В.К. Финн

Об определении эмпирических закономерностей посредством ДСМ - метода автоматического порождения гипотез[1]

Аннотация. В статье даются определения эмпирических закономерностей, обнаруженных в базах фактов посредством ДСМ-рассуждения. Эти закономерности подразделяются на эмпирические законы и тенденции.

Ключевые слова: ДСМ-метод автоматического порождения гипотез, ДСМ-рассуждения, степень противоречивости, абдуктивная сходимость, эмпирические закономерности (законы, тенденции).

«…что же представляет собой индуктивная, или человеческая логика – логика истины. Её дело рассмотреть методы мышления и обнаружить, какие степени доверия в них следовало бы установить, т.е. в какой пропорции они ведут к истине». И далее: «…она (индуктивная логика – В.Ф.) должна рассмотреть соответствующую обоснованность различных типов научных процедур, таких как поиск закона причинности согласно методам Милля, и современных математических методов, типа априорных доказательств, использованных при открытии теории относительности. Надлежащий план такого исследования следует искать у Милля».

Ф.П. Рамсей «Философские работы». Издательство Томского университета, 2003, стр. 149-150.

1. Некоторые понятия ДСМ - метода

В [1] и [2] были определены пять индуктивных методов Д.С. Милля [3] посредством средств современной логики. В [2] были также определены функционалы степени противоречивости ДСМ – рассуждений f^{σ} и F^{σ}, где $\sigma \in \{+, -, 0\}$.

В частности, $f^{+}(\Delta^{+}(p), \Delta^{-}(q) \cup \Delta^{0}(q)) = \dfrac{\left|\Delta^{+}(p) \cap (\Delta^{-}(q) \cup \Delta^{0}(q))\right|}{\left|\Delta^{+}(p)\right|}$, $F^{+}(\Omega^{+}(p), \Omega^{-}(q) \cup \Omega^{0}(q)) = \dfrac{\left|\Omega^{+}(p) \cap (\Omega^{-}(q) \cup \Omega^{0}(q))\right|}{\left|\Omega^{+}(p)\right|}$, где p и q – номера соответствующих Этапа I_p и Этапа I_q процесса ДСМ – рассуждений, а $\Delta^{\sigma}(p)$, $\Delta^{\sigma}(q)$; $\Omega^{\sigma}(p)$, $\Omega^{\sigma}(q)$ – множества гипотез с предикатом $\Rightarrow_2$ и $\Rightarrow_1$, соответственно ($\sigma \in \{+, -, 0\}$).

Аналогично определяются функционалы f^{σ} и F^{σ}, где $\sigma \in \{-, 0\}$.

В [2] были определены предикаты $\mathrm{R}^{\sigma}(p, q) \rightleftharpoons f^{\sigma}(\Delta^{\sigma}(p), \Delta^{\sigma_1}(q) \cup \Delta^{\sigma_2}(q)) = 0$, где $\sigma \neq \sigma_1$, $\sigma \neq \sigma_2$, $\sigma_1 \neq \sigma_2$, а $\sigma, \sigma_1, \sigma_2 \in \{+, -, 0\}$. Там же определен предикат $\overline{R}(p, q) \rightleftharpoons \mathrm{R}^{+}(p, q) \& \mathrm{R}^{-}(p, q) \& \mathrm{R}^{0}(p, q)$.

Аналогично определяются предикаты $\mathrm{K}^{\sigma}(p, q) \rightleftharpoons F^{\sigma}(\Omega^{\sigma}(p), \Omega^{\sigma_1}(q) \cup \Omega^{\sigma_2}(q)) = 0$, где $\sigma \neq \sigma_1$, $\sigma \neq \sigma_2$, $\sigma_1 \neq \sigma_2$, а $\sigma, \sigma_1, \sigma_2 \in \{+, -, 0\}$; а также определяется предикат $\overline{K}(p, q) \rightleftharpoons \mathrm{K}^{+}(p, q) \& \mathrm{K}^{-}(p, q) \& \mathrm{K}^{0}(p, q)$.

[1] Статья является дополнением к статьям В.К. Финн «Индуктивные методы Д.С. Милля в системах искусственного интеллекта». Часть I и Часть II, Искусственный интеллект и принятие решений, №3 и №4, 2010.

В [2] было показано, что $\overline{R}(p, q)$ и $\overline{K}(p, q)$ представляют, соответственно, отношения толерантности [4], т.е. являются рефлексивными и симметричными: $\forall p\overline{R}(p, p)$, $\forall p\forall q(\overline{R}(p, q)\to \overline{R}(q, p))$, $\forall p\overline{K}(p, p)$, $\forall p\forall q(\overline{K}(p, q)\to \overline{K}(q, p))$[2].

ДСМ – рассуждение в [2] называется (σ)-корректным, если имеют место $R^{\sigma}(p, q)$ и $K^{\sigma}(p, q)$, где $\sigma\in\{+, -, 0\}$. ДСМ – рассуждение называется тотально корректным, если имеют место $\overline{R}(p, q)$ и $\overline{K}(p, q)$.

Существенно отметить, что (σ)-корректность и тотальная корректность ДСМ – рассуждений определяется относительно заданной последовательности баз фактов (БФ) $\text{БФ}_0\subset\text{БФ}_1\subset\ldots\subset\text{БФ}_s$ и применяемой стратегии ДСМ-рассуждения $Str_{x,y}$, где $\text{БФ}_i=\text{БФ}_i^{+}\cup\text{БФ}_i^{-}\cup\text{БФ}_i^{0}\cup\text{БФ}_i^{\tau}$, $(0\le i\le s)$, x – имя M^{+}-предиката, а y – имя M^{-}-предиката, используемых для определения правил правдоподобного вывода $1^{\underline{\text{ого}}}$ рода (индукции) [2].

Предположим, далее, что для последовательности расширяемых баз фактов $\text{БФ}_0\subset\text{БФ}_1\subset\ldots\subset\text{БФ}_s$ имеет место абдуктивная сходимость [1, 5]: $\rho^{\sigma}(0)\le \rho^{\sigma}(1)\le\ldots\le\rho^{\sigma}(s)\ge\rho^{\sigma}$,

где $\rho^{\sigma}(i)=\dfrac{|\,\widetilde{\text{БФ}}_i^{\sigma}\,|}{|\,\text{БФ}_i^{\sigma}\,|}$ – степень каузальной полноты, $i = 0, 1, \ldots, s$, ρ – заданный порог, $\sigma\in\{+, -, 0\}$,

$\widetilde{\text{БФ}}_i^{\sigma}$ –подмножество (σ)-фактов из БФ_i^{σ} такое, что для него выполняются $\text{АКП}^{(\sigma)}$ – аксиомы каузальной полноты [1, 5].

$\text{АКП}^{(+)}$: $\forall X\forall Y\exists k\exists V_1\ldots\exists V_k\ \exists W_1\ldots\exists W_k$

$(J_{\langle1,0\rangle}(X\Rightarrow_1 Y)\to\exists n(\mathop{\&}\limits_{i=1}^{k} J_{(1,n)}(V_i\Rightarrow_2 W_i)\&(V_i\subset X)\&$

$(V_i\neq\varnothing)\&\ (W_i\neq\varnothing))\&\ (\bigcup\limits_{i=1}^{k} W_i=Y)))$.

$\text{АКП}^{(-)}$ определяется аналогично. $\text{АКП}^{(0)}$ содержится в квазиаксиоматической теории (КАТ) $\mathcal{T} = \langle\Sigma, \Sigma', \mathcal{R}\rangle$ в качестве элемента множества аксиом Σ [1, 5], если в открытом множестве фактов и гипотез Σ′ содержатся факты с истинностными значениями $\overline{\nu} = \langle 0, 0\rangle$, где первая компонента $\overline{\nu}$ есть фактическое противоречие, а вторая – число применений правил правдоподобного вывода (п.п.в.). $\text{АКП}^{(0)}$: $\forall Z\forall U((J_{\langle0,0\rangle}(Z\Rightarrow_1 U)\to\exists n\,\Pi_n^{0}(Z, U))$, где $\Pi_n^{0}(Z, U)$ – предикат, представляющий посылку в правиле правдоподобного вывода $2^{\underline{\text{ого}}}$ рода (п.п.в.-2) для аналогии [1].

Таким образом, $\widetilde{\text{БФ}}_i^{+} = \{J_{\langle1,0\rangle}(C^{(i)}\Rightarrow_1 Q^{(i)})|$ $\exists V_1\exists W_1\ \ldots\ \exists V_k\exists W_k\exists k\exists n(\mathop{\&}\limits_{j=1}^{k}(J_{(1,n)}(V_j\Rightarrow_2 W_j)\ \&$ $(V_j\subset C^{(i)})\&(V_j\neq\varnothing)\ \&\ (W_j\neq\varnothing)\ \&\ (\bigcup\limits_{j=1}^{k_i} W_j=Q^{(i)})))\}$,

аналогично определяется $\widetilde{\text{БФ}}_i^{-}$, а $\widetilde{\text{БФ}}_i^{0}$ = $\{J_{\langle1,0\rangle}(X\Rightarrow_1 Y)|\ \exists X\exists Y\exists n\exists\,\Pi_n^{0}(X, Y)\}$.

ДСМ-метод автоматического порождения гипотез (ДСМ-метод АПГ), который используется для обнаружения **эмпирических закономерностей** в БФ, состоит из следующих компонент:

(1) условий применимости,

(2) ДСМ – рассуждений,

(3) представления знаний в виде квазиаксиоматических теорий (КАТ) $\mathcal{T} = \langle\Sigma, \Sigma', \mathcal{R}\rangle$, где $\mathcal{R}$ – множество правил вывода (правдоподобных и дедуктивных);

(4) метатеоретического исследования ДСМ-рассуждений и предметных областей (моделей данных), содержащего дедуктивную имитацию ДСМ-рассуждений;

(5) интеллектуальных систем типа ДСМ (ИС-ДСМ),

(6) обнаружений эмпирических закономерностей завершающих процесс *knowledge discovery* в БФ посредством ИС-ДСМ.

Для обнаружения эмпирических закономерностей (ЭЗК) необходимо, чтобы базы фактов (БФ) содержали (+)-факты (позитивные), (–)-факты (негативные), (τ)-факты (неопределенности) и, возможно, (0)-факты (фактические противоречия) [1, 2, 5].

Кроме того, необходимо, чтобы в БФ_i содержались в неявном виде эмпирические зависимости причинно-следственного типа ((+)-причины и (–)-причины изучаемых эффектов). И, наконец, должна быть определена операция сходства для (σ)-фактов и (σ)-гипотез (предсказаний наличия или отсутствия эффектов или их фактической противоречивости). Таковы условия применимости ДСМ – метода АПГ [1].

Основной компонентой ДСМ-метода АПГ являются ДСМ-рассуждения [1, 2, 5, 6] (2), которые

[2] См. [2]: утверждения 4,5,6,7.

образуют синтез трех познавательных процедур – индукции, аналогии и абдукции. Эти процедуры имитируют естественный познавательный цикл «анализ данных (индукция) – предсказание (аналогия) – объяснение (абдукция)».

Охарактеризуем строение ДСМ – рассуждения.

1. **Шагом** ДСМ – рассуждения называется однократное применение правил правдоподобного вывода первого рода (п.п.в.-1 - индукции) или п.п.в.-2 (аналогии).

2. **Тактом** ДСМ – рассуждения называется упорядоченное последовательное применение п.п.в.-1 и п.п.в.-2.

3. **Этапом I** ДСМ-рассуждения будем называть последовательное применение тактов $(\text{п.п.в.-}1 \to \text{п.п.в.-}2)_1 \to (\text{п.п.в.-}1 \to \text{п.п.в.-}2)_2 \to \ldots \to (\text{п.п.в.-}1 \to \text{п.п.в.-}2)_n$ такое, что множество порожденных гипотез на такте n совпадает с множеством гипотез, порожденных на такте $n+1$, где n – номер первого такого совпадения [5, 1].

Этот такт с номером n называется тактом **стабилизации** Этапа I ДСМ-рассуждения.

4. **Этапом II** ДСМ-рассуждения называется проверка выполнимости аксиом каузальной полноты АКП$^{(\sigma)}$, где $\sigma \in \{+, -, 0\}$, и вычисление степени каузальной полноты $\rho^{\sigma}(i)$, характеризующей абдуктивную сходимость или расходимость **процесса** ДСМ-рассуждений для последовательности $БФ_0 \subset БФ_1 \subset \ldots \subset БФ_s$, результатом которого является принятие или непринятие порождаемых гипотез [1].

Третьей компонентой ДСМ-метода АПГ (3), являются КАТ $\mathcal{T}$ такие, что $\mathcal{T} = \langle \Sigma, \Sigma', \mathcal{R} \rangle$, где Σ содержат АКП$^{(\sigma)}$, а Σ' удовлетворяют условиям применимости ДСМ-рассуждений, сформулированным в компоненте (1) ДСМ-метода АПГ, а $\mathcal{R}$ содержат п.п.в.-1 и п.п.в.-2.

Четвертой компонентой ДСМ-метода АПГ (4) является метатеоретическое исследование ДСМ-рассуждений и предметных областей (моделей данных), к которым они применяются. Эти исследования содержат дедуктивную имитацию ДСМ-рассуждений, а также семантику ДСМ-рассуждений (процедурную, аргументационную и семантику управления рассуждением [6]) и препроцессинг для выбора стратегии ДСМ-рассуждения $Str_{x,y}$ [6, 1, 2], соответствующей предметной области и её базам фактов в ИС-ДСМ.

Заметим, что основные теоремы о дедуктивной имитации ДСМ-рассуждений содержатся в [7]. Таковыми являются теорема о непротиворечивости процедур ДСМ-рассуждения, теорема о единственности модели, теорема об обратимости правил правдоподобного вывода (п.п.в.-1 и п.п.в.-2) и теорема о выразимости процедур ДСМ-рассуждения (индукции, аналогии и абдукции) в логике предикатов 1ого порядка для конечных моделей и в слабой логике предикатов 2ого порядка для моделей произвольной мощности.

Пятой компонентой ДСМ-метода АПГ являются интеллектуальные системы типа ДСМ (ИС-ДСМ) [8]. ИС-ДСМ имеют следующую архитектуру: ИС-ДСМ = Решатель задач + (База фактов + База знаний) + комфортный интерфейс, где Решатель = Рассуждатель + Вычислитель + Синтезатор, База фактов представлена последовательностью $БФ_0 \subset БФ_1 \subset \ldots \subset БФ_s$ под управлением АКП$^{(\sigma)}$, База знаний (БЗ) реализует КАТ $\mathcal{T} = \langle \Sigma, \Sigma', \mathcal{R} \rangle$[3]. Главным модулем ДСМ-Решателя является **Рассуждатель**, реализующий синтез познавательных процедур, т.е. ДСМ-рассуждения (индукция + аналогия + абдукция). Это означает, что в Рассуждателе осуществляются шаги, такты и Этапы I и II ДСМ-рассуждения. Синтезатор реализует взаимодействие Расссуждателя и Вычислителя, а также формирует стратегии ДСМ – рассуждений для порождения гипотез и распознавания эмпирических закономерностей в последовательностях баз фактов $БФ_0 \subset БФ_1 \subset \ldots \subset БФ_s$ посредством проверки (σ)-корректности, тотальной корректности и абдуктивной сходимости ДСМ-рассуждений.

Шестой компонентой ДСМ-метода АПГ как раз и является распознавание эмпирических закономерностей в базах фактов, что завершает работу ДСМ-метода АПГ для обнаружения *knowledge discovery* в ИС-ДСМ [9].

2. Эмпирические закономерности: законы и тенденции

Итак, пусть заданы последовательность БФ $БФ_0 \subset БФ_1 \subset \ldots \subset БФ_s$ и стратегия ДСМ-рассуждения $Str_{x,y}$, реализуемая Рассуждателем ИС-ДСМ. Сформулируем ниже **явные определения** [10] предикатов $L^{\sigma}(Z, U, p, q)$ и $\widetilde{L}^{\sigma}(V, W, p, q)$, ха-

[3] Говоря точнее, отметим, что в БЗ входит часть Σ', образованная порожденными гипотезами, её дополнение есть БФ.

рактеризующих **семейства эмпирических закономерностей** (ЭЗК) и **основания ЭЗК** этого **семейства**.

Df. 1-1. $\forall Z \forall U \forall p \forall q(L^{+}(Z, U, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(Z \Rightarrow_1 U) \& \Pi^{+}_{n}(Z,U) \rightarrow J_{\langle 1,n+1\rangle}(Z \Rightarrow_1 U)) \& (0 \leq F^{+}(\Omega^{+}(p), \Omega^{-}(q) \cup \Omega^{0}(q)) \leq 0,2) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$, где $\Pi^{+}_{n}(Z,U)$ – предикат для представления аналогии в посылках п.п.в.-2 [1].

Df. 1-2. $\forall V \forall W \forall p \forall q(\widetilde{L}^{+}(V, W, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(V \Rightarrow_2 W) \& M^{+}_{x,n}(V,W) \& \neg M^{-}_{y,n}(V,W)) \rightarrow J_{\langle 1,n+1\rangle}(V \Rightarrow_2 W)) \& (0 \leq f^{+}(\Delta^{+}(p), \Delta^{-}(q) \cup \Delta^{0}(q)) \leq 0,2) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$, где $M^{+}_{x,n}(V,W)$ и $M^{-}_{y,n}(V,W)$ – предикаты для представления индукции в посылках п.п.в.-1, соответствующие $Str_{x,y}$.

Аналогично определяются предикаты $L^{\sigma}(Z, U, p, q)$ и $\widetilde{L}^{\sigma}(V, W, p, q)$, где $\sigma \in \{-, 0\}$. В этих предикатах имеются, соответственно, $\Pi^{-}_{n}(Z,U)$ и $\Pi^{0}_{n}(Z,U)$, $\neg M^{+}_{x,n}(V,W) \& M^{-}_{y,n}(V,W)$ и $M^{+}_{x,n}(V,W) \& M^{-}_{y,n}(V,W)$ [1].

Важным видом эмпирических закономерностей являются **эмпирические законы**, семейство которых определяется ниже посредством предикатов $L^{\sigma}_{1}(Z, U, p, q)$, где $\sigma \in \{+, -, 0\}$.

Df. 2-1. $\forall Z \forall p \forall q(L^{+}_{1}(Z, Q, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(Z \Rightarrow_1 Q) \& \Pi^{+}_{n}(Z, Q) \rightarrow J_{\langle 1,n+1\rangle}(Z \Rightarrow_1 Q)) \& \overline{K}(p, q) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$, где Q – константа, представляющая эффект в БФ.

Аналогично формулируется явное определение для $L^{-}_{1}(Z, Q, p, q)$.

Определим также **основания позитивных эмпирических законов** посредством предиката $L^{+}_{2}(V, W, p, q)$.

Df. 2-2. $\forall V \forall p \forall q(L^{+}_{2}(V, Q, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(V \Rightarrow_2 Q) \& M^{+}_{x,n}(V,Q) \& \neg M^{-}_{y,n}(V,Q)) \rightarrow J_{\langle 1,n+1\rangle}(V \Rightarrow_2 Q)) \& \overline{R}(p, q) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$.

Аналогично формулируется явное определение для $L^{-}_{2}(V, Q, p, q)$.

Позитивный эмпирический закон из семейства *Df.* 2-1 представим утверждением

(2-1) $\exists Z \forall p \forall q\, L^{+}_{1}(Z, Q, p, q)$.

Основание эмпирического закона из семейства *Df.* 2-2 представим утверждением

(2-2) $\exists V \forall p \forall q\, L^{+}_{2}(V, Q, p, q)$.

Реализациями позитивных эмпирического закона и его оснований будем называть, соответственно утверждения (2-1.1) $\forall p \forall q\, L^{+}_{1}(C, Q, p, q)$ и (2-2.1) $\forall p \forall q\, L^{+}_{2}(C', Q, p, q)$, где C и C′ – константы, представляющие носителя эффекта Q и его причину, соответственно.

Аналогично определяются негативные эмпирические законы и их основания, а также их реализации $\forall p \forall q\, L^{-}_{1}(C, Q, p, q)$ и $\forall p \forall q\, L^{-}_{2}(C', Q, p, q)$.

Очевидно, что для реализаций эмпирических законов и их оснований имеют место утверждения (*a*) $\forall p \forall q(L^{+}_{1}(C, Q, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(C \Rightarrow_1 Q) \& \Pi^{+}_{n}(C, Q) \rightarrow J_{\langle 1,n+1\rangle}(C \Rightarrow_1 Q)) \& \overline{K}(p, q) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$, (*b*) $\forall p \forall q(L^{+}_{2}(C', Q, p, q) \leftrightarrow \exists n \exists s(((J_{\langle\tau,n\rangle}(C' \Rightarrow_2 Q) \& M^{+}_{x,n}(C',Q) \& \neg M^{-}_{y,n}(C',Q)) \rightarrow J_{\langle 1,n+1\rangle}(C' \Rightarrow_2 Q)) \& \overline{R}(p, q) \& (\rho^{+} \leq \rho^{+}(s) \leq 1)))$.

Аналогичны утверждения (*c*) и (*d*) для $L^{-}_{1}(C, Q, p, q)$ и $L^{-}_{2}(C', Q, p, q)$.

Выразимы также эмпирические козаконы и их коснования: $\exists U \forall p \forall q(L^{\sigma}_{1}(C, U, p, q)$ и $\exists W \forall p \forall q(L^{\sigma}_{2}(C', W, p, q)$, где $\sigma \in \{+, -\}$ (возможно, что $C' \subset C$).

Заменив в определениях *Df.* 2-1 и *Df.* 2-2 подформулу $(\rho^{+} \leq \rho^{+}(s) \leq 1)$ на $\rho^{+}(s)=1$ и $\rho^{+} \leq \rho^{+}(s) < 1$ (например, $\rho^{+} = 0,8$), получим определение для сильного и слабого позитивного эмпирического закона и сильного и слабого основания позитивного эмпирического закона. Обозначим посредством $L^{+}_{1s}(Z, U, p, q)$ и $L^{+}_{1w}(Z, U, p, q)$ предикаты для представления сильного позитивного эмпирического закона и слабого позитивного эмпирического закона, соответственно. Аналогичные обозначения $L^{+}_{2s}(V, W, p, q)$ и $L^{+}_{2w}(V, W, p, q)$ введем для представления сильного и слабого основания позитивных эмпирических законов.

Таким образом, $\exists Z \forall p \forall q\, L^{+}_{1s}(Z, Q, p, q)$ и $\exists Z \forall p \forall q\, L^{+}_{1w}(Z, Q, p, q)$ выражают сильный и слабый эмпирический закон, соответственно. Аналогичное имеет место для оснований позитивных эмпирических законов $L^{+}_{2s}(V, W, p, q)$ и $L^{+}_{2w}(V, W, p, q)$, а также и для эмпирических негативных законов и их оснований, представимых посредством предикатов L^{-}_{1s}(Z, U, p, q), L^{-}_{1w}(Z, U, p, q) и L^{-}_{2s}(V, W, p, q) и L^{-}_{2w}(V, W, p, q).

Заменим теперь в *Df.* 1-1 подформулу $(0\le F^{+}(\Omega^{+}(p), \Omega^{-}(q)\cup\Omega^{0}(q))\le 0{,}2)$ на подформулу $(0< F^{+}(\Omega^{+}(p), \Omega^{-}(q)\cup\Omega^{0}(q))\le 0{,}2)$, а в *Df.* 1-2 заменим подформулу $(0\le f^{+}(\Delta^{+}(p), \Delta^{-}(q)\cup\Delta^{0}(q))\le 0{,}2)$ на подформулу $(0< f^{+}(\Delta^{+}(p), \Delta^{-}(q)\cup\Delta^{0}(q))\le 0{,}2)$ и получим, соответственно, явные определения предикатов L^{+}_{3}(Z, U, p, q) и L^{+}_{4}(V, W, p, q), характеризующих семейство эмпирических закономерностей и их оснований таких, что степени противоречивости множеств порождаемых гипотез посредством п.п.в.-1 и п.п.в.-2 не равна 0, но достаточно мала.

Тогда будем говорить, что предикаты L^{+}_{3}(Z, U, p, q) и L^{+}_{4}(V, W, p, q) характеризуют **семейство позитивных эмпирических тенденций и их оснований**, соответственно. А, именно:

Df. 3-1. $\forall \mathrm{Z}\forall \mathrm{U}\forall p\forall q(\mathrm{L}^{+}_{3}(\mathrm{Z}, \mathrm{U}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{Z}\Rightarrow_1\mathrm{U})\&\Pi^{+}_{n}(\mathrm{Z},\mathrm{U})\rightarrow J_{\langle 1,n+1\rangle}(\mathrm{Z}\Rightarrow_1\mathrm{U}))\ \& (0<F^{+}(\Omega^{+}(p), \Omega^{-}(q)\cup\Omega^{0}(q))\le 0{,}2)\ \&(\rho^{+}\le\rho^{+}(s)\le 1)))$.

Df. 3-2. $\forall \mathrm{V}\forall \mathrm{W}\forall p\forall q(\mathrm{L}^{+}_{4}(\mathrm{V}, \mathrm{W}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{V}\Rightarrow_2\mathrm{W})\&\ \mathrm{M}^{+}_{x,n}(\mathrm{V},\mathrm{W})\&\neg \mathrm{M}^{-}_{y,n}(\mathrm{V},\mathrm{W}))\rightarrow J_{\langle 1,n+1\rangle}(\mathrm{V}\Rightarrow_2\mathrm{W}))\&\ (0<f^{+}(\Delta^{+}(p), \Delta^{-}(q)\cup\Delta^{0}(q))\le 0{,}2)\ \&(\rho^{+}\le\rho^{+}(s)\le 1)))$.

Аналогично формулируется условие, характеризующее семейство негативных эмпирических тенденций и их оснований посредством предикатов L^{-}_{3}(Z, U, p, q) и L^{-}_{4}(V, W, p, q), соответственно.

Ниже сформулируем явные определения семейств позитивных (негативных) тенденций и их оснований:

Df. 4-1. $\forall \mathrm{Z}\forall p\forall q(\mathrm{L}^{\sigma}_{3}(\mathrm{Z}, \mathrm{Q}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{Z}\Rightarrow_1\mathrm{Q})\&\ \Pi^{\sigma}_{n}(\mathrm{Z},\mathrm{Q})\rightarrow J_{\langle\nu,n+1\rangle}(\mathrm{Z}\Rightarrow_1\mathrm{Q}))\ \& (0<F^{+}(\Omega^{+}(p), \Omega^{-}(q)\cup\Omega^{0}(q))\le 0{,}2)\ \&(\rho^{+}\le\rho^{+}(s)\le 1)))$, где

$$\nu = \begin{cases} 1, \textit{если } \sigma = + \\ -1, \textit{если } \sigma = - \end{cases}.$$

Df. 4-2. $\forall \mathrm{V}\forall p\forall q(\mathrm{L}^{+}_{4}(\mathrm{V}, \mathrm{Q}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{V}\Rightarrow_2\mathrm{Q})\&\ \mathrm{M}^{+}_{x,n}(\mathrm{V},\mathrm{Q})\&\neg \mathrm{M}^{-}_{y,n}(\mathrm{V},\mathrm{Q}))\rightarrow J_{\langle 1,n+1\rangle}(\mathrm{V}\Rightarrow_2\mathrm{Q}))\&\ (0<f^{+}(\Delta^{+}(p), \Delta^{-}(q)\cup\Delta^{0}(q))\le 0{,}2)\ \&(\rho^{+}\le\rho^{+}(s)\le 1)))$.

Эмпирические тенденции и их основания из соответствующего семейства представим утверждениями

(3-1) $\exists \mathrm{Z}\forall p\forall q\, \mathrm{L}^{\sigma}_{3}(\mathrm{Z}, \mathrm{Q}, p, q)$,

(3-2) $\exists \mathrm{V}\forall p\forall q\, \mathrm{L}^{\sigma}_{4}(\mathrm{V}, \mathrm{Q}, p, q)$, где $\sigma\in\{+, -\}$.

Соответственно, определим сильные и слабые эмпирические тенденции и их основания посредством утверждений

(4-1) $\exists \mathrm{Z}\forall p\forall q\, \mathrm{L}^{\sigma}_{3s}(\mathrm{Z}, \mathrm{Q}, p, q)$,

(4-2) $\exists \mathrm{Z}\forall p\forall q\, \mathrm{L}^{\sigma}_{3w}(\mathrm{Z}, \mathrm{Q}, p, q)$,

(5-1) $\exists \mathrm{V}\forall p\forall q\, \mathrm{L}^{\sigma}_{4s}(\mathrm{V}, \mathrm{Q}, p, q)$,

(5-2) $\exists \mathrm{V}\forall p\forall q\, \mathrm{L}^{\sigma}_{4w}(\mathrm{V}, \mathrm{Q}, p, q)$, где $\sigma\in\{+, -\}$.

Сильные и слабые реализации тенденций и их основания представляются утверждениями $\forall p\forall q\, \mathrm{L}^{\sigma}_{3s}(\mathrm{C}, \mathrm{Q}, p, q)$, $\forall p\forall q\, \mathrm{L}^{\sigma}_{3w}(\mathrm{C}, \mathrm{Q}, p, q)$, $\forall p\forall q\, \mathrm{L}^{\sigma}_{4s}(\mathrm{C}', \mathrm{Q}, p, q)$, $\mathrm{V}\forall p\forall q\, \mathrm{L}^{\sigma}_{4w}(\mathrm{C}', \mathrm{Q}, p, q)$, где $\sigma\in\{+, -\}$.

Ранее в этой статье была отмечена возможность использования АКП$^{(0)}$, содержащихся в КАТ, если в Σ' содержатся факты с истинностным значением $\bar{\nu} = \langle 0, 0\rangle$[4].

АКП$^{(0)}$: $\forall \mathrm{Z}\forall \mathrm{U}((J_{\langle 0,0\rangle}(\mathrm{Z}\Rightarrow_1\mathrm{U})\rightarrow\exists n\, \Pi^{0}_{n}(\mathrm{Z}, \mathrm{U}))$, а

$$\rho^{0}(i) = \frac{|\,\widetilde{\text{БФ}}^{0}_{i}\,|}{|\,\text{БФ}^{0}_{i}\,|}.$$

Для этого случая в КАТ явно определимы предикаты эмпирического конфликта и эмпирической тенденции к конфликту L^{0}_{1}(Z, U, p, q) и L^{0}_{3}(Z, U, p, q). Соответственно определимы основания эмпирических конфликтов и эмпирических тенденций посредством L^{0}_{2}(V, W, p, q) и L^{0}_{4} (V, W, p, q).

L^{0}_{1}(Z, U, p, q) и L^{0}_{2}(V, W, p, q) определяются ниже.

Df. 5-1. $\forall \mathrm{Z}\forall \mathrm{U}\forall p\forall q(\mathrm{L}^{0}_{1}(\mathrm{Z}, \mathrm{U}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{Z}\Rightarrow_1\mathrm{U})\&\ \Pi^{0}_{n}(\mathrm{Z}, \mathrm{U})\rightarrow J_{\langle 0,n+1\rangle}(\mathrm{Z}\Rightarrow_1\mathrm{U}))\ \&\ \overline{K}(p, q)\&\ (\rho^{0}\le\rho^{0}(s)\le 1)))$.

Df. 5-2. $\forall \mathrm{V}\forall \mathrm{W}\forall p\forall q(\mathrm{L}^{0}_{2}(\mathrm{V}, \mathrm{W}, p, q) \leftrightarrow \exists n\exists s(((J_{(\tau,n)}(\mathrm{V}\Rightarrow_2\mathrm{Q})\&\ \mathrm{M}^{+}_{x,n}(\mathrm{V},\mathrm{Q})\&\mathrm{M}^{-}_{y,n}(\mathrm{V},\mathrm{Q}))\rightarrow J_{\langle 0,n+1\rangle}(\mathrm{V}\Rightarrow_2\mathrm{Q}))\&\ \overline{R}(p, q)\&(\rho^{0}\le\rho^{0}(s)\le 1)))$.

---

[4] ИС-ДСМ с КАТ такого типа применимы для анализа социологических данных, в которых содержатся различного рода конфликты [11].

Явные определения предикатов $L^0_{1s}(Z, U, p, q)$, $L^0_{1w}(Z, U, p, q)$; $L^0_{2s}(V, W, p, q)$, $L^0_{2w}(V, W, p, q)$; $L^0_{3s}(Z, U, p, q)$, $L^0_{3w}(Z, U, p, q)$; $L^0_{4s}(V, W, p, q)$, $L^0_{4w}(V, W, p, q)$ формулируются аналогично рассмотренным выше случаям для предикатов L^σ_{is}, L^σ_{iw}, где $i = 1, 2, 3, 4$, а $\sigma \in \{+, -\}$.

Эмпирический конфликт и его основание представимы, соответственно, формулами

(6-1) $\exists Z\forall p\forall q\, L^0_1(Z, Q, p, q)$,

(6-2) $\exists V\forall p\forall q\, L^0_2(V, Q, p, q)$.

Их реализациями являются $\forall p\forall q\, L^0_1(C, Q, p, q)$ и $\forall p\forall q\, L^0_2(C', Q, p, q)$.

ИС-ДСМ с соответствующими $БФ_0 \subset БФ_1 \subset \ldots \subset БФ_s$ предикатами $L^0_1(Z, U, p, q)$ и $L^0_2(V, W, p, q)$ и $АКП^{(0)}$ в КАТ являются средством поддержки анализа конфликтов [12].

Введем еще одно важное для ДСМ-метода понятие **спецификации эмпирической закономерности**. Рассмотрим случай спецификации эмпирических законов – позитивного и негативного, соответственно. Будем говорить, что имеют место спецификации эмпирических законов, если истинны следующие утверждения

(1) $\forall Z\forall p\forall q((L^+_1(Z, Q, p, q)\&\neg L^+_2(C', Q, p, q)) \to (C' \subset Z))$,

(2) $\forall Z\forall p\forall q(L^-_1(Z, Q_0, p, q)\&\neg L^-_2(C'', Q_0, p, q) \to (C'' \subset Z))$, где C' и C'' – необходимые условия для эффектов Q и Q_0, соответственно.

Пусть истинны дополнительно утверждения (3) – (7), тогда будем говорить, что имеет место согласованная спецификация эмпирических позитивного и негативного законов.

(3) $\forall Z\forall p\forall q((L^+_1(Z, Q, p, q)\&(C' \subset Z)) \to \neg (C'' \subset Z))$,

(4) $\forall Z\forall p\forall q((L^-_1(Z, Q_0, p, q)\&(C'' \subset Z)) \to \neg (C' \subset Z))$,

(5) $\forall Z\forall p\forall q((L^+_1(Z, Q, p, q)\&(C' \subset Z)) \to \neg L^+_1(Z, Q_0, p, q))$,

(6) $\forall Z\forall p\forall q((L^-_1(Z, Q_0, p, q)\&(C'' \subset Z)) \to \neg L^-_1(Z, Q, p, q))$,

(7) $\exists Z\forall p\forall q(L^+_1(Z, Q, p, q)\&(C' \subset Z)\&\neg L^+_2(C', Q, p, q)))$,

(8) $\exists Z\forall p\forall q(L^-_1(Z, Q_0, p, q)\&(C'' \subset Z)\&\neg L^-_2(C'', Q_0, p, q))$.

Примером согласованной спецификации эмпирических законов, полученных ИС-ДСМ [13] является обнаружение в БФ маркера продолжительности жизни больных с опухолью меланомы после проведенных лечений. Наличие протеина S 100 со значением меньше 0,120 нг/мл (это константа C') влечет продолжительность жизни больше 5 лет (это константа Q).

Наличие же протеина S 100 со значением больше 0,120 нг/мл (это константа C'') влечет продолжительность жизни меньше 5 лет (это константа Q_0).

Этот результат важен тем, что получена необходимая информация для формирования групп риска, требующих особого медицинского наблюдения.

Воспользуемся теперь введенными выше определениями для эмпирических закономерностей (ЭЗК) и охарактеризуем множество всех порожденных гипотез как ординарных, так и «интересных» (неординарных) согласно П. Гаеку и Т. Гавранеку [14].

Для эффекта Q определим множество всех гипотез с носителями Z этого эффекта, порожденных посредством п.п.в.-2 (аналогии) ДСМ-метода АПГ. Сначала определим:

Df. 7-1. $V^+_{1,Q} = \{Z | \exists n J_{\langle 1, n\rangle}(Z \Rightarrow_1 Q)\}$.

Аналогично определим $V^\sigma_{1,Q}$, где $\sigma \in \{+, -\}$. $V^\sigma_{1,Q}$ соответствует множество порожденных гипотез Ω^σ_s на последнем Этапе I_s процесса ДСМ-рассуждения (для $\rho^\sigma \le \rho^\sigma(s) \le 1$), где $\sigma \in \{+, -, 0\}$.

Df. 7-2. $V^+_{2,Q} = \{V | \exists n J_{\langle 1, n\rangle}(V \Rightarrow_2 Q)\}$.

Аналогично определим $V^\sigma_{2,Q}$, где $\sigma \in \{+, -\}$. $V^\sigma_{2,Q}$ соответствует множество порожденных гипотез Δ^σ_s на последнем Этапе I_s процесса ДСМ-рассуждения (для $\rho^\sigma \le \rho^\sigma(s) \le 1$), где $\sigma \in \{+, -, 0\}$.

Df. 8-1. $\tilde{V}^+_{1,Q} = \{Z | \forall p\forall q\, L^+_1(Z, Q, p, q)\}$.

Аналогично определим $\tilde{V}^\sigma_{1,Q}$, где $\sigma \in \{+, -\}$.

$\tilde{V}^\sigma_{1,Q}$ соответствует множество гипотез $\tilde{\Omega}^\sigma_s$, где $\sigma \in \{+, -, 0\}$.

Df. 8-2. $\tilde{V}^+_{2,Q} = \{V | \forall p\forall q\, L^+_2(V, Q, p, q)\}$.

Аналогично определим $\widetilde{V}^{\sigma}_{2,Q}$, где $\sigma\in\{+,-\}$. $\widetilde{V}^{\sigma}_{2,Q}$ соответствует множество гипотез $\widetilde{\Delta}^{\sigma}_{s}$, где $\sigma\in\{+,-,0\}$.

Очевидно, что имеют место: $\widetilde{V}^{\sigma}_{1,Q}\subseteq V^{\sigma}_{1,Q}$ и $\widetilde{\Omega}^{\sigma}_{s}\subseteq\Omega^{\sigma}_{s}$, где $\sigma\in\{+,-,0\}$. $\widetilde{V}^{\sigma}_{2,Q}\subseteq V^{\sigma}_{2,Q}$ и $\widetilde{\Delta}^{\sigma}_{s}\subseteq\Delta^{\sigma}_{s}$, где $\sigma\in\{+,-,0\}$.

$\widetilde{\Delta}^{\sigma}_{s}$ и $\widetilde{\Omega}^{\sigma}_{s}$ являются «интересными» гипотезами согласно [14] такими, что они представляют весьма информативный результат knowledge discovery посредством ДСМ – метода АПГ и его компоненты ИС – ДСМ.

Следует отметить, что $\widetilde{\Delta}^{\sigma}_{s}$ и $\widetilde{\Omega}^{\sigma}_{s}$ сформированы относительно $БФ_0\subset БФ_1\subset\ldots\subset БФ_s$ посредством выбранной стратегии ДСМ-рассуждения $Str_{x,y}$ [1, 2, 6].

3. Квазиаксиоматические теории, эмпирические закономерности и knowledge discovery

Квазиаксиоматические теории (КАТ) $\mathcal{T} = \langle\Sigma, \Sigma', \mathcal{R}\rangle$ являются формализациями открытых теорий с правилами правдоподобного вывода [1, 2, 5-7]. КАТ являются основой методологии формирования баз знаний ИС-ДСМ.

Так как КАТ являются открытыми теориями, соответствующими процессу ДСМ – рассуждения, то КАТ представляет собой последовательность состояний $\mathcal{T}_0$, $\mathcal{T}_1$, …, $\mathcal{T}_s$, порождаемых этим процессом, где $\mathcal{T}_i = \langle\Sigma_i, \Sigma'_i, \mathcal{R}\rangle$, $i = 0, 1, \ldots, s$. Состояние КАТ $\mathcal{T}_0 = \langle\Sigma_0, \Sigma'_0, \mathcal{R}_0\rangle$ является **начальным**.

Эти состояния КАТ, образованные парами Σ_i, Σ'_i, соответствующими исполнениям Этапов I_p, где p=1,…, s, завершаются Этапом I_s, где имеет место $\rho^{\sigma}\leq\rho^{\sigma}(s)\leq 1$, т.е. осуществляется абдуктивная сходимость ДСМ-рассуждения. При этом Σ'_s содержит гипотезы из $\widetilde{\Delta}^{\sigma}_{s}$ и $\widetilde{\Omega}^{\sigma}_{s}$, а Σ_s расширяет Σ_0 из начального состояния посредством порожденных эмпирических закономерностей (ЭЗК) – эмпирических законов или тенденций (сильных или слабых). Таким образом, имеется последовательность состояний КАТ $\mathcal{T}_0$, $\mathcal{T}_1$, …, $\mathcal{T}_s$, где $\mathcal{T}_s = \langle\Sigma_s, \Sigma'_s, \mathcal{R}\rangle$ есть завершенный результат *knowledge discovery* относительно $БФ_0\subset БФ_1\subset\ldots\subset БФ_s$ и выбранной стратегии ДСМ-рассуждения $Str_{x,y}$ [1, 2, 6].

Эта относительность порождения ЭЗК, связанная не только с отдельными $Str_{x,y}$, но и с множеством возможных стратегий $\overline{Str}$ и с выбранным отношением частичного порядка $\succcurlyeq$ на множестве $\overline{Str}$. В силу конечности $\overline{Str}$ в нем существуют как максимальные $Str_{x,y}$, так и минимальные. Естественно, применять $max\ Str_{x,y}$ в целях обеспечения **точности** порождаемых гипотез, а также применять $min\ Str_{x,y}$ в целях получения **полноты** порождаемых гипотез. В связи с этим возникает необходимость выбора **оптимальных** $Str_{x,y}$, а также разработки **приближенных** $Str_{x,y}$ ДСМ-метода АПГ для функционирования ИС-ДСМ в реальное время относительно больших баз фактов.

Отметим также операциональность определений ЭЗК [2], аналогичную операциональности определений физических понятий согласно В.У. Бриджмену [15, 16]. В самом деле, предикаты $\overline{K}(p, q)$ и $\overline{R}(p, q)$ и термы $\rho^{\sigma}(s)$, где $\sigma\in\{+,-,0\}$, определены конструктивно посредством соответствующих процедур.

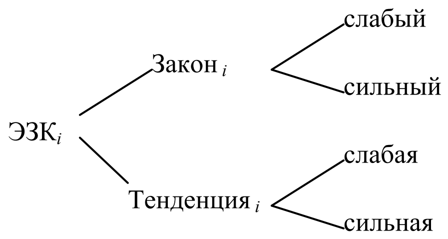
Весьма полезно обнаружить $V^{\sigma}_{2,Q}$, где $\sigma\in\{+,-\}$, такие, что $|V^{\sigma}_{2,Q}| = 1$. Это означает, что реализуется $Str_{x,y}$ с $e^{+}\in x$, $e^{-}\in y$, где e^{σ} – условие единственности причин для предикатов $M^{+}_{x,n}$ и $M^{-}_{y,n}$, соответственно [1].

Предметная область с единственными причинами может быть названа «миром Э. Дюркгейма», ибо он считал, что у социальных событий имеются **единственные** причины [17].

Следует рассмотреть общий случай предметной области такой, что в ней имеется характерное для нее множество эффектов $Q_1,\ldots, Q_r$. Тогда *knowledge discovery*, реализованное ИС-ДСМ характеризуется множествами $\widetilde{V}^{+}_{1,Q_i}$, $\widetilde{V}^{+}_{2,Q_i}$ и $\widetilde{\Omega}^{\sigma}_{i}$, $\widetilde{\Delta}^{\sigma}_{i}$, где $\sigma\in\{+,-,0\}$, а $i = 1, \ldots, r$.

Таким образом, в результате применения ИС-ДСМ к предметной области с эффектами Q_1, …, Q_r порождается множество ЭЗК = {$ЭЗК_1$, …, $ЭЗК_r$}, такое, что ($i = 1, \ldots, r$).

Эта схема выражает структуру результатов knowledge discovery относительно $БФ_0\subset БФ_1\subset\ldots\subset БФ_s$ и $Str_{x,y}$ посредством ИС-ДСМ и ДСМ- метода АПГ.



Литература

1. Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта. Часть I // Искусственный интеллект и принятие решений. №3, 2010, стр.3-21.
2. Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта. Часть II // Искусственный интеллект и принятие решений. №4, 2010.
3. Милль Д.С. Система логики силлогистической и индуктивной. – М.: ЛЕНАНД (УРСС), 2011.
4. Шрейдер Ю.А. Равенство, сходство, порядок. – М.: Издательство «Наука», Главная редакция физико-математической литературы, 1971.
5. Арский Ю.М., Финн В.К. Принципы конструирования интеллектуальных систем // Информационные технологии и вычислительные системы, №4, 2008, с.4-37.
6. Финн В.К. Индуктивный метод соединенного сходства – различия и процедурная семантика ДСМ – метода // Научно-техническая информация, Сер.2, №4, 2010, стр.1-17.
7. Аншаков О.М., Скворцов Д.П., Финн В.К. О дедуктивной имитации некоторых вариантов ДСМ-метода автоматического порождения гипотез // В кн.: ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания. Под общ. ред. О.М. Аншакова – М.: Книжный дом «Либроком», 2009, с.240-286.
8. Автоматическое порождение гипотез в интеллектуальных системах. Под общ. ред. В.К. Финна. – М.: Книжный дом «Либроком», 2009.
9. Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. Advances in Knowledge Discovery and Data Mining. – The AAAI Press, 1996.
10. De Bouvère K.L. A method in proofs of undefinability. – North-Holland Publ. Co., Amsterdam, 1950.
11. Финн В.К., Михеенкова М.А. О логических средствах концептуализации анализа мнений // Научно-техническая информация, сер.2, №6, 2002, стр.4-22.
12. Здравомыслов А.Г. Социология конфликта. – М.: АСПЕКТ ПРЕСС, 1995.
13. Н.Н. Михайлова, Е.С. Панкратова и др. О применении интеллектуальной компьютерной системы для анализа клинических данных больных меланомой // Российский биотерапевтический журнал, Том 9, №2, 2010, с.54.
14. Гаек П., Гавранек Т. Автоматическое образование гипотез. – М.: «Наука», Главная редакция физико-математической литературы, 1984.
Hájek P., Havránek, T. Mechanizing Hypothesis Formation. Mathematical Foundations for a General Theory. Berlin – Heidelberg – New York, 1978.
15. Bridgman P.W. The Logic of Modern Physics, N.Y., 1927.
16. Bridgman P.W. The Nature of Some our Physical Concepts, N.Y., 1952.
17. Дюркгейм Э. Метод социологии. – М.: «Наука», 1991, - Гл. 6: Правила, касающиеся доказательств, с.511-527. Durkheim E. Les règles de la Méthode de sociologie. Paris: Les Press universitaires de France, 16 édition, 1967.

Финн Виктор Константинович. Заведующий Сектором интеллектуальных систем Отдела теоретических и прикладных проблем информатики Всероссийского института научной и технической информации РАН (ВИНИТИ РАН). В 1957 году окончил Философский факультет, в 1966 году – Механико-математический факультет Московского государственного университета им. М.В. Ломоносова. Доктор технических наук, профессор, Заслуженный деятель науки РФ. Имеет 194 печатные работы. Область научных интересов: искусственный интеллект, математическая логика, интеллектуальные системы, машинное обучение, теория познания, логика и методология науки. E-mail: finn@viniti.ru.