{Q}; \mathtt{Prog})$; тогда $\left(\mathrm{U}_0^0 \Leftarrow \mathrm{U}_n^0\right)$ — искомая закономерность.[24] Если $\mathrm{Body}\,(\mathrm{C}_k\theta) \in \mathit{conj}\,(\mathtt{Data}\,(\mathfrak{B}))$ (либо $\mathrm{U}_n^0 \in \mathit{conj}\,(\mathtt{Data}\,(\mathfrak{B}))$ — для SLDF-подхода), то свойство А подтверждается нами в модели $\mathfrak{B}$ с оценкой γ (или η соответственно). Эта оценка может служить критерием принятия/отклонения гипотезы об истинности запроса: например, доверяем только прогнозам с оценкой не меньше $1 - \sigma$, здесь σ — *доверительная граница.* Рассуждения для SLDp и предсказаний глубины d (см. теоремы 5 и 2) проводятся по аналогии. Обратим внимание: в то время как организация в виде SLD накапливает размытость и нечёткость начального приближения ε, а значит способна наращивать ошибку в процессе вывода, семантическое μ-предсказание выдаёт нам ответ с той же степенью точности ε, с которой была задана вероятность μ.

7. Приложения на практике. Реляционный подход к интеллектуальному анализу данных

Предлагаемая вниманию концепция семантики и вычислимости (в частности, семантический μ-вывод как её инструмент) позволяет разработать достаточно общий подход к методам *интеллектуального анализа данных* (*Knowledge Discovery in Data Bases and Data Mining — KDD&DM*), чьей характерной чертой (будь то реализация в виде экспертных систем, распознавания образов, или какой-либо иной форме) является направленность на умение правильно предсказывать те или иные характеристики исследуемых объектов. Базирующееся на упомянутом выводе семантическое вероятностное предсказание (сокр. СВП) можно воспринимать как процесс обнаружения (индуктивного синтеза)

---

[24] Запись $\mathrm{U}_0^0 \Leftarrow \mathrm{U}_n^0$ подразумевает, что в самих запросах знак «$\Leftarrow$» отброшен; напомним, U_0^0 — результат подстановки основных термов в Q, а $\mathrm{U}_n^0 \in \mathit{conj}\ \mathtt{Fact}_{\mathrm{L}}^{0}[\mathtt{Prog}]$.

высоковероятных и максимально специфичных относительно доступной информации закономерностей, описывающих исходные данные в языке первого порядка с вероятностными оценками. Стоит отметить, к настоящему моменту практически лишь методы *вероятностного индуктивного логического программирования* (*Probabilistic Inductive Logic Programming — PILP*) способны организовать поиск закономерностей указанного сорта. Помимо прочего, в нашей работе установлено: СВП дает оценки как минимум не меньшие тех, что опираются в своём получении на (вероятностное) логическое программирование (см., например, [9]). Механизм СВП осуществлён в виде программной системы обнаружения знаний «`Discovery`», реализующей *реляционный подход* (*Relational Data Mining*) к интеллектуальному анализу данных — этот подход ранее был представлен в работах [1, 25, 26, 27, 28, 30, 31, 41]. Далее мы опишем основные черты реляционного подхода, его принципиальные отличия от прочих широко распространённых методов, возможности и контекст применения, а также коснёмся практических приложений.

В наше время имеется большое количество KDD&DM-методов и реализующих их программных систем. Коль скоро нас интересует природа информации, заключенной во входных данных (относительно предметной области), важной темой для обсуждения становится эмпирическое содержимое наличествующих данных. Анализ методов KDD&DM показывает [28], что любой из методов (явно или неявно) подразумевает заданными:

1. типы входных данных, с которыми работает рассматриваемый метод;
2. *онтологию* KDD&DM-метода (включая язык) для оперирования / интерпретации данных и результатов;
3. класс тестируемых (на данных) гипотез, выраженных в терминах языка и проверяемых средствами конкретного метода — его принято называть *пространством знаний*;
4. интерпретацию онтологии метода KDD&DM в онтологии предметной области. Это требование весьма важно и вместе с тем редко соблюдается в системах искусственного интеллекта. К примеру, дабы применять классификацию данных (записанных в форме числовых векторов) с помощью сфер-окрестностей, нам сперва необходимо проинтерпретировать сферы в онтологии предметной области; в противном случае мы столкнёмся с неспособностью интерпретировать/осознать полученные результаты классификации. Извлечённое средствами KDD&DM-метода знание суть множество подтверждённых гипотез, которые интерпретируемы в онтологии метода и предметной области (обсуждение вопроса может быть найдено в [28]).

Интерпретация результатов зависит от выбранной онтологии, определяющей наше видение "реального мира". Отнология и предметная область теснейшим образом связаны, хотя на один и тот же объект можно смотреть с разных позиций и областей. Итак, нельзя извлекать информацию из данных, не уделив должного внимания формированию онтологии. Между тем задание онтологий накладывает на методы KDD&DM следующие ограничения:

1. извлекаемая информация должна формулироваться посредством используемых отношений и операций, интерпретируемых как в онтологии самого метода, так и в онтологии предметной области; существующие

KDD&DM-методы обычно работают исключительно с определенным набором типов данных, задействуют заранее известные отношения и операции (соотвествующие конкретному методу), обладающие некими свойствами: возможности метода тем самым фактически фиксированы и допускают лишь незначительные вариации (скажем, в качестве вспомогательных параметров), т.е. реально задействована далеко не вся информация, содержащаяся в данных и выразимая в языке онтологии рассматриваемого метода;

2. методы KDD&DM часто нацелены на обнаружение только определенных типов гипотез в терминах извлеченной информации (примерами служат разнообразные классы нейронных сетей, решающих деревьев, аппроксимационных функций и т.п.); данное обстоятельство нередко существенно и не всегда обоснованно сужает область поиска решений.

Реляционный подход (и система «`Discovery`» [1, 25, 26, 27, 28, 41]) преодолевают ряд описанных выше трудностей за счёт:

1. расширения множества допустимых типов данных (объекты выступают в роли элементов носителей моделей первого порядка);
2. применения логики первого порядка для введения онтологии метода, интерпретируемой в онтологии предметной области; знание формулируется с помощью алгебраических систем (причём нетрудно перейти и к многосорным системам);
3. возможности привлекать дополнительное знание, выраженное в языке логики первого порядка;
4. способности извлекать различные классы гипотез, описываемые в терминах языка первого порядка.

Остановимся подробнее на некоторых из преимуществ реляционного подхода. Как уже было замечено, существующие KDD&DM-методы обычно не поддерживают такого режима исследования данных, когда тип обнаруживаемой закономерности может варьироваться достаточно произвольно. В отличии от них система «`Discovery`» позволяет работать с достаточно произвольными классами гипотез (теориями универсальных формул), выразимых в языке первого порядка. Кроме того, система «`Discovery`» имеет возможности обнаруживать гипотезы, формулируемые экспертом (например, финансистом) в терминах его предметной области (онтологии). Интерпретируемость получаемых в результате закономерностей крайне важна для наиболее ответственных задач (скажем, при финансовом прогнозировании). Действительно, если речь идет о крупном вложении капитала и у нас есть два прогноза об ожидаемой прибыли, полученные нейронными сетями и средствами реляционного подхода, то доверие будет к тому из них, который интерпретируем, т.е. понятен с точки зрения онтологии применяемых методов и предметной области: невозможно принимать ответственные решения, не понимая, каким образом они получены, а в такой ситуации нейронным сетям (подобно "черному ящику") трудно доверять, ведь речь идёт о переходе от реально наблюдаемых величин к числовым аналогам, и подобный переход отнюдь не всегда оказывается корректен (инвариантен относительно допустимых преобразований и т.д.). Реляционный подход и система «`Discovery`» призваны решать и другую проблему, ею является задача максимально полного извлечения знаний из данных — здесь следует обратить внимание на привлечение языка первого порядка, обладающего довольно богатыми выразительными возможностями, интерпретируемость онтологии метода в предметной области. Проведённые на практике эксперименты по сравнению «Discovery» с такими широко распространенными методами, как нейронные сети, решающие деревья, ассоциативные правила, FOIL-метод (First Order Inductive Logic), линейный дискриминантный анализ, показали, что программная система «Discovery» зачастую работает точнее указанных методов, причём нахождение множества μ-законов и выполняемый системой семантический вероятностный вывод осуществлялись достаточно эффективно для практически решаемых задач: при среднем объеме данных (десятки признаков, сотни объектов) требовалось от нескольких десятков минут до нескольких часов работы алгоритма на PC Pentium III, Pentium IV (более подробно см. [1, 25, 26, 27, 28, 29, 30, 31, 41]).

Список литературы

[1] Е.Е. Витяев. *Извлечение знаний из данных. Компьютерное познание. Модели когнитивных процессов.* Новосибирск, НГУ, 2006.

[2] E.E. Vityaev. *The logic of prediction.* Mathematical Logic in Asia, Proceedings of the 9th Asian Logic Conference (Novosibirsk, Russia), (S.S. Goncharov, R. Downey, H. Ono, editors), World Scientific, Singapore, 2006, 263–276.

[3] Е.Е. Витяев. *Синтез логики, вероятности и обучения в семантическом вероятностном выводе.* Сборник научных трудов IV Международной научно-практической конференции <Интегрированные модели и мягкие вычисления в искусственном интеллекте>, Коломна, **1** (Май 2007), 133–140.

[4] C.G. Hempel. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science.* Free Press, New York, 1966.

[5] C.G. Hempel. *Maximal Specificity and Lawlikeness in Probabilistic Explanation.* Philosophy of Science, **35**: 2 (1968), The University of Chicago Press, 116–133.

[6] J.W. Lloyd *Foundations of logic programming.* Springer-Verlag, 1987.

[7] S.S. Goncharov, Y.L. Ershov, D.I. Sviridenko. *Semantic programming.* Information processing, Proc. IFIP 10-th World Comput. Congress, Dublin, **10** (1986), 1093–1100.

[8] S.S. Goncharov. *Computability and computable models.* Mathematical Problems from Applied Logics. New Logics for the XXIst Century. II (D.M. Gabbay et al, editors), International Mathematical Series, Springer, New York, 2006, 99-214.

[9] R. Ng and V.S. Subrahmanian. *Probabilistic Logic Programming.* Information and Computation, **101**: 2 (1993), 150–201, .

[10] J.Y. Halpern. *An Analysis of First-Order Logics of Probability.* Artificial Intelligence, **46** (1990),311-–350.

[11] J.E. Fenstad. *The Structure of Probabilities Defined on First-Order Languages.* Studies in Inductive Logic and Probabilities, **2** (1980) (R.C. Jeffrey, editor), University of California Press, 251–262.

[12] L. De Raedt, K. Kersting. *Probabilistic Logic Learning.* ACM-SIGKDD Explorations, special issue on Multi-Relational Data Mining, **5**: 1, July 2003, 31–48.

[13] L. De Raedt and K. Kersting. *Probabilistic Inductive Logic Programming.* Invited paper in Proceedings of the 15th International Conference on Algorithmic Learning Theory (S. Ben-David, J. Case and A. Maruoka, editors), Padova, Italy, October 2004, 19–36.

[14] T. Lukasiewicz. *Probabilistic logic programming with conditional constraints.* ACM Transactions on Computational Logic (TOCL), **2**: 3 (2001), 264–312.

[15] G. Kern-Isberner and T. Lukasiewicz. *Combining Probabilistic Logic Programming with the Power of Maximum Entropy.* Artificial Intelligence, **157**: 1-2, August 2004, 139–202.

[16] R. Haenni. *Unifying Logical and Probabilistic Reasoning.* Progic'05, 2nd Workshop on Combining Probability and Logic, 2005.

[17] R. Haenni. *Probabilistic Logic and Probabilistic Networks.* Progic'07, 3rd Workshop on Combining Probability and Logic, 2007.

[18] M.A. Paskin. *Maximum Entropy Probabilistic Logic.* Technical Report UCB/CSD-01-1161, University of California, Berkeley, April 2002.
[19] S. Muggleton. *Stochastic logic programs.* In L. De Raedt, editor, Advances in Inductive Logic Programming. IOS Press, 1996.
[20] S. Muggleton. *Learning stochastic logic programs.* Electronic Transactions in Artificial Intelligence, **4**: 041, 2000.
[21] J. Vennekens and S. Verbaeten. *A General View on Probabilistic Logic Programming.* Proceedings 15th Belgian-Dutch Conference on Artificial Intelligence, 2003, 299–306.
[22] K. Kersting and L. De Raedt. *Bayesian Logic Programs.* Proceedings of the Work-in-Progress Track at the 10th International Conference on Inductive Logic Programming (J. Cussens and A. Frisch, editors), 2000, 138–155.
[23] D. Poole. *The Independent Choice Logic for modelling multiple agents under uncertainty.* Artificial Intelligence, **94**: 1-2 (1997), 7–56, .
[24] L.A. Zadeh. *Fuzzy Sets.* Information and Control, 8 (1965), 338–353.
[25] B.Y. Kovalerchuk, E.E. Vityaev. *Data Mining in Finance: Advances in Relational and Hybrid methods.* Kluwer Academic Publishers, 2000.
[26] E.E. Vityaev, B.Y. Kovalerchuk. *Empirical Theories Discovery based on the Measurement Theory.* Mind and Machine, **14**: 4 (2004), 551–573.
[27] B.Y. Kovalerchuk, E.E. Vityaev. *Symbolic Methodology for Numeric Data Mining.* Intelligent Data Analysis, special issue on "Philosophies and Methodologies for Knowledge Discovery and Intelligent Data Analysis" (K. Rennolls, E. Vityaev, editors), IOS Press, **12**: 2 (2008), 165–188.
[28] E.E. Vityaev, B.Y. Kovalerchuk. *Relational Methodology for Data Mining and Knowledge Discovery.* Intelligent Data Analysis, special issue on "Philosophies and Methodologies for Knowledge Discovery and Intelligent Data Analysis" (K. Rennolls, E. Vityaev, editors), IOS Press, **12**: 2 (2008), 189–210.
[29] Е.Е. Витяев, А.А. Москвитин. *Введение в теорию открытий. Программная система Discovery.* Вычислительные системы: Логические методы в информатике, Вып. 148, Новосибирск, 1993, 117–163.
[30] B.Y. Kovalerchuk, E.E. Vityaev, J.F. Ruiz. *Consistent and Complete Data and "Expert"Mining in Medicine.* Medical Data Mining and Knowledge Discovery, Springer, 2001, 238–280.
[31] E.E. Vityaev, B.Y. Kovalerchuk. *Data Mining For Financial Applications.* Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers (O. Maimon and L. Rokach, editors), Springer, 2005, 1203–1224.
[32] А.А. Боровков. *Теория вероятностей*, М.: Наука, 1986.
[33] А.А. Боровков. *Математическая статистика (Оценки параметров. Проверка гипотез)*, М.: Наука, 1984.
[34] Г.Дж. Кейслер, Ч.Ч. Чэн. *Теория моделей*, М.: Мир, 1977.
[35] Ю.Л. Ершов, Е.А. Палютин. *Математическая логика*, СПб.: Изд-во <Лань>, 2004.
[36] Н.Н. Непейвода. *Прикладная логика*, Новосибирск: НГУ, 2000.
[37] R. Goldblatt. *Logics of time and computation.* CSLI Lecture Notes **7**, 1992.
[38] A. Chagrov, M. Zakharyaschev. *Modal logic.* Oxford logic guides: 35 (D. Gabbay, A. Macintyre and D. Scott, editors), Clarendon press, Oxford, 1997.
[39] Д. Рутковская, М. Пилиньский, Л. Рутковский. *Нейронные сети, генетические алгоритмы и нечёткие системы*, М.: Изд-во <Горячая линия-Телеком>, 2006.
[40] А.Г. Кусраев, С.С. Кутателадзе *Введение в булевозначный анализ*, М.: Наука, 2005.
[41] Scientific Discovery website. *http://www.math.nsc.ru/AP/ScientificDiscovery*

Смердов Станислав Олегович
Новосибирский государственный университет,
ул. Пирогова 2,
630090, Новосибирск, Россия
E-mail address: netid@ya.ru

Витяев Евгений Евгеньевич
Институт математики им. С. Л. Соболева СО РАН,
пр. академика Коптюга 4,
630090, Новосибирск, Россия
E-mail address: evgenii.vityaev@math.nsc.ru