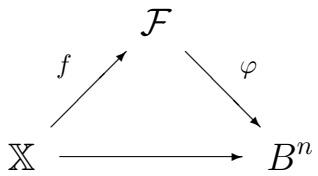
Е.А. Колмаков

Метрическое обобщение алгоритмов классификации на основе анализа формальных понятий

Введение

Анализ формальных понятий (*англ.* Formal Concept Analysis) — это направление прикладной теории решёток, позволяющее формализовать некоторые модели машинного обучения. С его помощью решаются задачи обработки и представления знаний, интеллектуального анализа данных. В частности, задачи классификации и кластеризации. Существует множество алгоритмов классификации на основе АФП [6], и многие из них предполагают, что объекты $x \in \mathbb{X}$ описываются при помощи бинарных (двоичных) признаков. Однако часто в прикладных задачах объекты описываются с помощью вещественных чисел, графов и других признаков. Некоторые алгоритмы используют исходное признаковое описание напрямую, например, узорные структуры [4, 5], но многие классификаторы используют признаки только после шкалирования. Для использования таких методов необходим переход (осуществляемый некоторым отображением φ) от исходного признакового пространства $\mathcal{F}$ к пространству бинарных векторов фиксированной длины, то есть булеву кубу B^n:



Обычно в задачах классификации исходное признаковое пространство $\mathcal{F}$ наделено дополнительной структурой, например, метрического пространства. Во многих случаях при отображении φ значительная часть метрической информации теряется и используется в новом признаковом пространстве только в слабой форме.

В этой работе предлагается модель алгоритмов классификации, обобщающая некоторые из существующих классификаторов [1, 3], рассматриваются её аналогии с ABO [7] и метрическими алгоритмами классификации. Предложенная модель использует метрическую

информацию из $\mathcal{F}$ наряду с объектно-признаковыми зависимостями. Кроме того, на произвольной конечной решётке вводится псевдометрика, которая имеет простой смысл в терминах понятий и может быть использована для их сравнения при построении и модификации классификаторов, использующих решётку формальных понятий.

1 Основные определения

1.1 Постановка задачи классификации

Будем рассматривать задачу классификации (обучения по прецедентам) в следующей постановке. Задано некоторое множество *объектов* $\mathbb{X} = \bigsqcup_{y \in Y} \mathbb{X}_y$, которое разделено на конечное число попарно непересекающихся подмножеств $\mathbb{X}_y$, называемых *классами*. Принадлежность элементов множества объектов к этим классам известна только для конечного подмножества $X \subset \mathbb{X}$, называемого *обучающей выборкой*.

Каждый объект $x \in \mathbb{X}$ описывается с помощью конечного набора *признаков* $\{f_i\}_{i=1}^n$, то есть отображений $f_i \colon X \to D_i$, при этом $f_i(x)$ — значение i-го признака объекта x. Множество $D_1 \times \cdots \times D_n$ обозначается $\mathcal{F}$ и называется исходным *признаковым пространством*. Обычно $\mathbb{X}$ отождествляется с $\mathcal{F}$.

По информации о разбиении множества объектов на классы, содержащейся в обучающей выборке, необходимо построить *алгоритм классификации* $a \colon \mathbb{X} \to Y$, который для каждого нового объекта $x \in \mathbb{X}$ указывал бы метку $a(x) \in Y$ содержащего его класса, либо некоторое выделенное значение, которое обозначает отказ от классификации и соответсвует ситуации, в которой алгоритм не смог решить вопрос о принадлежности объекта к одному из заданных классов.

1.2 Анализ формальных понятий

В этой статье мы пользуемся стандартной терминологией теории решёток и АФП. В этом разделе кратко даны основные определения, описаны классификаторы на основе АФП и используемые обозначения.

Пусть G и M — произвольные непустые множества, называемые *множеством объектов* и *множеством признаков*, а $I \subseteq G \times M$ — соответствие между G и M. Упорядоченная тройка $\mathbb{K} = (G, M, I)$ называется *формальным контекстом*. В случае конечных множеств объектов и признаков формальный контекст может быть задан с помощью объектно-признаковой матрицы.

Для любых $A \subseteq G$ и $B \subseteq M$ определим отображения $(\cdot)'$ следующим образом:

$$A' = \{m \in M | gIm \text{ для всех } g \in A\}, \quad B' = \{g \in G | gIm \text{ для всех } m \in B\}.$$

Эти отображения задают *соответствие Галуа* между множествами 2^G и 2^M. Мы пишем g' и m' вместо $\{g\}'$ и $\{m\}'$ для любых $g \in G, m \in M$.

Пара (A, B), где $A \subseteq G, B \subseteq M$ и $A' = B, B' = A$ называется *формальным понятием* контекста $\mathbb{K}$ с *формальным объёмом* A и *формальным содержанием* B. Определим "проекции" $ext : (A, B) \mapsto A$ и $int : (A, B) \mapsto B$. Множество формальных понятий контекста $\mathbb{K}$ образует полную решётку $\mathfrak{B}(\mathbb{K})$, называемую *решёткой формальных понятий*.

Пусть $\langle L, \wedge, \vee \rangle$ — решётка и $x \in L$. Через $x^{\triangle}(x^{\triangledown})$ мы обозначаем *порядковый идеал* (*фильтр*), *порождённый элементом* x. Через $At(L), J(L)$ и $M(L)$ — множество всех *атомов*, $\bigvee$*-неразложимых* и $\bigwedge$*-неразложимых* элементов решётки L соответсвенно. Функция $f \colon L \to \mathbb{R}$ называется *супермодулярной*, если $\forall x, y \in L$:

$$f(x) + f(y) \leqslant f(x \wedge y) + f(x \vee y).$$

Понятие C называется *непротиворечивым*, если все объекты из $ext(C)$ принадлежат одному классу. Обе процедуры классификации алгоритма GALOIS описаны в [3]. Классификатор GALOIS(1) вычисляет близость $S(x, C)$ между объектом и каждым непротиворечивым понятием, а затем x присваивается метка класса, соответствующего понятию C с наибольшим значением $S(x, C)$. Процедура GALOIS(2) строит множество всех непротиворечивых понятий, удовлетворяющих $int(C) \subseteq x'$, объекту x присваивается метка того класса, понятий которого больше всего в этом множестве.

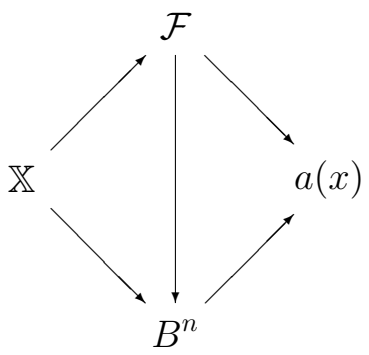
Предположим, что в задаче классификации заданы множества положительных и отрицательных примеров относительно некоторого *целевого свойства* $z \notin M$, а также множество недоопределённых примеров — объектов, для которых неизвестно значение предиката обладания свойством z. Такие входные данные могут быть заданы с помощью трёх контекстов: $\mathbb{K}_+ = (G_+, M, I_+)$, $\mathbb{K}_- = (G_-, M, I_-)$ и $\mathbb{K}_\tau = (G_\tau, M, I_\tau)$ — *положительный, отрицательный* и *недоопределённый контексты* соответственно. Здесь G_+, G_-, и G_τ суть, соответственно, множества положительных, отрицательных и недоопределенных примеров, M есть множество структурных признаков,

$I_\varepsilon \subseteq G_\varepsilon \times M, \varepsilon \in \{+, -, \tau\}$ суть отношения, которые определяют структурные признаки положительных, отрицательных и недоопределенных примеров. Операторы Галуа в этих контекстах обозначаются верхними индексами $+, -$ и τ соответственно. Формальное понятие (A_+, B_+) положительного контекста называется *положительным понятием*, при этом A_+ называется. Аналогично для *отрицательных* и *недоопределённых понятий*. Формальное содержание B_+ положительного понятия (A_+, B_+) называется *положительной гипотезой*, если оно не является подмножеством формального содержания g^- какого-либо отрицательного примера $g \in G_-$. Аналогично для *отрицательных гипотез*.

Если формальное содержание недоопределенного примера $g \in G_\tau$ содержит положительную (отрицательную) гипотезу, то говорят, что эта гипотеза является *гипотезой в пользу положительной* (*отрицательной*) *классификации* g. Под алгоритмом классификации на основе гипотез мы подразумеваем следующую процедуру, описанную в [1]. Если недоопределённый пример $g \in G_\tau$ содержит положительную гипотезу и ни одной отрицательной, то он классифицируется положительно, аналогично для отрицательных гипотез. Если g не содержит ни положительных, ни отрицательных гипотез или содержит как положительную, так и отрицательную гипотезы, то происходит отказ от классификации.

2 Обобщение и модификация алгоритмов

Общим недостатком классификаторов на основе АФП, использующих бинарные признаки, является то, что они забывают метрическую структуру исходного признакового пространства $\mathcal{F}$. Основная идея этой работы — использовать исходную метрическую информацию наряду с теоретико-порядковыми отношениями между объектами и признаками, которые задаются решёткой понятий. Тогда процесс построения классификатора $a(x)$ может быть представлен следующей схемой:



Важным является то, что пространства $\mathcal{F}$ и B^n вместе с дополнительными структурами на них (в первом пространстве — это метрика, во втором — это формальный контекст, задаваемый обучающей выборкой) используются одновременно, что предоставляет более богатый набор методов для построения алгоритма классификации и позволяет сохранить использовать исходную метрическую информацию напрямую.

2.1 Введение метрических оценок

Пусть $\mathcal{H}_+$ и $\mathcal{H}_-$ — построенные по обучающей выборке множества понятий, содержания которых являются положительными и отрицательными гипотезами, $S(x, C)$ — мера близости между объектом x и множеством объектов C (основанная на функции расстояния из исходного признакового пространства $\mathcal{F}$).

Пусть $(\mathcal{F}, \rho)$ — метрическое пространство. Тогда разумно выбирать меру близости $S(x, C)$ как некоторую невозрастающую функцию от $\rho(x, C)$. Расстояние от точки x до множества C может быть определено разными способами, например:

- $\rho(x, C) = \inf_{c \in C} \rho(x, c)$,
- $\rho(x, C) = \rho(x, c_*)$, где c_* — некоторый "центр масс" системы точек C,
- $\rho(x, C) = \frac{1}{|C|} \sum_{c \in C} \rho(x, c)$ и другие.

Определим оценки за положительный и отрицательный классы:

$$\Gamma_+(x) = \sum_{C \in \mathcal{H}_+} I(x, C) S(x, ext(C)), \quad \Gamma_-(x) = \sum_{C \in \mathcal{H}_-} I(x, C) S(x, ext(C)),$$

где $I(x, C) = [int(C) \subseteq x']$ и $[\cdot]$ — индикаторная функция. Тогда итоговый классификатор будет иметь следующий вид:

$$a(x) = \text{sign } \Gamma(x) = \text{sign}\big(\Gamma_+(x) - \Gamma_-(x)\big).$$

Что можно сказать о качестве введённого классификатора $a(x)$?

Утверждение 1. *Если алгоритм классификации на основе гипотез верно распознаёт объект, то это же относится и к классификатору* $a(x) = \text{sign } \Gamma(x)$.

Поскольку алгоритм $a(x)$ отказывается от классификации в меньшем числе случаев, чем раньше, но число ошибок (среди классифицированных объектов) может увеличиться. Можно предложить следующую модификацию:

$$a_R(x) = \begin{cases} \text{sign } \Gamma(x), & |\Gamma(x)| > R; \\ \text{отказ от классификации}, & \text{иначе.} \end{cases}$$

Здесь R — некоторая положительная константа. $a_R(x)$ классифицирует увереннее, но с ростом R увеличивается число отказов. Такой подход влечёт за собой проблему выбора порога R.

2.2 Аналогия с алгоритмами вычисления оценок

В приведённом выше алгоритме при вычислении оценок используются множества положительных и отрицательных гипотез, то есть подмножества решётки понятий специального вида. Возникает идея обобщения данных оценок на подмножества произвольного вида, тем или иным образом характеризующие отдельные классы $y \in Y$.

Пусть $\mathcal{C}$ — множество понятий, каждое понятие $C \in \mathcal{C}$ характеризует некоторый класс $y \in Y$ и только его, то есть

$$\mathcal{C} = \bigsqcup_{y \in Y} \mathcal{C}_y, \text{ где } Y \text{ — множество классов.}$$

В этом случае будем называть $\mathcal{C}$ *системой опорных понятий*.

Определим оценку объекта x за класс y следующим образом:

$$\Gamma_y(x) = \sum_{C \in \mathcal{C}_y} S(x, C).$$

Итоговый классификатор имеет стандартный вид:

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x).$$

Оценки такого вида схожи с оценками за классы, которые используются в алгоритмах вычисления оценок [7], а множества $\mathcal{C}_y$ являются аналогами опорных множеств. Приведём конкретные примеры системы опорных понятий $\mathcal{C}$ и функции близости $S(x, C)$ и рассмотрим получающиеся при этом классификаторы:

- $\mathcal{C} = \mathcal{H}_+ \bigsqcup \mathcal{H}_-$ — множество положительных и отрицательных гипотез.

$S(x,C) = [int(C) \subseteq x']\hat{S}(x, ext(C))$, где $\hat{S}(x, ext(C))$ — некоторая функция близости. Соответствующий классификатор был описан выше.

- $\mathcal{C} = \bigsqcup_{y \in Y} \mathcal{C}_y$ — множество непротиворечивых понятий.
  $S(x,C) = |\big(M \setminus int(C)\big) \cup x'|$ — модифицированный GALOIS(1).
  $S(x,C) = [int(C) \subseteq x']$ — алгоритм GALOIS(2).

2.3 Аналогия с метрическими алгоритмами классификации

Пусть $\mathcal{C} = \{C_1, \ldots, C_n\}$ — множество опорных понятий. Предположим, что в исходном признаковом пространстве $\mathcal{F}$ введена мера расстояния ρ и на её основе определена функция расстояния между объектами и понятиями $\rho(x, C)$. Упорядочим $\mathcal{C}$ по возрастанию расстояний от C_i до классифицируемого объекта x:

$$\rho(x, C_x^{(1)}) \leqslant \rho(x, C_x^{(2)}) \leqslant \cdots \leqslant \rho(x, C_x^{(n)}),$$

$C_x^{(i)}$ — i-ый сосед объекта x среди $\mathcal{C}$, $y_x^{(i)}$ — класс, характеризуемый понятием $C_x^{(i)}$.

Определим оценку объекта x за класс y:

$$\Gamma_y(x) = \sum_{i=1}^{n} w_i(x)[y_x^{(i)} = y],$$

$w_i(x)$ — вес i-го соседа объекта x (неотрицательная и невозрастающая по i функция).

Введённые оценки полностью аналогичны оценкам, используемым в метрических алгоритмах классификации, за исключением того, что в качестве соседей выступают не объекты, а опорные понятия. Таким образом, выбирая подходящие веса $w_i(x)$, получаем аналоги всех известных метрических методов (kNN, Parzen window, potential functions и других), но в терминах понятий. Например:

- $w_i(x) = [i \leqslant k]$ — метод k ближайших соседей;
- $w_i(x) = [i \leqslant k]w_i$ — метод k взвешенных ближайших соседей, здесь w_i — вес, зависящий только от номера соседа;
- $w_i(x) = K(\frac{\rho(x, C_x^{(i)})}{h(x)})$ — метод Парзеновского окна переменной ширины, здесь $K(z)$ — невозрастающая положительная функция на $[0,1]$, $h(x)$ — ширина окна (например, $h(x) = \rho(x, C_x^{(k+1)})$).

Предложенные в этом и предыдущих пунктах методы имеют следующие достоинства:

- использование метрической информации из исходного признакового пространства при вычислении близости $S(x, C)$ или расстояния $\rho(x, C)$ наряду с объектно-признаковыми зависимостями, задаваемыми решёткой понятий, что позволяет снизить число отказов от классификации и ошибок;

- предложенные алгоритмы являются обобщением некоторых из уже существующих и могут быть использованы для их модификации (например, алгоритма GALOIS или классификации на основе гипотез).

Следует отметить, что проблема сложности построения решётки понятий остаётся нерешённой. Возможный вариант решения — выбор системы опорных понятий $\mathcal{C}$ таким образом, чтобы избежать построения всей решётки. Также, за счёт использования дополнительной информации об объектах, возникают вопросы, касающиеся выбора метрики в исходном признаковом пространстве, выбора меры близости $S(x, C)$ и системы опорных понятий $\mathcal{C}$.

2.4 Псевдометрика на решётке понятий

Другой подход, использующий понятие близости в алгоритмах АФП, заключается в введении функции расстояния на множестве всех понятий. При выводе логических правил в алгоритме Rulearner [2] наиболее важными характеристиками элемента x решётки понятий при его сравнении с другими понятиями были значение функции $\mathrm{cover}(x) = |J(L) \cap x^{\nabla}|$ и множество $M(x) = M(L) \cup x^{\triangle}$. В случае приведённого контекста, это согласуется с тем, что понятие характеризуется своим объёмом (отдельные объекты соответствуют $\bigvee$-неразложимым элементам решётки) и содержанием (отдельные признаки соответствуют $\bigwedge$-неразложимым элементам решётки).

Таким образом, $\mathrm{cover}(x)$ соответствует числу объектов из обучающей выборки, покрываемых понятием x, а множество $M(x)$ — признакам, которые характеризуют данное понятие. Воспользуемся этими соображениями для введения функции близости на произвольной конечной решётке. В силу утверждений, двойственных к теоремам 3.1 и 3.3 из статьи [8], справедлива следующая

Теорема 1. *Пусть* $\langle L, \wedge, \vee \rangle$ *— решётка, и функция* $f\colon L \to \mathbb{R}$ *изотонна и супермодулярна. Тогда* $d_f(x, y) = f(x) + f(y) - 2f(x \wedge y)$

является псевдометрикой на этой решётке.

Рассмотрим произвольную конечную решётку $\langle L, \wedge, \vee \rangle$, непустое подмножество $D \subseteq L$ и функцию $f\colon L \to \mathbb{Z}_+$, определённую следующим образом:

$$f(x) = |D(x)|, \text{ где } D(x) = D \cap x^{\triangledown}.$$

Утверждение 2. *Функция $f(x)$ является изотонной и супермодулярной.*

Доказательство. Изотонность f следует из следующей цепочки импликаций:

$$x \leqslant y \Rightarrow x^{\triangledown} \subseteq y^{\triangledown} \Rightarrow D(x) \subseteq D(y) \Rightarrow f(x) = |D(x)| \leqslant |D(y)| = f(y).$$

Перейдём к доказательству супермодулярности:

$$f(x) + f(y) = |D(x)| + |D(y)| = |D(x) \cup D(y)| + |D(x) \cap D(y)| \leqslant$$
$$\leqslant f(x \vee y) + f(x \wedge y).$$

Докажем последнее неравенство. Включение $D(x) \cup D(y) \subseteq D(x \vee y)$ следует из включений:

$$x \leqslant x \vee y \Rightarrow D(x) \subseteq D(x \vee y),$$

$$y \leqslant x \vee y \Rightarrow D(y) \subseteq D(x \vee y),$$

Равенство $D(x) \cap D(y) = D(x \wedge y)$ следует из того, что $x^{\triangledown} \cap y^{\triangledown} = (x \wedge y)^{\triangledown}$. □

Таким образом, согласно теореме 3, функция $f(x)$ индуцирует псевдометрику $d_f(x, y)$ на решётке, определяемую следующей формулой:

$$d_g(x, y) = f(x) + f(y) - 2f(x \wedge y).$$

Значение функции $d_f(x, y)$ имеет простой смысл.

Утверждение 3. $d_f(x, y) = |D(x) \oplus D(y)|$, *где* $A \oplus B = (A \setminus B) \cup (B \setminus A)$.

Доказательство. На последнем шаге доказательства утверждения 2 установлено соотношение $D(x) \cap D(y) = D(x \wedge y)$.

$$\begin{aligned} f(x) + f(y) - 2f(x \wedge y) &= |D(x)| + |D(y)| - 2|D(x \wedge y)| = \\ &= |D(x)| + |D(y)| - 2|D(x) \cap D(y)| = \\ &= |D(x) \cup D(y)| + |D(x) \cap D(y)| - 2|D(x) \cap D(y)| = \\ &= |D(x) \cup D(y)| - |D(x) \cap D(y)| = |D(x) \oplus D(y)|. \end{aligned}$$

□

Следствие. *Если* $\langle L, \wedge, \vee \rangle$ *— конечная булева алгебра и* $D = At(L)$, *то* $d_f(x, y)$ *— это в точности расстояние Хэмминга.*

Для сравнения формальных понятий разумно выбирать $D = J(L)$ или $D = At(L)$. В терминах этой псведометрики два понятия тем ближе, чем меньше примеров (то есть понятий вида (g'', g'), где $g \in G$) покрывается только одним из данных понятий и не покрывается другим. Более того, функция cover(x) выражается через функцию $d_f(x, y)$ следующим образом: cover(x) $= d_f(x, 0)$.

Одним из недостатков введённой меры расстояния является то, что число элементов из $D(x \wedge y)$ никак не учитывается, что в некоторых случаях может привести к неадекватным оценкам расстояния. Возможные модификации:

1. Учёт числа атомов из пересечения путём «нормировки», например:
$$d(x, y) = \frac{|D(x) \oplus D(y)|}{|D(x) \cup D(y)|}.$$

2. Добавление весов элементам D. Например, пусть w_d — доля гипотез (или непротиворечивых понятий), покрывающих $d \in D$. Тогда $d(x, y)$ примет вид:
$$d(x, y) = \sum_{d \in D(x) \oplus D(y)} w_d.$$

Расстояние между понятиями можно также применять для модификации классификаторов на основе АФП. Пусть, например, для классификации объекта x используется алгоритм на основе гипотез. Предположим, что имеется две гипотезы H_1^+, H_2^+ в пользу положительной классификации x и две гипотезы H_1^-, H_2^- в пользу отрицательной классификации x. В этом случае стандартный алгоритм отказывается от классификации. Предположим, что нам дополнительно известны расстояния $d(H_1^+, H_2^+), d(H_1^-, H_2^-)$ и $d(H_1^+, H_2^+) \gg d(H_1^-, H_2^-)$. Тогда имеет смысл отнести x к положительному классу, поскольку далёкие по отношению к введённой мере понятия являются менее “коррелированными” (так как они покрывают большое число различных примеров), а значит, их голоса более значимы.

Мера расстояния между понятиями также может быть использована для уменьшения размера системы опорных понятий (например,

гипотез), используемой классификатором. Это способствует увеличению обобщающей способности классификатора, уменьшению переобучения и удалению “шумовых” понятий.

3 Эксперименты

Для тестирования некоторых из предложенных методов и сравнения их с алгоритмами, рассмотренными в третьем разделе, были использованы два набора данных из UCI Machine Learning Repository [9]. Помимо алгоритмов из третьего раздела тестировались:

1. Модификация алгоритма GALOIS(1) (см. второй пример из пункта 3.2).

2. Модификации алгоритма классификации на основе гипотез с помощью введения метрических оценок (см. пункт 3.1). Была выбрана функция близости $S(x, C) = K(\rho(x, C), a)$. В качестве $K(r, a)$ использовалась одна из функций:

$$K_1(r, a) = \frac{1}{1 + \exp(ar)}, \qquad K_2(r, a) = \frac{1}{a + r}.$$

Для вычисления $\rho(x, C)$ использовались следующие варианты:

$$\rho_1(x, C) = \inf_{c \in C} \rho(x, c), \quad \rho_2(x, C) = \frac{1}{|C|} \sum_{c \in C} \rho(x, c),$$

$$\rho_3(x, C) = \sup_{c \in C} \rho(x, c).$$

Введём следующие обозначения:

ν_c — доля объектов, на которых не произошло отказа от классификации;

$\nu_r = 1 - \nu_c$ — доля отказов от классификации;

e_t — общая доля ошибок классификации (отказ также считается ошибкой);

e_r — доля ошибок классификации среди всех классифицированных объектов.

SPECT Heart Data Set

Алгоритм	ν_c	ν_r	e_t	e_r
GALOIS(1)	1	0	0.1604	0.1604
Modified GALOIS(1)	1	0	0.0856	0.0856
GALOIS(2)	1	0	0.0802	0.0802
Rulearner	0.7487	0.2513	0.2727	0.0286
Hypotheses-based	0.5936	0.4064	0.6150	0.1842
$K = K_1, a = 0.0125, \rho = \rho_1$	0.8021	0.1979	0.3155	0.1467
$K = K_1, a = 0.0125, \rho = \rho_2$	0.8021	0.1979	0.2888	0.1133
$K = K_1, a = 0.0125, \rho = \rho_3$	0.8021	0.1979	0.2834	0.1067
$K = K_1, a = 1, \rho = \rho_2$	0.7273	0.2727	0.3422	0.0956
$K = K_2, a = 1, \rho = \rho_1$	0.8021	0.1979	0.2941	0.1200
$K = K_2, a = 1, \rho = \rho_2$	0.8021	0.1979	0.3209	0.1533

Таблица 1. Результаты тестирования алгоритмов. Задача SPECT.

Данные разделены на обучающую (80 объектов) и контрольную (187 объектов) выборки. Для этой задачи существует два набора данных: SPECT и SPECTF Heart Data Set. В первом каждый объект описывается с помощью 22 бинарных признаков, а во втором — с помощью 44 числовых. В качестве функции расстояния во втором случае использовалась обычная евклидова метрика. Результаты тестирования представлены в таблице 1.

Liver Disorders Data Set

Данные были разделены на обучающую (150 объектов) и контрольную выборки (195 объектов). Изначально каждый объект описывался с помощью 6 числовых признаков. Была проведена простая процедура бинаризации: каждый признак линейным преобразованием переводился в $[0, 1]$, этот отрезок делился на 5 частей, и каждому признаку ставился в соответствие вектор из B^5 с единицей в k-ой позиции, где k — номер интервала, в который попал признак после такого преобразования. Таким образом, после бинаризации каждый признак лежит в 5-ом слое булева куба B^{30}. В исходном признаковом пространстве использовалась евклидова метрика. Результаты тестирования представлены в таблице 2.

Алгоритм	ν_c	ν_r	e_t	e_r
GALOIS(1)	1	0	0.4605	0.4605
Modified GALOIS(1)	1	0	0.5590	0.5590
GALOIS(2)	1	0	0.4359	0.4359
Rulearner	0.9795	0.0205	0.4564	0.4450
Hypotheses-based	0.2923	0.7077	0.8256	0.4035
$K = K_1, a = 1, \rho = \rho_1$	0.8821	0.1179	0.5231	0.4593
$K = K_1, a = 0.01, \rho = \rho_2$	0.8974	0.1026	0.5436	0.4914
$K = K_1, a = 0.25, \rho = \rho_3$	0.8872	0.1128	0.5385	0.4798
$K = K_2, a = 200, \rho = \rho_1$	0.8974	0.1026	0.4769	0.4171
$K = K_2, a = 150, \rho = \rho_2$	0.8974	0.1026	0.4564	0.3943
$K = K_2, a = 150, \rho = \rho_3$	0.8974	0.1026	0.4667	0.4057

Таблица 2. Результаты тестирования алгоритмов. Задача Liver Disorders.

Целью проведённых экспериментов было не решение конкретных задач классификации, а сравнение методов классификации на основе АФП между собой. В связи с этим применялась очень простая процедура бинаризации, чем объясняется высокая ошибка всех алгоритмов во второй задаче, и параметр a выбирался с помощью перебора по сетке небольшого размера. Выбор подходящей процедуры бинаризации признаков является отдельной темой для исследований и может существенно улучшить качество классификации. Например, можно использовать интервалы переменной длины.

На основе полученных результатов можно сделать следующие выводы об эффективности предложенных алгоритмов:

1. Во всех случаях число отказов от классификации, по сравнению с алгоритмом классификации на основе гипотез, существенно уменьшено, при этом в первой задаче средняя относительная доля ошибок e_r существенно ниже, а во второй незначительно больше, чем e_r алгоритма классификации на основе гипотез.

2. Общая доля ошибок классификации e_t модификаций алгоритма на основе гипотез в первой задаче сравнима с e_t алгоритма Rulearner, а во второй — с e_t алгоритма GALOIS. В обеих задачах она значительно ниже, чем у алгоритма классификации на основе гипотез.

3. В первой задаче модификация алгоритма GALOIS(1) улучшила качество классификации почти на 8%, что достаточно существенно. Однако во второй задаче эта же модификация ухудшила качество классификации, что говорит о чувствительности этого метода к процедуре шкалирования.

4. Выбор конкретной функции $K(r, a)$ и варианта вычисления расстояния $\rho(x, C)$ почти не влияют на количество отказов от классификации, однако влияют на число ошибок. Таким образом, с помощью правильного их подбора можно улучшить качество классификации.

4 Заключение

В этой статье был формально изучен и экспериментально исследован новый подход к построению классификаторов, совмещающий подход АФП и классификаторов, использующих понятие близости. Описанная модель классификаторов обобщает некоторые из существующих классификаторов на основе АФП и использует как исходную метрическую информацию, так и порядковые объектно-признаковые зависимости. Также введена псевдометрика между элементами произвольной конечной решётки, которая имеет понятную интерпретацию в терминах понятий и которую можно использовать для сравнения понятий с целью модификации алгоритмов классификации на основе АФП.

Возможные дальнейшие исследования могут быть направлены на исследование алгоритмов классификации, получаемых при конкретном выборе системы опорных понятий $\mathcal{C}$ и меры близости $S(x, C)$, и возможности выбора системы $\mathcal{C}$ так, чтобы избежать построения всей решётки понятий целиком или исключить отказы от классификации, а также на разработку процедур классификации с использованием введённой на понятиях псевдометрики.

Литература

1. *S.O. Kuznetsov.* Complexity of Learning in Concept Lattices from Positive and Negative Examples. Discrete Applied Mathematics, 2004, No. 142(1–3), pp. 111-125.

2. *M. Sahami.* Learning classification Rules Using Lattices. N. Lavrac and S. Wrobel eds., pp. 343–346, Proc ECML, Heraclion, Crete, Greece (April 1995).

3. *C. Caprineto, G. Romano*. GALOIS An order-theoretic approach to conceptual clustering. In proceedings of ICML93, pp. 33–40, Amherst, USA (July 1993).

4. *M. Kaytoue, S.O. Kuznetsov, A. Napoli, S. Duplessis*. Mining gene expression data with pattern structures in formal concept analysis. Information Sciences, Volume 181, Issue 10, 15 May 2011, pp. 1989-2001, Information Science, 2011.

5. *S.O. Kuznetsov*. Scalable Knowledge Discovery in Complex Data with Pattern Structures. In: P. Maji, A. Ghosh, M.N. Murty, K. Ghosh, S.K. Pal, Eds., Proc. 5th International Conference Pattern Recognition and Machine Intelligence (PReMI'2013), Lecture Notes in Computer Science (Springer), Vol. 8251, pp. 30-41, 2013.

6. *O. Prokasheva, A. Onishchenko, S. Gurov*. Classification methods based on Formal Concept Analysis. FCAIR 2013 – Formal Concept Analysis Meets Information Retrieval. Workshop co-located with the 35th European Conference on Information Retrieval (ECIR 2013). March 24, 2013, Moscow, Russia. National Research University Higher School of Economics, pp. 95-104. ISSN 1613-0073

7. *Ю.И. Журавлёв*. Об алгебраическом подходе к решению задач распознавания или классификации. — Проблемы кибернетики: — 1978. — Т.33. — С.5-68.

8. *Dan A. Simovici*. Betweenness, Metrics and Entropies in Lattices. Proceedings of the 38th International Symposium on Multiple Valued Logic. 22-24 May, 2008, Dallas, TX, USA. IEEE Computer Society Washington, pp. 26-31. ISSN 0195-623X

9. *Bache, K. & Lichman, M.* (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.