ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2021, № 1, с. 160–168

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 51.01+517.11

МЕТОД АБДУКТИВНОГО ВЫВОДА В ЗАДАЧАХ ОБЪЯСНЕНИЯ НАБЛЮДАЕМОГО

© 2021 г. С. Н. Васильев
ИПУ им. В.А. Трапезникова РАН, Москва, Россия
e-mail: vassilyev_sn@mail.ru

Поступила в редакцию 10.07.2020 г.
После доработки 12.07.2020 г.
Принята к публикации 28.09.2020 г.

В проблематику искусственного интеллекта, управления и принятия решений при неполной или недостоверной информации входит широкий класс задач абдуктивного объяснения наблюдаемого, включая задачи в терминах “причина-эффект”. Работа посвящена обоснованию метода логического формирования гипотез, объясняющих наблюдаемое. Предлагаются средства представления знаний и вывода гипотез. Вводится язык, обладающий свойством подстановочности. Свойства языка и вводимых в нем исчислений обеспечивают гипотезирование путем сочетания дедукции с абдукцией. В отличие от известных логических методов абдукции, предложенные средства позволяют выводить гипотезы (миноранты), необходимые и достаточные для формального объяснения наблюдаемого. На основе минорант в сочетании с базовой теорией предметной области формируются достоверные причины наблюдаемого или релевантные обстоятельства, выводящие на эти причины. При этом в ситуациях наличия также эмпирических данных эти причины и обстоятельства могут формироваться и в правдоподобных версиях. Рассматриваются примеры из техники и медицины.

DOI: 10.31857/S000233882101011X

Введение. В область искусственного интеллекта (ИИ), управления и принятия решений при неполной или недостоверной информации входят задачи так называемого объяснения на основе моделей и методов абдуктивных рассуждений (Explanation by abductive reasoning), в том числе в терминах “причина-эффект”. Здесь под “эффектом” или далее также под “наблюдаемым” могут пониматься аномалии, симптомы и вообще некоторые выделенные (особые) состояния системы как объекта анализа и управления.

Понятие абдукции введено логиком и философом Ч.С. Пирсом (C.S. Peirce, 1839–1914 гг.) в форме логических рассуждений, сравнимых, но отличных от дедукции и индукции. Сегодня абдукция чаще всего определяется следующим образом: на основе заданной теории T предметной области и утверждения G о наблюдаемом эффекте, подлежащем объяснению, найти такую не противоречащую теории T гипотезу-объяснение Δ, что из T и Δ выводимо G (модификации см., например, в [1–3]).

В работах по абдуктивному гипотезированию на теорию T предметной области часто накладываются ограничения, при которых дедукция максимально упрощается, например содержание теории должно иметь форму “if-and-only-if”, т.е. форму семейства эквивалентностей

$$\bigwedge_{i=1}^{m} (c_i \leftrightarrow E_{i1} \vee \ldots \vee E_{in}),$$

где в пропозициональном случае c_i – переменные-причины, а E_{ij}, $j = \overline{1,n}$, – элементарные конъюнкции, составленные из переменных-эффектов и их отрицаний, возможно, при дополнительных требованиях к множеству задействованных переменных, частично упорядоченному импликацией [1, 2].

В данной работе не предполагается какой-либо предварительной трансформации теории T, а в качестве языка представления знаний в разд. 1 вводится S-язык L_S как подмножество пропозиционального языка, но без ослабления его выразительной силы. Допуская использование отрицания переменных, язык L_S шире пропозиционального варианта языка позитивно-образованных формул [4, 5], но обладает тем же свойством подстановочности (см. разд. 1). Свойства языка L_S и вводимых в нем исчислений $\mathbb{C}_\alpha$, $\mathbb{C}_\beta$ обеспечивают гибкое сочетание дедукции с абдуктивным достоверным или правдоподобным выводом.

Вывод в $\mathbb{C}_\beta$ наблюдаемого эффекта G выполняется вместе с синтезом формулы P, называемой гипотезой-минорантой и объясняющей эффект G лишь формально. С точностью до эквивалентности гипотеза P логически минимальна (любая гипотеза-причина Δ, объясняющая G в смысле выше приведенного определения, должна быть не слабее $P: \Delta \to P$). Гипотеза-миноранта выступает контрольным условием и опорным средством формирования достоверных гипотез Δ, а при наличии сверхтеории T эмпирических знаний миноранта является также и средством вывода правдоподобных гипотез Δ. Эмпирические знания – это знания об априорном множестве потенциально возможных причин-кандидатов и/или о правдоподобных причинно-следственных связях. Из гипотезы P переходом к достаточным условиям, например путем упрощения P или использования эмпирических знаний, получаются либо причины Δ-эффекта G (достоверные или правдоподобные), либо поначалу лишь релевантные обстоятельства (факты, события, "улики"), из которых в рамках теории T средствами исчислений $\mathbb{C}_\alpha$ и $\mathbb{C}_\beta$ оказываются выводимыми и сами причины.

Заметим, что в проблематике интеллектуализации автономных (беспилотных) агентов с реактивными правилами "условие-действие" понятие "эффект" может иметь также смысл целевого состояния агента [6]. Эта актуальная тема абдуктивной информационной поддержки планирования действий агентов выходит за рамки данной статьи. Актуальной задачей остается также проблематика ранжирования альтернативных объяснений наблюдаемого для сужения множества гипотез до наиболее достоверных с использованием вероятностных, логических и других оценок.

В разд. 1 определяется язык представления знаний, а в разд. 2 – правила преобразования его формул. В разд. 3 вводятся исчисления $\mathbb{C}_\alpha$, $\mathbb{C}_\beta$ и обосновываются их свойства. В разд. 4 применение этих исчислений иллюстрируется примерами из техники и медицины.

1. Представление знаний. В пропозициональном языке L стандартной (классической) семантики выделим подмножество L_S. Ввиду свойства подстановочности формул языка L_S, которое будет рассмотрено ниже, назовем его *S-языком* (от английского *Substitution*). Далее конъюнктами называем элементарные конъюнкции литералов (пропозициональных переменных и/или их отрицаний) и пропозициональные константы *false*, *true*. Если в элементарную конъюнкцию (ЭК) входит пара контрарных литералов (переменная и ее отрицание), то по умолчанию такие ЭК заменяются на константу *false*. Логические связки $\wedge$, $\vee$, $\neg$, $\to$ в S-языке понимаются стандартно.

О п р е д е л е н и е 1 (*формулы языка* L_S и их типы):

1) ЭК A или константа *false*, – *S-формулы*; им приписывается *тип* $\wedge$; это простейшие S-формулы;

2) если F_i – S-формулы типа $\wedge$ $(i = \overline{1,m})$, а G – ЭК или *true*, то $G \to \left(\bigvee_{i=1}^{m} F_i\right)$ – *S-формулы типа* $\to$;

3) если F_i – S-формулы типа $\to$ $(i = \overrightarrow{1,m})$, а G – ЭК или *true*, то $G \wedge \left(\bigwedge_{i=1}^{m} F_i\right)$ – *S-формулы типа* $\wedge$;

4) других формул в языке L_S нет.

Если $m \neq 1$ и некоторое F_i совпадает с *false*, то возможно упрощение S-формулы очевидным (логически эквивалентным) преобразованием; такие упрощения далее будут подразумеваться.

Каноническим представлением S-формулы F называем ее запись в виде формулы с конъюнктом *true* в корневой вершине ее структуры:

$$F = true \to \bigvee_{i=1}^{m} \Omega_i, \quad \Omega_i = A_i \wedge \left(\bigwedge_{j=1}^{n_i} (B_{ij} \to \Phi_{ij})\right), \tag{1.1}$$

где $m \geq 1$, $n_i \geq 0$ (если $n_i = 0$, то $\Omega_i = A_i$).

Подформулы некоторой S-формулы F, образуемые в соответствии с индуктивным определением S-формул, называются *главными подформулами* (ГП). Так, в (1.1) при $n_i \neq 0$ подформулы A_i

и B_{ij} не являются главными, а при $n_i = 0$ подформула A_i – главная. Выражение $F_1 \subseteq F$ означает, что F_1 – ГП S-формулы F.

Если $H \subseteq F \in L_S$, $H = D \to \bigvee_{i=1}^{n} H_i$, то подформула $|H|$из H вида $|H| = \bigvee_{i=1}^{n} H_i$ не является главной. Через $|H|_{-i^*}$обозначается $\bigvee_{i=1, i\neq i^*}^{n} H_i$ (если H – типа $\wedge$, то смысл $|H|$ и $|H|_{-i^*}$ аналогичен с заменой $\vee$ на $\wedge$). Эти обозначения будут использованы в разд. 3 для упрощения записи формул.

Каждой вершине древовидной структуры S-формулы приписываем тип, соответствующий типу ГП, корнем которой эта вершина является. По определению 1 типы $\wedge$ и $\to$ вершин вдоль ветви структуры S-формулы F чередуются.

Формулы языка L, не принадлежащие подмножеству L_S, понимаем стандартно. Например, если $F_1, F_2 \in L_S$, то выражение $F_1 \to F_2$, вообще говоря не принадлежащее L_S, понимается как в L.

В отличие от формул языка L для S-формул F, F_1, F_2 справедливо свойство *подстановочности*: если $F_1 \subseteq F$ и $F_2 \to F_1$, то $F(F_2/F_1) \to F$, где $F(F_2/F_1)$ – результат подстановки в F вместо F_1 не менее сильной формулы F_2.

Язык L_S полон относительно выразительной силы языка L. Семантика S-формул очевидна.

2. Преобразования S-формул. Рассмотрим S-формулу F в каноническом представлении (1.1). Конъюнкт A_i называется *базой*, а конъюнкты B_{ij} (когда $n_i \neq 0$) – *вопросами к базе* A_i. Если $B_{ij} \subseteq A_i$, то вопрос B_{ij} к базе A_i называем *уместным*. Тождественно ложную S-формулу (*true* $\to$ *false*) далее для краткости обозначаем $\perp$.

Пусть в (1.1) $F \neq \perp$, $n_i \neq 0$ и Φ_{ij} имеют вид

$$\Phi_{ij} = \bigvee_{k=1}^{l_{ij}} (C_{ijk} \wedge \Psi_{ijk}), \quad \text{где} \quad l_{ij} \geq 1, \quad \text{или} \quad \Phi_{ij} = false. \tag{2.1}$$

Предположим, что

$$\exists i^* \in \overline{1,m}, \quad n_{i^*} \neq 0, \quad j^* \in \overline{1,n_{i^*}} \quad B_{i^*j^*} \subseteq A_{i^*}, \tag{2.2}$$

т.е. вопрос $B_{i^*j^*}$ к базе A_{i^*} уместен.

Определение 2 (α-*преобразование*). Пусть S-формула F имеет вид (1.1), (2.1), (2.2). Тогда, если $\Phi_{i^*j^*} \neq false$, то α-преобразованием называется такое отображение $\alpha : L_S \to L_S$, что

$$\begin{gathered}\alpha(F) = true \to \left(\bigvee_{i=1, i\neq i^*}^{m} \Omega_i\right) \vee \Omega_{i^*}^*, \quad \Omega_{i^*}^* = \bigvee_{k=1}^{l_{i^*j^*}} \Omega_{i^*j^*k}^*, \\ \Omega_{i^*j^*k}^* = \left[\{A_{i^*} \cup C_{i^*j^*k}\} \wedge \Psi_{i^*j^*k} \wedge \left(\bigwedge_{j=1, j\neq j^*}^{n_{i^*}} (B_{i^*j} \to \Phi_{i^*j})\right)\right].\end{gathered} \tag{2.3}$$

Если же $\Phi_{i^*j^*} = false$, то при $m > 1$

$$\alpha(F) = true \to \left(\bigvee_{i=1, i\neq i^*}^{m} \Omega_i\right),$$

а при $m = 1$ $\alpha(F) = \perp$.

В (2.3) и далее, когда это не вызывает коллизий, конъюнкты рассматриваются как множества входящих в эти конъюнкты элементов. Если $\Phi_{i^*j^*} \neq false$, то в случае появления в конъюнкте $\{A_{i^*} \cup C_{i^*j^*k}\}$ из (2.3) контрарных литералов применяем следующие очевидные упрощения формулы $\alpha(F)$: при $l_{i^*j^*} > 1$ удаляется $\Omega_{i^*j^*k}^*$; при $l_{i^*j^*} = 1$ и $m = 1$ $\alpha(F)$ заменяется на $\perp$; при $l_{i^*j^*} = 1$ и $m > 1$ удаляется $\Omega_{i^*j^*1}^*$. Удаляются и дубли ГП при совпадении $\Omega_{i^*}^*$ с Ω_i, $i \neq i^*$.

Пусть $\alpha(F) \neq \perp$ и выполняется условие типа (2.2). Применяя α-преобразование к $\alpha(F)$, получим S-формулу $\alpha^2(F)$ и т.д. до тех пор, пока очередной результат еще отличен от $\perp$ и выполняется (2.2). Пусть на некотором шаге r_1 $(r_1 \geq 0)$ $\alpha^{r_1}(F) \neq \perp$, но α-преобразование уже не применимо ввиду невыполнения (2.2), т.е. $\forall i \in \overline{1, m^1}$, где m^1 – количество баз в $\alpha^{r_1}(F)$, база A_i будет:

1) либо висячей ($n_i = 0$) и ввиду упомянутых упрощений отличной от *false*, либо

2) невисячей ($n_i \geq 1$) и такой, что $\forall j \in \overline{1, n_i}$ вопросы B_{ij} к базе A_i неуместны (под выражениями A_i, B_{ij} понимаются новые конъюнкты в формуле $\alpha^{r_1}(F)$).

Определение 3 (β-*преобразование*). Пусть I_1, I_2 – множества индексов $i \in \overline{1, m^1}$, для которых базы A_i удовлетворяют условиям 1) и 2) соответственно, и

$$\alpha^{r_1}(F) = true \rightarrow \bigvee_{i=1}^{m^1} \Omega_i^1,$$

где

$$\forall i \in I_1 \;\; \Omega_i^1 = A_i \quad \text{и} \quad \forall i \in I_2 \;\; \Omega_i^1 = A_i \wedge \left(\bigwedge_{j=1}^{n_i} (B_{ij} \rightarrow \Phi_{ij}^1) \right). \tag{2.4}$$

Назовем β-*преобразованием* такое отображение $\beta : L_S \rightarrow L_S$, что

$$\begin{gathered} \beta(\alpha^{r_1}(F)) = true \rightarrow \bigvee_{i=1}^{m^1} \Omega_i^*, \quad \forall i \in \overline{1, m^1} = I_1 \cup I_2 \quad \Omega_i^* = A_i \wedge P_i^1, \\ \forall i \in I_1 \quad P_i^1 = (A_i \rightarrow false), \quad \forall i \in I_2 \quad P_i^1 = \left(A_i \rightarrow \bigvee_{j=1}^{n_i} B_{ij} \right). \end{gathered} \tag{2.5}$$

Формирование условия

$$P^1 = P(\alpha^{r_1}(F)) = \bigwedge_{i=1}^{m^1} P_i^1$$

называем далее (α, β)-*гипотезированием*.

3. *S*-исчисления и их свойства. На основе α- и β-преобразований формул языка L_S, рассматриваемых в каноническом представлении, введем два исчисления.

Пусть $F^1 = \beta(\alpha^{r_1}(F))$. Применим к F^1 α-преобразование максимально возможное число раз r_2 $(r_2 \geq 1)$. Если $\alpha^{r_2}(F^1) \neq \perp$, то применим β-преобразование, получив $F^2 = \beta(\alpha^{r_2}(F^1))$ и т.д. Пусть $F^n = \beta(\alpha^{r_n}(F^{n-1}))$ и впервые $\alpha^{r_{n+1}}(F^n) = \perp$. В результате n-кратного (α, β)-гипотезирования получим условие

$$P = \bigwedge_{k=1}^{n} P^k, \quad P^k = P(\alpha^{r_k}(F^{k-1})),$$

именуемое *гипотезой-минорантой* (ГМ).

Определение 4 (*S-исчисления* $\mathbb{C}_\alpha$, $\mathbb{C}_\beta$). *S*-исчисления $\mathbb{C}_\alpha = (\perp, \{\alpha\})$ и $\mathbb{C}_\beta = (\perp, \{\alpha, \beta\})$ – это исчисления в языке L_S с аксиомой $\perp$ и указанными преобразованиями в качестве правил вывода. Здесь выводы формулы F – конечные последовательности формул, начинающиеся с формулы F, заканчивающиеся формулой $\perp$ и с промежуточными формулами, получаемыми в $\mathbb{C}_\alpha$ с помощью только α-преобразований, а в $\mathbb{C}_\beta$ – с помощью α- и β-преобразований как в построенной ранее для формулы F последовательности: $F^0, F^1, F^2, \ldots, F^{n+1}$, где $F^0 = F$. Вывод *S*-формулы F суть ее опровержение.

Теорема 1. *Формула F языка L_S противоречива тогда и только тогда, когда F выводима в исчислении $\mathbb{C}_\alpha$.*

Доказательство. Пусть формула F – выводима в $\mathbb{C}_\alpha$. Поскольку в процессе вывода формула F конечным числом применений преобразования α приводится к виду $\perp$, то для доказательства противоречивости F достаточно доказать, что α – логически эквивалентное преобразование. Пусть F имеет канонический вид (1.1), (2.1), (2.2) или в сокращенном виде

$$F = (true \to |F|) = (true \to |F|_{-i^*} \vee \Omega_{i^*}), \quad \Omega_{i^*} = (A_{i^*} \wedge (B_{i^*j} \to \Phi_{i^*j}) \wedge |\Omega_{i^*}|_{-j^*}),$$

$$\Phi_{i^*j^*} = \bigvee_{k=1}^{l_{i^*j^*}} (C_{i^*j^*k} \wedge \Psi_{i^*j^*k}).$$

Отсюда в силу условия (2.2) и подстановочности S-формул эквивалентными преобразованиями сначала удаляем $B_{i^*j^*}$. После этого, воспользовавшись дистрибутивностью $\wedge$ относительно $\vee$, получим

$$F \Leftrightarrow \left(true \to |F|_{-i^*} \vee \left(\bigvee_{k=1}^{l_{i^*j^*}} \{A_{i^*} \cup C_{i^*j^*k}\} \wedge \Psi_{i^*j^*k} \wedge |\Omega_{i^*}|_{-j^*} \right)\right) \Leftrightarrow$$

$$\Leftrightarrow (true \to |F|_{-i^*} \vee \Omega_{i^*}^*) = \alpha(F).$$

Для доказательства полноты исчисления (т.е. того, что любая противоречивая формула выводима) покажем от противного, что если формула F не выводима, то она выполнима, иначе говоря, для F найдется модель I, т.е. интерпретация I, в которой F истинна. Пусть F имеет вид (1.1) и не выводима в исчислении $\mathbb{C}_\alpha$.

Рассмотрим каждую базу A_i, $i \in \overline{1,m}$. Предположим, что A_i – висячая вершина структуры формулы F (т.е. $\Omega_i = A_i$, $|\Omega_i| = \varnothing$). Тогда $A_i \neq false$ ввиду упрощений, упомянутых в конце разд. 1 (в связи с определением 1), и того, что $F \neq \perp$. Поэтому интерпретация I вида $I = A_i$ является моделью для Ω_i, а следовательно и для F.

Пусть вершина A_i – невисячая. Если у нее нет уместных вопросов B_{ij}, $j \in \overline{1,n_i}$, то формируем интерпретацию

$$I = A_i \wedge \left(\bigwedge_{j=1}^{n_i} \neg B_{ij} \right),$$

где $\neg B_{ij}$ – отрицание конъюнкта B_{ij}. Очевидно, что F истинна в ней, так как конъюнкты B_{ij}, с которых начинаются все члены конъюнкции

$$|\Omega_i| = \bigwedge_{j=1}^{n_i} (B_{ij} \to \Phi_{ij}),$$

– посылки.

Пусть теперь у невисячей базы A_i множество уместных вопросов B_{ij} не пусто. Сформируем из них очередь $B_i = (B_{ij_1}, \ldots, B_{ij_q})$, $\{j_1, \ldots, j_q\} \subseteq \overline{1,n_i}$. Если после просмотра всех баз модель не получена, то очередь каждой базы непустая. Применим α-преобразование к первым вопросам B_{ij_1} всех очередей B_i параллельно. При этом базы A_i приобретут вид $\{A_i \cup C_{ij_1k}\}$, $k \in \overline{1,l_{ij_1}}$, а ГП типа $(B_{ij_1} \to \Phi_{ij_1})$ в них укоротятся до Ψ_{ij_1k}; при $l_{ij_1} > 1$ базы размножаются в количестве l_{ij_1}.

Повторим выполненный просмотр баз с намерением обнаружить модель I для F. Если, действуя как и раньше, найдем, то доказательство выполнимости F будет завершено, иначе снова для каждой базы из уместных вопросов формируем очередь для применения к ним α-преобразования. В случае неисчерпанности у базы прежней очереди и появления у ней новых уместных вопросов старую очередь с конца наращиваем этими вопросами. Ко всем первым элементам очередей снова применяется α-преобразование и т.д.

В силу невыводимости F и конечности формулы процесс завершится получением такой формулы $\alpha^n(F)$, у которой появится либо висячая база (т.е. без каких-либо вопросов), либо невисячая без уместных вопросов. В обоих случаях конъюнкт базы или его расширение, аналогичное выше использованному, будет искомой моделью I для F. Теорема 1 доказана.

Пусть $\{T, G, \Delta\} \subset L_S$, T – теория предметной области (контент некоторой базы знаний), G – наблюдаемый эффект. *Объяснением* наблюдаемого признается гипотеза Δ, удовлетворяющая требованиям:

а) из $T \wedge \Delta$ в исчислении $\mathbb{C}_\beta$ выводимо G,

б) $T \wedge \Delta$ – непротиворечиво.

Теорема 2. *Пусть $F^0 = (T \wedge \neg G)^{L_S}$ – образ отрицания формулы $(T \to G)$ в языке L_S. Пусть также $P \in L_S$ и P построено (α, β)-гипотезированиями в процессе вывода формулы F^0 в исчислении $\mathbb{C}_\beta$. Тогда $P \leftrightarrow (T \to G)$ и для любой гипотезы Δ, объясняющей G, справедливо $\Delta \to P$.*

Доказательство. Докажем от противного, что $P \to (T \to G)$. Пусть истинно P, а $(T \to G)$ ложно, т.е. имеет место $P \wedge F^0$. Условие P имеет вид

$$P = \bigwedge_{k=1}^{n} P^k, \quad P^k = P(F_\alpha^{k-1}) = \bigwedge_{i=1}^{m^k} P_i^k,$$

где $F_\alpha^{k-1} = \alpha^{r_k}(F^{k-1})$, $F^k = \beta(F_\alpha^{k-1})$.

Пусть F_α^{k-1} имеет вид

$$F_\alpha^{k-1} = \left(true \to \bigvee_{i=1}^{m^k} \Omega_i \right). \tag{3.1}$$

При β-преобразовании формулы F_α^{k-1} все P_i^k встраиваются в Ω_i конъюнктивно (см. (2.4), (2.5)), усиливая Ω_i до Ω_i^*, т.е. $F^k = F_\alpha^{k-1}(\Omega_i^*/\Omega_i,\ \forall i \in \overline{1, m^k})$, $\Omega_i^* = P_i^k \wedge \Omega_i$.

Так как $F_\alpha^{k-1} \leftrightarrow F^{k-1}$,

$$P^k \wedge F^{k-1} \leftrightarrow P^k \wedge \left(\bigvee_{i=1}^{m^k} \Omega_i \right) \leftrightarrow \bigvee_{i=1}^{m^k} (P^k \wedge \Omega_i) \leftrightarrow F^k \quad (\forall k \in \overline{1, n}),$$

а $P = \bigwedge_{k=1}^{n} P^k$, то $P \wedge F^0 \to F^n$. Поскольку $F^n \leftrightarrow F_\alpha^n = \bot$, то полученное противоречие доказывает, что $P \to (T \to G)$: в теории T условие P – объяснение наблюдаемого G.

Рассмотрим теперь необходимость условия P, т.е. что $(T \to G) \to P$. Снова действуя от противного, предположим, что справедливо $(T \to G) \wedge \neg P$, т.е. $\neg F^0 \wedge \neg P$. Проверим, что это приводит к противоречию. Вначале докажем, что

$$\forall k \in \overline{1, n} \quad (\neg F_\alpha^{k-1} \wedge \neg P^k \to false). \tag{3.2}$$

Пусть F_α^{k-1} имеет вид (3.1). Так как

$$\neg F_\alpha^{k-1} \leftrightarrow \bigwedge_{i=1}^{m^k} \neg \Omega_i, \quad \neg P^k \leftrightarrow \bigvee_{i=1}^{m^k} \neg P_i^k,$$

то для доказательства (3.2) достаточно вывести противоречие из предположений о том, что справедливо $\forall i \in \overline{1, m^k} \neg \Omega_i$ и $\exists i^* \in \overline{1, m^k} \neg P_{i^*}^k$. Возможны только два случая, разбором которых обосновывается (3.2):

1) $\Omega_{i^*} = A_{i^*}$, $P_{i^*}^k = (A_{i^*} \to false)$, при этом $\neg \Omega_i \wedge \neg P_{i^*}^k \leftrightarrow \neg A_{i^*} \wedge A_{i^*} \leftrightarrow false$;

2) $\Omega_{i^*} = A_{i^*} \wedge \left(\bigwedge_{j=1}^{n_{i^*}} (B_{i^* j} \to \Phi_{i^* j}) \right)$, $P_{i^*}^k = \left(A_{i^*} \to \bigvee_{j=1}^{n_{i^*}} B_{i^* j} \right)$, при этом

$$\neg \Omega_{i^*} \wedge \neg P_{i^*}^k \leftrightarrow \left[\neg A_{i^*} \vee \left(\bigvee_{j=1}^{n_{i^*}} (B_{i^* j} \wedge \neg \Phi_{i^* j}) \right) \right] \wedge \left[A_{i^*} \wedge \left(\bigwedge_{j=1}^{n_{i^*}} (\neg B_{i^* j}) \right) \right] \leftrightarrow false.$$

Так как $F^k = \beta(F_\alpha^{k-1})$, $F_\alpha^{k-1} \leftrightarrow F^{k-1}$ и $\beta(F_\alpha^{k-1}) \to F_\alpha^{k-1}$, то $\forall k \in \overline{1,n}\ (F^k \to F^{k-1})$. Отсюда и из (3.2) следует, что $\neg F^0 \to P^1 \wedge \neg F^1, \neg F^1 \to P^2 \wedge \neg F^2\ ,\ldots,\ \neg F^{n-1} \to P^n \wedge \neg F^n$. Поэтому $\neg F^0 \to \bigwedge_{k=1}^{n} P^k$, т.е. $(T \to G) \to P$, что и требовалось доказать.

Поскольку всякая гипотеза, объясняющая наблюдаемое G, должна удовлетворять по определению условию $T \wedge \Delta \to G$, то отсюда и из доказанного свойства необходимости гипотезы-миноранты P вытекает, что $\Delta \to P$. Теорема 2 доказана.

4. Примеры применения исчислений. Рассмотрим примеры гипотез-минорант P и достаточных условий их выполнимости, полученных путем удаления из минорант элементов, несовместимых с теорией T (примеры 1 и 2), или путем привлечения эмпирических знаний (пример 2). Эмпирические знания могут расширять знания базовой теории об априорном множестве потенциально возможных причин-кандидатов и, кроме того, могут дополнять базовую теорию достоверных причинно-следственных связей также лишь правдоподобными причинно-следственными связями. Преобразованиями миноранты P получаются либо сами причины Δ-эффекта G (достоверные или правдоподобные), либо релевантные обстоятельства (пример 1), из которых в рамках теории T средствами исчислений $\mathbb{C}_\alpha$ или $\mathbb{C}_\beta$ оказываются выводимыми и сами причины. Не предполагается какой-либо предварительной трансформации теории T до формы “if-and-only-if” (см. Введение) с ацикличностью множества “причинных переменных”, частично упорядоченного импликацией [1, 2].

Пример 1. Рассмотрим простой пример из области диагностики автомобиля (знания приведенного ниже типа сформулированы в [7] для иллюстрации концепции экспертных систем в ИИ).

Пусть некоторые пропозициональные переменные языка L_S имеют следующий смысл: Ф – “*Фары горят*”; Б – “*в Баке топливо есть*”; К – “*топливо поступает в Карбюратор*”; Д – “*топливо поступает в Двигатель*”; В – “*двигатель Вращается*”; И – “*автомобиль Исправен*”; А – “*проблема – в Аккумуляторе*”, З – “*проблема – в свечах Зажигания*”, С – “*проблема – в Стартере*”. Допустим, что автомобиль неисправен и выявлены факты Ф, В, К. Предположим, что, помимо этих фактов, базовая теория T включает знания, которые в естественном языке имеют вид: *Если топливо поступает в двигатель и двигатель вращается, то проблема – в свечах зажигания. Если в баке топливо есть и топливо поступает в карбюратор, то топливо поступает в двигатель. Если двигатель не вращается, то при горящих фарах проблема – в стартере, при негорящих – в аккумуляторе. Если автомобиль исправен, то проблем в свечах зажигания, стартере и аккумуляторе нет.*

Базовая теория T в языке L_S представима формулой:

$$\begin{aligned} T = \{Ф, В, Б\} \wedge [\ (\{Д, В\} \to З) \wedge (\{Б, К\} \to Д) \wedge \\ \wedge[\neg В \to (true \wedge (\ (Ф \to С) \wedge (\neg Ф \to А)\ )\ )] \wedge \\ \wedge(И \to \{\neg З, \neg С, \neg А\})\ ]. \end{aligned}$$

Наблюдаемое G состоит в неисправности автомобиля: $G = \neg И$, а вывод объяснения в S-исчислении $\mathbb{C}_\beta$ начинается с формулы $F^0 = (T \wedge \neg G)^{L_S}$. Ее база $\{Ф, В, Б, И\}$ при получении F^1 (однократным применением α-преобразования к F^0) дополнится элементами множества $\{\neg З, \neg С, \neg А\}$.

Применением β-преобразования получится гипотеза-миноранта

$$P = P^1 = P(\alpha^{r_1}(F^0)) = P(F^1) = (\{Ф, В, Б, И, \neg З, \neg С, \neg А\} \to (Д \vee К \vee \neg В)).$$

После представления ее в форме элементарной дизъюнкции и удаления из нее членов $\neg Ф, \neg В, \neg Б, \neg И$, контрарных элементам базы из F^1, а также тривиального члена $З \vee С \vee А$ получается объяснение $Д \vee К$ в классе релевантных обстоятельств. Действительно, далее в рамках исчисления $\mathbb{C}_\alpha$ промежуточным элементом вывода формулы $(T \wedge \neg G \wedge (Д \vee К))^{L_S}$, включившей полученное релевантное обстоятельство, оказывается факт $З$ как причина неисправности.

Пример 2. Рассмотрим другой иллюстративный пример, а именно из области медицины [1], не приводя для краткости детальной семантики переменных (статья [1] посвящена представлению знаний в диагностике с достоверными и эмпирическими знаниями).

Пусть e, h – симптомы заболевания. Известны болезни t, d, a. Допустим, что $\{t \to e,\ a \to h\}$ – множество достоверных причинно-следственных связей базовой теории T в терминах симптомов e,

h и болезней t, a: “*болезнь t – причина симптома e* “; “*болезнь a – причина симптома h* “. Пусть $\{d \to h, a \to e\}$ – правдоподобные (недостоверные) причинно-следственные связи (ППСС) как эмпирические данные “прошлого опыта”; например, ППСС $d \to h$ (соответственно $a \to e$) – это высказывание: “*болезнь d* (соответственно a) *иногда может являться причиной симптома h* (соответственно e)”.

Пример интересен, в частности, и тем, что словарь базовой теории T не охватывает словаря ППСС (нет болезни d). Тем не менее, из гипотезы-миноранты, выводимой средствами (α, β)-гипотезирования, получаются все логически возможные в рассматриваемом контексте диагнозы.

Если наблюдается симптом $G = e$, то $F^0 = (T \wedge \neg G)^{L_S} = (true \to \neg e \wedge [(t \to e) \wedge (a \to h)])$.

В исчислении $\mathbb{C}_\beta$ при выводе F^0 формируется гипотеза-миноранта $P = \bigwedge_{k=1}^{2} P(\alpha^{r_k}(F^{k-1})) =$
$= (\neg e \to (t \vee a)) \wedge (\{\neg e, a, h\} \to t)$.

При этом $r_1 = 0$, $r_2 = 3$, $r_3 = 2$ и $\alpha^{r_3}(F^2) = \perp$. Все элементарные дизъюнкции конъюнктивной нормальной формы (КНФ) $(e \vee t \vee a) \wedge (e \vee t \vee \neg a \vee \neg h)$ гипотезы-миноранты P следуют из t, т.е. болезнь t – объяснение симптома e.

Этот диагноз не использует ППСС. С учетом же их и того, что оба члена КНФ следуют из a и $a \to e$ соответственно, выводится еще одно (лишь правдоподобное) объяснение: “причина – в болезни a, если верна ППСС $a \to e$”.

Пусть теперь наблюдаются одновременно не один, а два симптома: $G = e \wedge h$. Выявление причин использует вывод формулы

$$F^0 = (T \wedge \neg G)^{L_S} = (true \to true \wedge [(t \to e) \wedge (a \to h) \wedge (e \to \neg h)]).$$

При этом формируется гипотеза-миноранта

$$P = \bigwedge_{k=1}^{3} P^k = P_1^2 \wedge \left(\bigwedge_{i=1}^{3} P_i^2\right) \wedge P_1^3 \leftrightarrow$$

$$\leftrightarrow ((t \vee a \vee e) \wedge (\neg t \vee a \vee \neg e \vee h) \wedge (t \vee \neg a \vee e \vee \neg h) \wedge (t \vee a \vee \neg e \vee h)) \quad (4.1)$$

и на ее основе – три следующих объяснения:

1) “комплекс из двух причин $\{t, a\}$” (достоверно);

2) “если верна ППСС $a \to e$, то причина – в a” (правдоподобно);

3) “комплекс из двух причин d, t и ППСС $d \to h$” (правдоподобно).

Третье объяснение получается из (4.1) после расширения сигнатуры формулы (4.1) дополнительной причинной переменной d путем замены каждой элементарной дизъюнкции на две другие, получающиеся добавлением в одну переменной d, а во вторую – литералы $\neg d$. Каждая из элементарных дизъюнкций расширенной КНФ будет следовать из причин d, t и ППСС $d \to h$, что и оказывается третьим объяснением наблюдаемого комплекса симптомов $G = e \wedge h$. Заметим, что расширение словаря для применимости ППСС $d \to h$ с формированием этого объяснения было необходимо для выполнения лишь второго члена гипотезы-миноранты из (4.1).

Заключение. Работа посвящена проблематике абдуктивного объяснения наблюдаемых эффектов (аномалий, симптомов). Изложен и обоснован метод логического формирования объяснений. Предложены средства представления знаний и вывода объясняющих гипотез. Введен язык, обладающий свойством подстановочности. Свойства языка и вводимых в нем исчислений обеспечивают абдукцию путем сочетания дедукции с гипотезированием. В отличие от известных логических средств абдукции, предложенным методом выводятся гипотезы-миноранты, являющиеся логически необходимыми и достаточными условиями для формального объяснения наблюдаемого. Эту миноранту любая объясняющая гипотеза должна иметь своим логическим следствием.

Гипотеза-миноранта выступает контрольным условием и опорным средством формирования достоверных гипотез, а в случае наличия сверхбазовой теории также эмпирических знаний, то и средством формирования также правдоподобных гипотез. Эмпирические знания могут расширять знания базовой теории об априорном множестве потенциально возможных причин-кандидатов и, кроме того, могут дополнять базовую теорию достоверных причинно-следственных связей также ППСС. Из гипотезы-миноранты переходом к достаточным условиям путем ее упрощения и использования эмпирических знаний получаются либо причины наблюдаемого эффекта (достоверные или правдоподобные), либо поначалу лишь релевантные обстоятельства (факты, события, “улики”), из которых в рамках базовой теории средствами разработанных исчислений оказываются выводимыми и сами причины. Не предполагается какой-либо предварительной трансформации базовой теории, нередко используемой в литературе для упрощения вывода гипотез. Предложенный метод абдуктивного вывода обоснован и проиллюстрирован примерами.

Актуальной задачей остается проблематика ранжирования альтернативных объяснений наблюдаемого для сужения множества гипотез до наиболее достоверных с использованием вероятностных, логических и других оценок. Актуальной задачей на будущее, особенно в проблематике интеллектуализации автономных (беспилотных) агентов, является также задача абдуктивной информационной поддержки планирования действий агентов с развитием полученных здесь результатов.

СПИСОК ЛИТЕРАТУРЫ

1. *Poole D.* Representing Diagnosis Knowledge // J. Annals of Mathematics and Artificial Intelligence. 1994. V. 11. P. 33–50.
2. *Kowalski R.* Logic Programming // Computational Logic. Edition: In the History of Logic series. Eds. D. Gabbay and J. Woods. Amsterdam: Elsevier. 2014. P. 523–569.
3. *Финн В.К.* Об эвристиках ДСМ-исследований // Научно-техническая информация. Сер. 2. 2019. № 10. С. 1–34.
4. *Васильев С.Н.* Формализация знаний и управление на основе позитивно-образованных языков // Информационные технологии и вычислительные системы. 2008. № 1. С. 3–19.
5. *Васильев С.Н., Жерлов А.К.* Об исчислениях типово-кванторных формул // Докл. АН. 1995. Т. 343. № 5. С. 583–585.
6. *Kowalski R., Sadri F.* Reactive Computing as Model Generation // New Generation Computing. 2015. V. 33. P. 33–67.
7. *Люгер Дж.Ф.* Искусственный интеллект: стратегии и методы решения сложных проблем. М.: Вильямс, 2003. 864 с.