УДК 512.58

Д.В. Виноградов

Проекции полурешеток: язык теории категорий*

Дается описание теоретико-категорного подхода к проекциям полурешеток, альтернативного классическому походу через проекции узорных структур. В частном случае декартовых произведений и их проекций на подпроизведения, этот подход образует фундамент для последовательного варианта ВКФ-метода машинного обучения, основанного на бинарной операции сходства.

Ключевые слова: *полурешетка, монада, категория алгебр, предел функтора*

DOI: 10.36535/0548-0027-2021-06-4

ВВЕДЕНИЕ

Оригинальный метод машинного обучения, основанный на использовании анализа формальных понятий (АФП) [1] – современном разделе теории решеток, разработан нами [2] в 2012-2020 гг. Он получил название ВКФ-метод в честь профессора В.К. Финна, многие идеи которого об индуктивной логике Д.С. Милля [3] были использованы при его создании.

Однако осталась некоторая незавершенность – в оригинальном ВКФ-методе предполагается, что все признаки, описывающие объекты, наблюдаются одновременно. Это допущение вряд ли следует признать адекватным практике, когда лицо, принимающее решение получает информацию последовательно некоторыми блоками. Многие альтернативные парадигмы – деревья решений, рекуррентные нейросети, вероятностные автоматы – учитывают такую последовательность.

В настоящей работе будет описано теоретико-категорное представление для последовательного варианта ВКФ-метода, когда полурешетка описаний объектов отображается в прямое произведение полурешеток, соответствующих значениям признаков, описывающих обучающие примеры и порождаемые из них сходства.

Более обще можно рассматривать последовательность расширяющихся описаний объектов и гипотез с гомоморфизмами между полурешетками таких описаний.

Последовательный вариант ВКФ-метода допускает цепочку расширений описаний обучающих примеров, контр-примеров и гипотез с помощью добавления дополнительных признаков. Но можно рассмотреть и более общий вариант, когда имеется семейство полурешеток с гомоморфизмами между ними. В более узком случае это обычные проекции из декартова произведения полурешеток в некоторые (или даже во все) его подпроизведения.

Первой работой, в которой вопрос расширения (или сужения) полурешеток для интеллектуального анализа данных, основанного на операции сходства, был рассмотрен, является статья Б. Гантера и С.О. Кузнецова [4]. К сожалению, эта работа содержит большое число неверных утверждений (Proposition 1, Theorem 2). Ложность Proposition 1 была обнаружена М.В. Самохиным [5]. Наиболее простой контр-пример был построен А.В. Бузмаковым [6]. Для устранения этой проблемы авторами [7] были наложены различные дополнительные условия на определение проекции узорных структур. Нам они кажутся неестественными. Ложность ключевой Theorem 2 была установлена T.B. Kaiser и S.E. Schmidt [8], причем ложность этой теоремы сохраняется и для дополненных определений проекций.

Мы полагаем, что теоретико-категорный язык, опирающийся на классический учебник [9] и представленный далее, гораздо адекватнее для описания структур полурешеток и их проекций как декартово-замкнутой категории алгебр над соответствующей монадой.

ПОЛУРЕШЕТКИ КАК АЛГЕБРЫ НАД МОНАДОЙ

Локальное сходство является бинарной операцией на множестве X, объемлющем множество объектов, т. е. представляет собой отображение $\wedge: X \times X \longrightarrow X$. Элементы множества X будем называть *фрагментами*.

Для независимости результата нахождения сходства нескольких объектов операция сходства должна удовлетворять аксиомам *полурешетки*: тождествам ассоциативности, идемпотентности и коммутативности.

Очевидно, что в таком случае можно определить *глобальную операцию сходства*, которая по любому подмножеству $S \subseteq X$ порождает наибольшую нижнюю грань $\wedge S \in X$. Другими словами, имеется отображение $\wedge: \mathcal{P}X \to X$, где $\mathcal{P}X$ – *множество-степень* (= множество всех подмножеств) для X.

* Работа выполнена при частичной поддержке гранта РФФИ № 18-29-03063мк

Для представления проекций между полурешетками удобно использовать теоретико-категорные понятия [9].

Напомним, что категория $\mathbf{C}$ состоит из семейства *объектов* и семейства *стрелок*, причем каждому объекту $c \in \mathbf{C}$ сопоставляется стрелка $1_c: c \to c \in \mathbf{C}$, а каждой паре стрелок $f: a \to b, g: b \to c \in \mathbf{C}$ сопоставляется стрелка $g \cdot f: a \to c \in \mathbf{C}$, причем $f \cdot 1_a = f, 1_b \cdot f = f$ и композиция ассоциативна $h \cdot (g \cdot f) = (h \cdot g) \cdot f$, если определена в $\mathbf{C}$, т. е. коммутативны диаграммы

$$\begin{tikzcd} a \arrow[r, "f"] \arrow[d, "g\cdot f"'] & b \arrow[d, "h\cdot g"] \arrow[dl, "g"'] \\ c \arrow[r, "h"] & d \end{tikzcd} \qquad \begin{tikzcd} a \arrow[r, "1_a"] \arrow[dr, "f"'] & a \arrow[d, "f"] \arrow[dr, "f"] & \\ & b \arrow[r, "1_b"] & b \end{tikzcd}$$

Диаграмма называется *коммутативной*, если все композиции стрелок с общим началом и общим концом равны между собой.

Напомним, что *функтор* $T: \mathbf{J} \to \mathbf{C}$ из категории $\mathbf{J}$ в категорию $\mathbf{C}$ состоит из *функции объектов* T, которая каждому объекту $j \in \mathbf{J}$ сопоставляет объект $Tj \in \mathbf{C}$, и *функции стрелок* T, которая каждой стрелке $u: j \to i \in \mathbf{J}$ сопоставляет стрелку $Tu: Tj \to Ti \in \mathbf{C}$, причем $T(1_j) = 1_{Tj}, T(v \cdot u) = Tv \cdot Tu$ (каждый раз, когда композиция $v \cdot u$ определена в $\mathbf{J}$).

Категория $\mathbf{Set}$ множеств и отображений между ними допускает эндофунктор $\mathcal{P}: \mathbf{Set} \to \mathbf{Set}$, который отображает множество X во множество-степень $\mathcal{P}X = \{A | A \subseteq X\}$, а отображение $f: X \to Y$ в $\mathcal{P}f: \mathcal{P}X \to \mathcal{P}Y$, где $\mathcal{P}f(A) = \{f(x) | x \in A\} \subseteq Y$, для любого $A \subseteq X$.

В дальнейшем нам встретится *забывающий функтор* $F: \mathbf{Lat} \to \mathbf{Set}$, который отображает полурешетку $\langle X, \wedge \rangle$ в ее носитель X, а гомоморфизм полурешеток $f: \langle X_1, \wedge \rangle \to \langle X_2, \wedge \rangle$ в отображение $Ff: X_1 \to X_2$.

Напомним, что *естественное преобразование* $\tau: R \to T$ функтора $R: \mathbf{J} \to \mathbf{C}$ в функтор $T: \mathbf{J} \to \mathbf{C}$ – это функция, которая каждому объекту $j \in \mathbf{J}$ сопоставляет стрелку $\tau_j: Rj \to Tj \in \mathbf{C}$ таким образом, что для каждой стрелки $u: j \to i \in \mathbf{J}$ следующая диаграмма коммутативна:

$$\begin{tikzcd} j \arrow[d, "u"] \\ i \end{tikzcd} \qquad \begin{tikzcd} Rj \arrow[r, "\tau_j"] \arrow[d, "Ru"'] & Tj \arrow[d, "Tu"] \\ Ri \arrow[r, "\tau_i"] & Ti \end{tikzcd}$$

Существует естественное преобразование $\eta: I_{\mathbf{Set}} \to \mathcal{P}$ тождественного функтора $I_{\mathbf{Set}}$ в функтор $\mathcal{P}$ с компонентами $\eta_X: X \to \mathcal{P}X$, которые отображают каждый $x \in X$ в одноэлементное подмножество $\eta_X(x) = \{x\} \in \mathcal{P}X$.

Имеется также естественное преобразование $\cup: \mathcal{P}\mathcal{P} \to \mathcal{P}$ с компонентами $\cup_X: \mathcal{P}\mathcal{P}X \to \mathcal{P}X$, которые отображают каждое семейство $S \subseteq \mathcal{P}X$ подмножеств в их объединение $\cup_X (S) = \cup \{A | A \in S\} \in \mathcal{P}X$.

Каждый эндофунктор $T: \mathbf{C} \to \mathbf{C}$ определяет композиции $T^2 = T \cdot T: \mathbf{C} \to \mathbf{C}$ и $T^3 = T^2 \cdot T: \mathbf{C} \to \mathbf{C}$. Пусть $\mu: T^2 \to T$ – естественное преобразование с компонентами $\mu_c: T^2 c \to Tc$ для каждого $c \in \mathbf{C}$. Тогда $T\mu: T^3 \to T^2$ обозначает естественное преобразование с компонентами $(T\mu)_c = T(\mu_c): T^3 c \to T^2 c$, а преобразование $\mu T: T^3 \to T^2$ имеет компоненты $(\mu T)_c = \mu_{Tc}: T^3 c \to T^2 c$.

Напомним, что *монада* $T = \langle T, \eta, \mu \rangle$ в категории $\mathbf{C}$ состоит из функтора $T: \mathbf{C} \to \mathbf{C}$ и двух естественных преобразований $\eta: I_{\mathbf{C}} \to T$ и $\mu: T^2 \to T$, делающих коммутативными следующие диаграммы:

$$\begin{tikzcd} T^3 \arrow[r, "T\mu"] \arrow[d, "\mu T"'] & T^2 \arrow[d, "\mu"] \\ T^2 \arrow[r, "\mu"] & T \end{tikzcd} \qquad \begin{tikzcd} I_{\mathbf{C}}T \arrow[r, "\eta T"] \arrow[dr, "id"'] & T^2 \arrow[d, "\mu"] & TI_{\mathbf{C}} \arrow[l, "T\eta"'] \arrow[dl, "id"] \\ & T & \end{tikzcd}$$

Легко проверить, что тройка $\langle \mathcal{P}, \eta, \cup \rangle$ задает монаду в категории $\mathbf{Set}$. Необходимые тождества:

$$\cup \cdot (\mathcal{P} \cup) = \cup \cdot (\cup \mathcal{P}): \mathcal{P}^3 \to \mathcal{P} \text{ и}$$

$$\cup \cdot (\eta \mathcal{P}) = id = \cup \cdot (\mathcal{P}\eta): \mathcal{P} \to \mathcal{P}$$

представляют собой равенство

$$\cup_{i \in \{I_j | j \in J\}} S_i = \cup_{j \in J} \left\{ \cup_{i \in I_j} S_i \right\}$$

и равенства

$$\cup \{A | A \subseteq S\} = S = \cup \left\{ \{x\} \middle| x \in S \right\},$$

соответственно.

Для монады $\langle \mathcal{P}, \eta, \cup \rangle$ в категории $\mathbf{Set}$ можно определить категорию *алгебр* $\mathbf{Set}^{\mathcal{P}}$ как множество пар $\langle X, \wedge \rangle$, где объект (множество) X называется *носителем алгебры*, а морфизм $\wedge: \mathcal{P}X \to X$ называется *структурным отображением*, причем должны выполняться тождества $\wedge \cdot \mathcal{P}\wedge = \wedge \cdot \cup_X: \mathcal{P}^2 X \to X$ и $\wedge \cdot \eta_X = id_X: X \to X$.

$$\begin{tikzcd} \mathcal{P}^2(X) \arrow[r, "\mathcal{P}\wedge"] \arrow[d, "\cup_X"'] & \mathcal{P}(X) \arrow[d, "\wedge"] \\ \mathcal{P}(X) \arrow[r, "\wedge"] & X \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[dr, "id"'] & \mathcal{P}(X) \arrow[d, "\wedge"] \\ & X \end{tikzcd}$$

<u>**Лемма 1.**</u> *Класс алгебр $\langle X, \wedge \rangle$ над монадой $\langle \mathcal{P}, \eta, \cup \rangle$ в категории* $\mathbf{Set}$ *совпадает с полными полурешетками.*

<u>*Доказательство*</u>. Структурное отображение $\wedge: \mathcal{P}X \to X$ задает частичный порядок $x \leq y \Leftrightarrow [\wedge \{x, y\} = x]$. Антисимметричность: $x \leq y$ и $y \leq x$ влекут $x = \wedge \{x, y\} = \wedge \{y, x\} = y$. Рефлексивность: $\wedge \{x\} = \wedge \cdot \eta_X(x) = x$. Транзитивность: $x \leq y$ и $y \leq z$ влекут

$$\wedge \{x, z\} = \wedge \{\wedge \{x, y\}, z\} = \wedge \left\{ \{x, y\} \cup \{z\} \right\} =$$

$$\wedge \left\{ \{x\} \cup \{y, z\} \right\} = \wedge \left\{ x, \wedge \{y, z\} \right\} = \wedge \{x, y\} = x.$$

Покажем, что $\wedge S$ – точная нижняя грань для $S \subseteq X$. Для всякого $x \in S$ верно $S \cup \{x\} = S$. Тождество $\wedge \cdot \mathcal{P}\wedge = \wedge \cdot \cup_X$ влечет $\wedge \{\wedge S, x\} = \wedge S$, что означает $\wedge S \leq x$. Пусть для всех $x \in S$ выполняется $\wedge \{y, x\} = y$ (т.е. $y \leq x$).

Тогда

$$\wedge\{\wedge S, y\} = \wedge(S \cup \{y\}) = \wedge\left(\cup\left\{\{x, y\} \middle| x \in S\right\}\right) = \wedge\{y\} = y.$$

То, что полная полурешетка является алгеброй над монадой $\langle \mathcal{P}, \eta, \cup \rangle$, легко проверяется.

ПРЕДЕЛЫ ПОЛУРЕШЕТОК

Обозначим категорию алгебр над монадой $\langle \mathcal{P}, \eta, \cup \rangle$ в категории $\mathbf{Set}$ через $\mathbf{Lat}$. Эта категория является *декартово-замкнутой*, т. е. в ней существуют пределы малых функторов, определение которых мы сейчас напомним.

Диагональный функтор $\Delta: \mathbf{C} \to \mathbf{C}^{\mathbf{J}}$ отображает каждый объект $c \in \mathbf{C}$ в постоянный функтор $\Delta c: \mathbf{J} \to \mathbf{C}$, который на любом объекте $j \in \mathbf{J}$ принимает значение c, а на любой стрелке из $\mathbf{J}$ – значение 1_c. Если $f: a \to c$ – некоторая стрелка из $\mathbf{C}$, то Δf – естественное преобразование $\Delta f: \Delta a \to \Delta c$, которое на любом объекте $j \in \mathbf{J}$ принимает значение $f: a \to c$.

$$\mathbf{J}: \quad \begin{tikzcd} i & j \arrow[l] \arrow[r] & k \arrow[r, shift left] \arrow[r, shift right] & l \end{tikzcd}$$

$$\mathbf{C}: \quad \begin{tikzcd} a \arrow[d, "f"] & a \arrow[d, "f"] & a \arrow[l, "1_a"'] \arrow[r, "1_a"] \arrow[d, "f"] & a \arrow[r, shift left, "1_a"] \arrow[r, shift right, "1_a"'] \arrow[d, "f"] & a \arrow[d, "f"] \\ c & c & c \arrow[l, "1_c"'] \arrow[r, "1_c"] & c \arrow[r, shift left, "1_c"] \arrow[r, shift right, "1_c"'] & c \end{tikzcd}$$

Назовем *конусом* с основанием F и вершиной $c \in \mathbf{C}$ естественное преобразование $\tau: \Delta c \to F$ из постоянного функтора Δc в некоторый функтор $F: \mathbf{J} \to \mathbf{C}$.

$$\mathbf{J}: \quad \begin{tikzcd} i & j \arrow[l, "u"'] \arrow[r] & k \arrow[r, shift left] \arrow[r, shift right] & l \end{tikzcd}$$

$$\begin{array}{l} \mathbf{C}: \\ \\ \mathbf{C}: \end{array} \quad \begin{tikzcd} c \arrow[d, "\tau_i"] & c \arrow[l, "1_c"'] \arrow[r, "1_c"] \arrow[d, "\tau_j"] & c \arrow[r, shift left, "1_c"] \arrow[r, shift right, "1_c"'] \arrow[d, "\tau_k"] & c \arrow[d, "\tau_l"] \\ F(i) & F(j) \arrow[l, "Fu"'] \arrow[r] & F(k) \arrow[r, shift left] \arrow[r, shift right] & F(l) \end{tikzcd}$$

Поскольку значения функтора $\Delta c: \mathbf{J} \to \mathbf{C}$ сводятся к $c \in \mathbf{C}$, то естественное преобразование τ для каждого объекта $j \in \mathbf{J}$ состоит из такой стрелки $\tau_j: c \to F(j)$, что для любой стрелки $u: j \to i \in \mathbf{J}$ коммутативна диаграмма

$$\begin{tikzcd} & c \arrow[dl, "\tau_i"'] \arrow[d, "\tau_j"] \arrow[dr, "\tau_k"] & & \\ F(i) & F(j) \arrow[l, "Fu"'] \arrow[r] & F(k) & \ldots \end{tikzcd}$$

Предел функтора $F: \mathbf{J} \to \mathbf{C}$ – это универсальная стрелка (= естественное преобразование) $\langle r, \nu \rangle$ из Δ в F. Объект $r \in \mathbf{C}$ обычно обозначается $r = \lim F$.

Естественное преобразование $\nu: \Delta r \to F$ универсально среди естественных преобразований $\tau: \Delta c \to F$, где $c \in \mathbf{C}$.

Универсальное свойство преобразования ν состоит в следующем: это конус с основанием F и вершиной $\lim F$; для любого конуса τ с основанием F и вершиной c существует единственная такая стрелка $t: c \to \lim F$, что $\tau_j = \nu_j \cdot t$.

$$\begin{tikzcd} c \arrow[r, dashed, "t"] \arrow[d, "\tau_j"'] \arrow[dr, "\tau_i" near end] & \lim F \arrow[d, "\nu_i"] \arrow[dl, "\nu_j"' near start] \\ F(i) & F(j) \arrow[l, "Fu"] \end{tikzcd}$$

Примером предела является *декартово произведение*, где категория $\mathbf{J} = \langle \{1,2\}, \emptyset \rangle$ содержит два объекта и пустое множество стрелок.

Тогда имеем универсальное свойство декартова произведения на языке теории категорий:

$$\begin{tikzcd} & L \arrow[dl, "f_1"'] \arrow[d, dashed, "\langle f_1, f_2 \rangle"] \arrow[dr, "f_2"] & \\ L_1 & L_1 \times L_2 \arrow[l, "\pi_1"] \arrow[r, "\pi_2"'] & L_2 \end{tikzcd}$$

Так как $\mathbf{Lat} = \mathbf{Set}^{\mathcal{P}}$, то предел $\lim F$ будет состоять из *нитей* – таких элементов декартова произведения $\alpha \in \prod_{j \in \mathbf{J}} L_j$ с координатами α_j, что $Fu(\alpha_j) = \alpha_i$ для всех $u: j \to i \in \mathbf{J}$. Операция сходства $\wedge$ на нитях является ограничением операции сходства из декартова произведения, т. е. вычисляется покомпонентно.

Стандартным способом описания обучающих объектов и порождаемых из них кандидатов в гипотезы в ВКФ-методе является набор признаков. Если значения признака дискретны, то их можно закодировать с помощью битовых строк с побитовым умножением в качестве бинарной операции сходства между ними, используя технику анализа формальных понятий [1]. В [10] разработан алгоритм такого кодирования и доказана его корректность.

Случай признаков с непрерывными значениями был исследован в [11] и предложен алгоритм кодирования битовыми строками с побитовым умножением как сходством на основании теоретико-информационного подхода. Попутно был исследован вопрос об обнаружении зависимостей (логического типа) между различными непрерывными признаками на основе логистической регрессии.

Для семейства $\{L_j = \langle X_j, \wedge_j \rangle | j \in J\}$ категории $\mathbf{Lat}$, индексированного некоторым множеством J, рассмотрим диаграмму всех проекций $\pi_U^V: \prod_{j \in V} L_j \to \prod_{j \in U} L_j$, где $U \subseteq V \subseteq J$. Пределом этой диаграммы можно считать $\prod_{j \in J} L_j$. Полурешетка L_j представляет собой значения j-го признака. При этом операция сходства вычисляется покомпонентно.

$$\begin{tikzcd} L_1 & L_1 \times L_2 \arrow[l, "\pi_1^{1,2}"'] & (L_1 \times L_2) \times L_3 \arrow[l, "\pi_{1,2}^{1,2,3}"'] & \ldots \arrow[l] \end{tikzcd}$$

В случае проекций между декартовыми произведениями полурешеток, нить должна содержать последовательность элементов декартовых произведений, т. е. многократное дублирование координат, но, устраняя повторы, можно отождествить предел с $\prod_{j \in J} L_j$ с покоординатными операциями сходства.

ОПРЕДЕЛЕНИЕ ГЛОБАЛЬНОГО СХОДСТВА

Для определения *глобального сходства* (или, в терминологии анализа формальных понятий [1], *поляры*), лежащего в основе ДСМ-метода поддержки научных исследований [3] и ВКФ-метода машинного обучения [2], нам потребуется установить свободность алгебры $\langle \mathcal{P}S, \cup \rangle$ на множестве S обучающих примеров через сопряжение забывающего и порождающего функторов.

Напомним, что *сопряжение* между категориями $\mathbf{C}$ и $\mathbf{B}$ – это тройка $\langle F, G, \phi \rangle : \mathbf{C} \rightharpoonup \mathbf{B}$, где $F: \mathbf{B} \to \mathbf{C}$ и $G: \mathbf{C} \to \mathbf{B}$ – функторы, а биекция ϕ каждой стрелке $h: Gc \to b$ сопоставляет стрелку $\phi h: c \to Fb$, сопряженную к h справа, причем для всех стрелок $f: c' \to c$ и $g: b \to b'$ выполнены условия естественности

$$\phi(g \cdot h) = Fg \cdot \phi h, \phi(h \cdot Gf) = \phi h \cdot f.$$

Это равносильно естественности преобразования ϕ^{-1}, то есть для всех $f: c' \to c$, $g: b \to b'$ и $k: c \to Fb$ выполняются

$$\phi^{-1}(k \cdot f) = \phi^{-1} k \cdot Gf, \phi^{-1}(Fg \cdot k) = g \cdot \phi^{-1} k.$$

Для любого объекта $c \in \mathbf{C}$ рассмотрим $\eta_c: c \to FGc$ как образ стрелки $1_{Gc}: Gc \to Gc$ при отображении ϕ. Это задает естественное преобразование тождественного функтора $I_{\mathbf{C}}: \mathbf{C} \to \mathbf{C}$ в функтор $FG: \mathbf{C} \to \mathbf{C}$.

Биекция ϕ может быть выражена в терминах стрелок η_c, а именно $\phi h = Fh \cdot \eta_c$ для всех $h: Gc \to b$.

Действительно, ввиду условий естественности имеем

$$\phi(h) = \phi(h \cdot 1_{Gc}) = Fh \cdot \phi 1_{Gc} = Fh \cdot \eta_c.$$

Аналогично, имеется естественное преобразование ε функтора $GF: \mathbf{B} \to \mathbf{B}$ в тождественный функтор $I_{\mathbf{B}}$ как набор стрелок $\varepsilon_b: GFb \to b$ – образов стрелок $1_{Fb}: Fb \to Fb$ при обратной биекции ϕ^{-1}.

Наоборот, биекция ϕ^{-1} может быть выражена в терминах стрелок ε_b по формуле $\phi^{-1} k = \varepsilon_b \cdot Gk$ для всех $k: c \to Fb$.

<u>Теорема 1.</u> Для монады $\langle \mathcal{P}, \eta, \cup \rangle$ в категории $\mathbf{Set}$ имеет место сопряжение $\langle F, G, \eta, \varepsilon \rangle : \mathbf{Set} \rightharpoonup \mathbf{Lat}$, где $F\langle X, \wedge \rangle = X$ – забывающий функтор $F: \mathbf{Lat} \to \mathbf{Set}$, $GX = \langle \mathcal{P}X, \cup_X \rangle$ – порождающий функтор $G: \mathbf{Set} \to \mathbf{Lat}$, $\eta_X: X \to \mathcal{P}X$ – естественное преобразование $I_{\mathbf{Set}} \to FG$\$, а $\varepsilon_{\langle X, \wedge \rangle} = \wedge$ – естественное преобразование $GF \to I_{\mathbf{Lat}}$.

<u>Доказательство</u>. Функтор $F\langle X, \wedge \rangle = X$ забывает структурное отображение $\wedge: \mathcal{P}X \to X$ полурешетки.

Для каждого множества X пара

$$\langle \mathcal{P}X, \cup_X \rangle \ (\cup_X: \mathcal{P}^2 X \to \mathcal{P}X)$$

является свободной полной полурешеткой над X в силу закона ассоциативности и наличия левой и правой единицы в монаде $\langle \mathcal{P}, \eta, \cup \rangle$. Следовательно, соответствие $GX = \langle \mathcal{P}X, \cup_X \rangle$ действительно определяет функтор $G: \mathbf{Set} \to \mathbf{Lat}$, как и утверждалось.

Тогда $FGX = F\langle \mathcal{P}X, \cup_X \rangle = \mathcal{P}X$, поэтому единица $\eta_X: X \to \mathcal{P}X$ данной монады является естественным преобразованием $I_{\mathbf{Set}} \to FG$.

С другой стороны, $GF\langle X, \wedge \rangle = GX = \langle \mathcal{P}X, \cup_X \rangle$. При этом тожество $\bigwedge \cdot \mathcal{P} \bigwedge = \bigwedge \cdot \cup_X: \mathcal{P}^2 X \to X$ означает, что структурное отображение $\wedge: \mathcal{P}X \to X$ является гомоморфизмом полурешеток $GF\langle X, \wedge \rangle = \langle \mathcal{P}X, \cup_X \rangle \to \langle X, \wedge \rangle$.

В итоге получаем естественное преобразование

$$\varepsilon : GF \to I_{\mathbf{Lat}}.$$

Остальные тождества для сопряжения имеют вид $\cup_X \cdot (\mathcal{P}\eta_X) = id_{\mathcal{P}X}: \mathcal{P}X \to \mathcal{P}X$ и $\bigwedge \cdot \eta_X = id_X: X \to X$. Первое совпадает с равенством $S = \cup \left\{ \{x\} \middle| x \in S \right\}$, а второе – с условием $\wedge \cdot \eta_X(x) = x$ рефлексивности порядка, установленного в Лемме 1.

По доказанной теореме сопряжение

$$\langle F, G, \eta, \varepsilon \rangle : \mathbf{Set} \rightharpoonup \mathbf{Lat}$$

определяет изоморфизм

$$\phi: \mathbf{Lat}(\langle \mathcal{P}S, \cup_S \rangle, \langle X, \wedge \rangle) \to \mathbf{Set}(S, X),$$

для которого функция описаний обучающих примеров $f: S \to X$ в решетке $\langle X, \wedge \rangle$ отображается в

$$' = \phi^{-1}(f) = \wedge \cdot \mathcal{P}f: \langle \mathcal{P}S, \cup_S \rangle \to \langle X, \wedge \rangle.$$

Поэтому имеет место отображение $': \mathcal{P}S \to X$, для которого $\{o\}' = \wedge \cdot \eta(o) = f(o)$ для каждого $o \in S$.

Следовательно, каждому подмножеству $A \subseteq S$, называемому *списком родителей*, сопоставляется элемент $A' \in X$, называемый *полярой* или *сходством* списка родителей.

Хотя исходная конструкция совпадает с понятием «узорной структуры» по Гантеру–Кузнецову [4], категорная формализация заменяет понятие проекции между такими структурами на гомоморфизм алгебр над соответствующей монадой, что обеспечивает инвариантность конструкции поляры относительно преобразований имен объектов и полурешеток значений.

Рассмотрим теперь частный случай последовательного ВКФ-метода или, более обще – обогащение описаний обучающих примеров для ДСМ-метода и ВКФ-метода. На множестве S (положительных) обучающих объектов зададим отображения $f_j: S \to FL_j = X_j$, сопоставляющие каждому объекту значение j-го признака в полурешетке $L_j = \langle X_j, \wedge_j \rangle$.

Комбинирование решеток значений признаков можно описать некоторым функтором $F: \mathbf{J} \to \mathbf{Lat}$, где объект $j \in \mathbf{J}$ представляет один из признаков (вообще говоря, сколь угодно сложный), а стрелка $u: j \to i \in \mathbf{J}$ – гомоморфизм полурешеток $Fu: L_j \to L_i$.

Сопряжение $\langle F, G, \eta, \varepsilon \rangle : \mathbf{Set} \rightharpoonup \mathbf{Lat}$ отображает каждое

$$f_j: S \to FL_j = X_j$$

в

$$'_j = \phi^{-1}(f_j) = \wedge \cdot \mathcal{P}f: \langle \mathcal{P}S, \cup_S \rangle \to L_j = \langle X_j, \wedge_j \rangle.$$

По универсальности предела $\langle \lim F, \nu \rangle$ существует единственная стрелка $': \mathcal{P}S \to \lim F$, для которого $'_j = \nu_j \cdot '$ для всех $j \in \mathbf{J}$.

Таким образом, каждому списку $A \subseteq S$ обучающих примеров сопоставляется их сходство $A' \in \lim F$.

Теперь можно, используя классический анализ формальных понятий [1], развивать обобщенный ВКФ-метод.

ЗАКЛЮЧЕНИЕ

В настоящей работе описан теоретико-категорный подход к обобщенному ВКФ-методу машинного обучения, основанного на операции сходства. Эту работу следует рассматривать как фундамент для установления терминологии и общего подхода к методологии применения ВКФ-метода к различным предметным областям.

Одним из достоинств классического ДСМ-метода была возможность его применения к выборкам малого объема. Однако выборки большого объема не могли быть учтены из-за высокой вычислительной (экспоненциальной, в общем случае) сложности алгоритмов порождения сходств. Вероятностный характер алгоритмов ВКФ-метода позволил осуществлять вычисления за полиномиальное время. Но тогда возникает вопрос о надежности порождаемых гипотез.

Развитие средств накопления и обработки больших массивов данных (*Big Data*) дает возможность порождать правдоподобные гипотезы о причинах изучаемых феноменов с большой надежностью. При этом ключевой становится проблема выбора адекватного представления обучающих примеров. При изменении представления возникают отображения соответствующих решеток сходств. Естественность (в смысле теории категорий) этих отображений позволяет сравнивать результаты работы базового алгоритма ВКФ-метода (основанного на АФП [1]), примененного к одной и той же выборке, но с различными представлениями. Впрочем, польза от такого теоретико-категорного описания имеется даже для классического ДСМ-метода.

* * *

Приносим благодарность Л.А. Якимовой за полезные обсуждения, а также своим коллегам по ВЦ им. А.А. Дородницына РАН ФИЦ ИУ РАН и РГГУ (особенно, проф. Е.М. Бениаминову) за поддержку и конструктивные дискуссии. Критические замечания и предложения проф. С.О. Кузнецова были с благодарностью учтены при подготовке окончательной версии работы.

СПИСОК ЛИТЕРАТУРЫ

1. Ganter B., Wille R. Formal Concept Analysis. Transl. from German. – Berlin: Springer–Verlag, 1999. – 284 p.
2. Vinogradov D.V. Machine Learning Based on Similarity Operation // Communications in Computer and Information Science. – 2018. – Vol. 934. – P. 46-59
3. ДСМ-метод автоматического порождения гипотез: логические и эпистемологические основания / ред.: В.К. Финн, О.М. Аншаков. – М.: Эдиториал-УРСС, 2009. – 432 с.
4. Ganter B., Kuznetsov S.O. Pattern Structures and Their Projections // Proc. 9th Conference ICSS 2001, Lecture Notes in Computer Science. – 2001. – Vol. 2120. – P. 129-142
5. Самохин М.В. Машинное обучение на узорных структурах: дис. … канд. техн. наук. – М.: ВИНИТИ РАН, 2006. – 124 с.
6. Бузмаков А.В. Моделирование процессов с состояниями сложной структуры на основе решеток замкнутых описаний; дис. … канд. техн. наук. – М.: НИУ-ВШЭ, 2014. – 154 с.
7. Buzmakov A.V., Kuznetsov S.O., Napoli A. Revisiting Pattern Structure Projections // Proc. International Conference FCA 2015 / eds. J. Baixeries, C. Sacarea, M. Ojeda-Aciego. Lecture Notes in Computer Science – 2015. – Vol. 9113. – P. 200-215
8. Kaiser T.B., Schmidt S.E. Some Remarks on the Relation between Annotated Ordered Sets and Pattern Structures and Their Projections // Proc. Proc. Conference Pattern Recognition and Machine Intelligence. Lecture Notes in Computer Science – 2011. – Vol. 6744. – P. 43-48
9. Маклейн С. Категории для работающего математика / пер. с англ. – М.: Физматлит – 2004. – 352 с.
10. Виноградов Д.В. О представлении объектов битовыми строками для ВКФ-метода // Научная и техническая информация. Сер. 2. – 2018. – № 5. – С. 1-4; Vinogradov D.V. On Object Representation by Bit Strings for the VKF-Method // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 3. – P. 113-116
11. Vinogradov D. Continuous Attributes for FCA-based Machine Learning // Proc. 8th Workshop FCA4AI 2020 co-located with ECAI 2020 / eds. S.O. Kuznetsov, A. Napoli, S. Rudolf. CEUR Workshop Proceedings – 2020. – Vol. 2729. – P. 103-111

Материал поступил в редакцию 15.02.21

Сведения об авторе

ВИНОГРАДОВ Дмитрий Вячеславович – доктор физико-математических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН; профессор Российского Государственного Гуманитарного Университета, Москва
e-mail: vinogradov.d.w@gmail.com