

Анализ формальных понятий в задачах классификации

С. И. Гуров

Факультет Вычислительной математики и кибернетики,
МГУ имени М.В. Ломоносова

Кафедра Математических методов прогнозирования

e-mail: sgur@cs.msu.ru

Просеминар кафедры ММП

19 марта 2013 г. / Москва

Содержание

- 1 Введение. Задача классификации и методы её решения
- 2 Частично упорядоченные множества, решётки и отображения
- 3 Решётка формальных понятий
- 4 Классификация на основе гипотез

Раздел

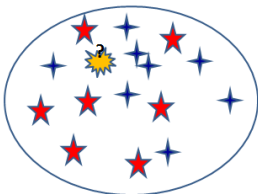
- 1 Введение. Задача классификации и методы её решения
- 2 Частично упорядоченные множества, решётки и отображения
- 3 Решётка формальных понятий
- 4 Классификация на основе гипотез

Классификация по прецедентам: постановка задачи




- 1 Множество объектов \mathcal{X} разделено на несколько подмножеств (*классов*).
- 2 Информация о таком разбиении содержится только в указании о принадлежности к данным классам элементов конечной *обучающей последовательности (выборки)* из \mathcal{X} , элементы которой называют *прецедентами*.
- 3 Объекты имеют описание на некотором формальном языке, указывающем степень обладания объектами конечным числом признаков из множества $M = \{x_1, \dots, x_n\}$.

Классификация: пространство объектов

Распознавание образов



пространство объектов

-  - объекты класса A
-  - объекты класса B
-  - объект неизвестного класса

прецедент	класс
Объект 1	A
Объект 2	B
...	...
Объект L	A
Объект X	?

- Поиск полезных ископаемых
- Медицинская диагностика
- Прогнозирование
- ...

Рис. 1. Информационная модель классификации

Классификация: признаковая матрица

Часто используется описание в виде *объектно-признаковой (0, 1)-матрицы* M , в которой объектам соответствуют строки, признакам — столбцы, а элементы матрицы кодируют наличие/отсутствие признаков у объектов.

Класс K_i	x_1	x_2	\dots	x_n
Объект 1	1	0	\dots	1
Объект 2	0	1	\dots	1
Объект 3	1	1	\dots	0
\dots	\dots	\dots	\dots	\dots
Объект m_i	0	1	\dots	0

— для каждого из

классов $K_1, \dots, K_s, s \geq 2$. Далее $s = 2$.

Классификация: язык описания и решающее правило

Задача обучения

По матрице M сформулировать *решающее правило*, которое по описанию нового объекта из \mathcal{X} указывало бы имя класса, его содержащего. Решающее правило должно максимизировать некоторой функционал, определяющей *качество классификации*.

Таким функционалом в подавляющем числе случаев является минимум (*не абсолютный!*) числа ошибок классификации, однако может также учитываться, например, и доля отказов.

Классификация: подходы к решению задачи

- метрические методы (*NN*, ...);
- разделяющие поверхности (*SVM*, ...);
- потенциальные функции;
- логические методы;
- коллективные решающие правила (*области компетенции, голосование, алгебраический подход*);
- структурные методы;
- ...
- реляционный подход (**АФП (FCA)**, ...)

Wille R., Ganter B. Formal concept analysis. Berlin; Heidelberg; New York: Springer-Verl., 1999.

Классификация: подходы к решению задачи

- метрические методы (*NN*, ...);
- разделяющие поверхности (*SVM*, ...);
- потенциальные функции;
- логические методы;
- коллективные решающие правила (*области компетенции, голосование, алгебраический подход*);
- структурные методы;
- ...
- реляционный подход (**АФП (FCA)**, ...)

Wille R., Ganter B. Formal concept analysis. Berlin; Heidelberg; New York: Springer-Verl., 1999.

Раздел

- 1 Введение. Задача классификации и методы её решения
- 2 Частично упорядоченные множества, решётки и отображения**
- 3 Решётка формальных понятий
- 4 Классификация на основе гипотез

Отношения

Пусть A и B — непустые множества; их *декартовым произведением* (символически $A \times B$) называют совокупность всех пар вида (a, b) , где $a \in A, b \in B$.

Определение

Бинарное отношение или (*соответствие между A и B*) есть подмножество $A \times B$.

Нотация: $(a, b) \in \rho \subseteq A \times B \Leftrightarrow a\rho b$.

Если $A = B$ то отношение $\rho \subseteq A \times A$ называется *однородным*.

Антисимметричность ρ : $(a\rho b) \& (b\rho a) \Rightarrow a = b$.

Отношения

Пусть A и B — непустые множества; их *декартовым произведением* (символически $A \times B$) называют совокупность всех пар вида (a, b) , где $a \in A, b \in B$.

Определение

Бинарное отношение или (*соответствие между* A и B) есть подмножество $A \times B$.

Нотация: $(a, b) \in \rho \subseteq A \times B \Leftrightarrow a\rho b$.

Если $A = B$ то отношение $\rho \subseteq A \times A$ называется *однородным*.

Антисимметричность $\rho: (a\rho b) \& (b\rho a) \Rightarrow a = b$.

Отношения

Пусть A и B — непустые множества; их *декартовым произведением* (символически $A \times B$) называют совокупность всех пар вида (a, b) , где $a \in A$, $b \in B$.

Определение

Бинарное отношение или (*соответствие между* A и B) есть подмножество $A \times B$.

Нотация: $(a, b) \in \rho \subseteq A \times B \Leftrightarrow a\rho b$.

Если $A = B$ то отношение $\rho \subseteq A \times A$ называется *однородным*.

Антисимметричность ρ : $(a\rho b) \& (b\rho a) \Rightarrow a = b$.

Отношения

Пусть A и B — непустые множества; их *декартовым произведением* (символически $A \times B$) называют совокупность всех пар вида (a, b) , где $a \in A$, $b \in B$.

Определение

Бинарное отношение или (*соответствие между* A и B) есть подмножество $A \times B$.

Нотация: $(a, b) \in \rho \subseteq A \times B \Leftrightarrow a\rho b$.

Если $A = B$ то отношение $\rho \subseteq A \times A$ называется *однородным*.

Антисимметричность ρ : $(a\rho b) \& (b\rho a) \Rightarrow a = b$.

Частично упорядоченное (ч.у.) множество — это...

пара $\mathbf{P} = \langle P, \leq \rangle$, где P — непустое множество, а \leq — *рефлексивное, антисимметричное и транзитивное* однородное отношение на нём.
 $a \leq b$ — « a *предшествует* b » или « b *содержит* a ».

Примеры

- $\langle 2^P, \subseteq \rangle$ — упорядоченность подмножеств множества P *по включению*;
- $\mathbf{N}_1 = \langle \mathbb{N}, \leq \rangle$ и $\mathbf{N}_2 = \langle \mathbb{N}, | \rangle$ — упорядоченность натуральных чисел естественным (*линейным*) порядком и по делимости. 6 *непосредственно предшествует* 7 в \mathbf{N}_1 и $12, 18, 30, \dots$ в \mathbf{N}_2 .

Частично упорядоченное (ч.у.) множество — это...

пара $\mathbf{P} = \langle P, \leq \rangle$, где P — непустое множество, а \leq — *рефлексивное, антисимметричное и транзитивное* однородное отношение на нём.
 $a \leq b$ — « a *предшествует* b » или « b *содержит* a ».

Примеры

- $\langle 2^P, \subseteq \rangle$ — упорядоченность подмножеств множества P *по включению*;
- $\mathbf{N}_1 = \langle \mathbb{N}, \leq \rangle$ и $\mathbf{N}_2 = \langle \mathbb{N}, | \rangle$ — упорядоченность натуральных чисел естественным (*линейным*) порядком и по делимости. 6 *непосредственно предшествует* 7 в \mathbf{N}_1 и $12, 18, 30, \dots$ в \mathbf{N}_2 .

Частично упорядоченное (ч.у.) множество — это...

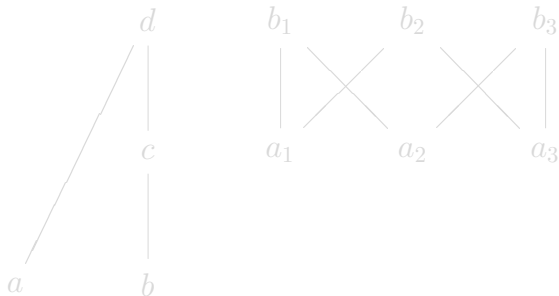
пара $\mathbf{P} = \langle P, \leq \rangle$, где P — непустое множество, а \leq — *рефлексивное, антисимметричное и транзитивное* однородное отношение на нём.
 $a \leq b$ — « a *предшествует* b » или « b *содержит* a ».

Примеры

- $\langle 2^P, \subseteq \rangle$ — упорядоченность подмножеств множества P *по включению*;
- $\mathbf{N}_1 = \langle \mathbb{N}, \leq \rangle$ и $\mathbf{N}_2 = \langle \mathbb{N}, | \rangle$ — упорядоченность натуральных чисел естественным (*линейным*) порядком и по делимости. 6 *непосредственно предшествует* 7 в \mathbf{N}_1 и $12, 18, 30, \dots$ в \mathbf{N}_2 .

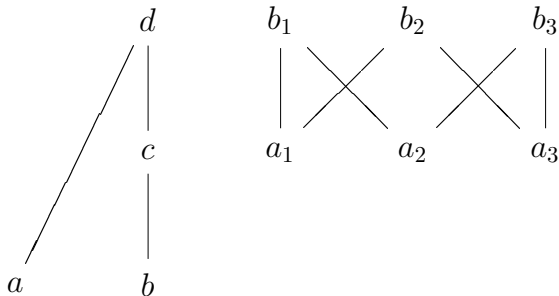
Диаграммы Хассе - это...

... способ наглядного представления ч.у. множеств: если элемент a **предшествует** элементу b , то a рисуют **ниже** b и соединяют их **отрезком**, если это предшествование **непосредственное**.



Диаграммы Хассе - это...

... способ наглядного представления ч.у. множеств: если элемент a **предшествует** элементу b , то a рисуют **ниже** b и соединяют их **отрезком**, если это предшествование **непосредственное**.



Особые элементы ч.у. множеств: определения

Определение

Элемент $u \in P$ ч.у. множества $\langle P, \leq \rangle$ называют:

- *максимальным*, если $u \leq x \Rightarrow u = x$,
- *минимальным*, если $u \geq x \Rightarrow u = x$,
- *наибольшим*, если $x \leq u$,
- *наименьшим*, если $x \geq u$

для любых $x \in P$.

Наибольший элемент (если он существует) единственен и является также и *максимальным*.

Аналогично для наименьших и минимальных элементах.

Особые элементы ч.у. множеств: определения

Определение

Элемент $u \in P$ ч.у. множества $\langle P, \leq \rangle$ называют:

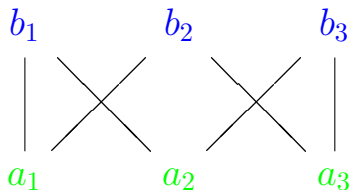
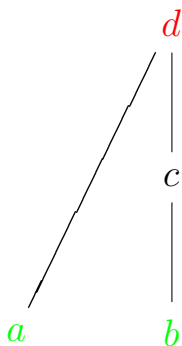
- *максимальным*, если $u \leq x \Rightarrow u = x$,
- *минимальным*, если $u \geq x \Rightarrow u = x$,
- *наибольшим*, если $x \leq u$,
- *наименьшим*, если $x \geq u$

для любых $x \in P$.

Наибольший элемент (если он существует) единственен и является также и **максимальным**.

Аналогично для наименьших и минимальных элементах.

Особые элементы ч.у. множеств: примеры



наибольший элемент максимальные элементы
минимальные элементы

Конусы ч.у. множества: определение

Определение

Пусть \mathbf{P} — ч.у. множество и $A \subseteq P$. Множества A^Δ и A^∇ определяемые условиями

$$A^\Delta = \{x \in P \mid \forall(a \in A) (a \leq x)\};$$

$$A^\nabla = \{x \in P \mid \forall(a \in A) (x \leq a)\}$$

называются *верхним* и *нижним конусами множества A* , а их элементы — *верхними* и *нижними гранями множества A* соответственно.

Антимонотонность конусов подмножеств по включению: $A \subseteq B \Rightarrow B^\nabla \subseteq A^\nabla$ и $B^\Delta \subseteq A^\Delta$.

Конусы ч.у. множества: определение

Определение

Пусть \mathbf{P} — ч.у. множество и $A \subseteq P$. Множества A^Δ и A^∇ определяемые условиями

$$A^\Delta = \{x \in P \mid \forall(a \in A) (a \leq x)\};$$

$$A^\nabla = \{x \in P \mid \forall(a \in A) (x \leq a)\}$$

называются *верхним* и *нижним конусами множества A* , а их элементы — *верхними* и *нижними гранями множества A* соответственно.

Антимонотонность конусов подмножеств по включению: $A \subseteq B \Rightarrow B^\nabla \subseteq A^\nabla$ и $B^\Delta \subseteq A^\Delta$.

Точные грани ч.у. множества:

Определение

Пусть \mathbf{P} — ч.у. множество и $A \subseteq P$.

- Если в A^Δ существует **наименьший** элемент, то он называется *точной верхней гранью множества A* (символически $\sup A$).
- Если в A^∇ существует **наибольший** элемент, то он называется *точной нижней гранью множества A* (символически $\inf A$).

Точные грани ч.у. множества:

Определение

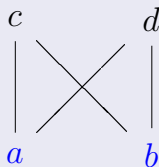
Пусть \mathbf{P} — ч.у. множество и $A \subseteq P$.

- Если в A^Δ существует **наименьший** элемент, то он называется *точной верхней гранью множества A* (символически $\sup A$).
- Если в A^∇ существует **наибольший** элемент, то он называется *точной нижней гранью множества A* (символически $\inf A$).

Точные грани ч.у. множества:

Примеры

- 1 Пусть $A = \{a, b\}$ в ч.у. множестве



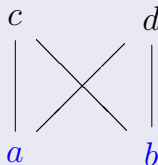
тогда $A^\Delta = \{c, d\}$ и $\sup A$ отсутствует.

- 2 Если S — совокупность подмножеств некоторого множества, то (по включению) $\sup S$ совпадает с объединением, а $\inf S$ — с пересечением всех подмножеств из S .

Точные грани ч.у. множества:

Примеры

- 1 Пусть $A = \{a, b\}$ в ч.у. множестве



- тогда $A^\Delta = \{c, d\}$ и $\sup A$ отсутствует.
- 2 Если S — совокупность подмножеств некоторого множества, то (по включению) $\sup S$ совпадает с объединением, а $\inf S$ — с пересечением всех подмножеств из S .

Решёточно упорядоченные (р.у.) множества:

Определение

Ч.у. множество, в котором для любых элементов a и b существуют $\inf \{a, b\}$ и $\sup \{a, b\}$ называют *решёточно упорядоченным*.

В конечном р.у. множестве \mathbf{P} точные грани существуют для любого подмножества элементов.

$\sup \mathbf{P} \stackrel{\text{def}}{=} 1$ — наибольший элемент (единица) P
 $\inf \mathbf{P} \stackrel{\text{def}}{=} 0$ — наименьший элемент (нуль) P
— *универсальные грани*.

Элементы, непосредственно следующие за нулём — *атомы*.

Решёточно упорядоченные (р.у.) множества:

Определение

Ч.у. множество, в котором для любых элементов a и b существуют $\inf \{a, b\}$ и $\sup \{a, b\}$ называют *решёточно упорядоченным*.

В конечном р.у. множестве \mathbf{P} точные грани существуют для любого подмножества элементов.

$\left. \begin{array}{l} \sup \mathbf{P} \stackrel{\text{def}}{=} \mathbf{1} \text{ — наибольший элемент (единица) } P \\ \inf \mathbf{P} \stackrel{\text{def}}{=} \mathbf{0} \text{ — наименьший элемент (нуль) } P \end{array} \right\}$
 — *универсальные грани*.

Элементы, непосредственно следующие за нулём — *атомы*.

Р.у. множество = (алгебраическая) решётка

Определение

Решёткой \mathbf{L} называется множество L с заданными на нём бинарными операциями: \vee (объединения) и \wedge (пересечения), подчиняющимися законам

$$\text{Com } \vee : x \vee y = y \vee x;$$

$$\text{Com } \wedge : x \wedge y = y \wedge x;$$

$$\text{Ass } \vee : x \vee (y \vee z) = (x \vee y) \vee z;$$

$$\text{Ass } \wedge : x \wedge (y \wedge z) = (x \wedge y) \wedge z;$$

$$\text{Id } \vee : x \vee x = x;$$

$$\text{Id } \wedge : x \wedge x = x,$$

$$\text{Abs1} : x \wedge (x \vee y) = x;$$

$$\text{Abs2} : x \vee (x \wedge y) = x,$$

$$a \vee b = \sup \{a, b\}; \quad a \wedge b = \inf \{a, b\}$$

Р.у. множество = (алгебраическая) решётка

Определение

Решёткой \mathbf{L} называется множество L с заданными на нём бинарными операциями: \vee (объединения) и \wedge (пересечения), подчиняющимися законам

$$\text{Com } \vee : x \vee y = y \vee x;$$

$$\text{Com } \wedge : x \wedge y = y \wedge x;$$

$$\text{Ass } \vee : x \vee (y \vee z) = (x \vee y) \vee z;$$

$$\text{Ass } \wedge : x \wedge (y \wedge z) = (x \wedge y) \wedge z;$$

$$\text{Id } \vee : x \vee x = x;$$

$$\text{Id } \wedge : x \wedge x = x,$$

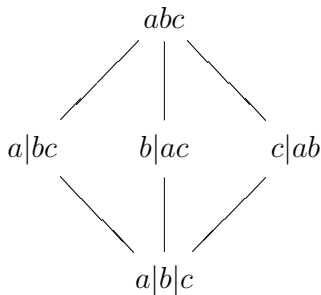
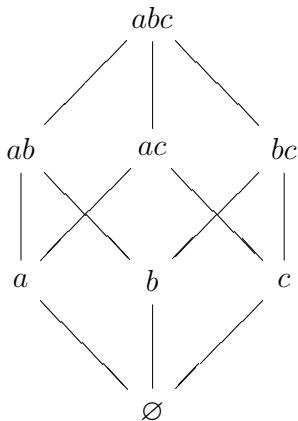
$$\text{Abs1} : x \wedge (x \vee y) = x;$$

$$\text{Abs2} : x \vee (x \wedge y) = x,$$

$$\mathbf{a} \vee \mathbf{b} = \sup \{\mathbf{a}, \mathbf{b}\}; \quad \mathbf{a} \wedge \mathbf{b} = \inf \{\mathbf{a}, \mathbf{b}\}$$

Решётки:

булеан и беллиан множества $\{a, b, c\}$



Отображения между ч.у. множествами

Нотация: φ и ψ — функции; $\varphi(x) = x\varphi$;

$$(\varphi * \psi)(x) = \psi(\varphi(x)) = x\varphi\psi.$$

Определение

Пусть \mathbf{P}_1 и \mathbf{P}_2 — ч.у. множества и $x, y \in P$.
Отображение $\varphi: P_1 \rightarrow P_2$ называется
соответственно

- *изотонным (порядковым гомоморфизмом)*, если $x \leq y \Rightarrow x\varphi \leq y\varphi$;
- *обратно изотонным*, если $x\varphi \leq y\varphi \Rightarrow x \leq y$;
- *антиизотонным*, если $x \leq y \Rightarrow x\varphi \geq y\varphi$.

Отображения между ч.у. множествами

Нотация: φ и ψ — функции; $\varphi(x) = x\varphi$;

$$(\varphi * \psi)(x) = \psi(\varphi(x)) = x\varphi\psi.$$

Определение

Пусть \mathbf{P}_1 и \mathbf{P}_2 — ч.у. множества и $x, y \in P$.
Отображение $\varphi: P_1 \rightarrow P_2$ называется
соответственно

- *изотонным (порядковым гомоморфизмом)*, если $x \leq y \Rightarrow x\varphi \leq y\varphi$;
- *обратно изотонным*, если $x\varphi \leq y\varphi \Rightarrow x \leq y$;
- *антиизотонным*, если $x \leq y \Rightarrow x\varphi \geq y\varphi$.

Оператор замыкания на ч.у. множестве

Определение

Изотонное отображение φ ч.у. множества \mathbf{P} в себя, такое что $x \leq x\varphi$ и $x\varphi = x\varphi\varphi$ называется *оператором замыкания* (на \mathbf{P}).

Если $x = x\varphi$, то элемент x называется *φ -замкнутым*.

Соответствия Галуа: определение

Определение

Пусть \mathbf{P} и \mathbf{Q} — ч.у. множества. Пара отображений (φ, ψ) , где $\varphi : P \rightarrow Q$, а $\psi : Q \rightarrow P$, удовлетворяющая свойствам

- 1 φ и ψ антиизотонны;
- 2 $\varphi\psi$ и $\psi\varphi$ — операторы замыкания (на \mathbf{P} и \mathbf{Q} соответственно).

называется *соответствием Галуа* между \mathbf{P} и \mathbf{Q} .

Справедливы и более сильные соотношения

$$p \leq q\psi \Leftrightarrow q \leq p\varphi \quad \text{и} \quad \varphi = \varphi\psi\varphi, \quad \psi = \psi\varphi\psi.$$

Раздел

- 1 Введение. Задача классификации и методы её решения
- 2 Частично упорядоченные множества, решётки и отображения
- 3 Решётка формальных понятий**
- 4 Классификация на основе гипотез

Понятие: философское отступление...

Понятие — это ...

... целостная совокупность суждений,
утверждающих об отличительных признаках вещи

Примеры:

искусство, наука, ...

Объём понятия — это ...

... *совокупность всех вещей*, обладающих
зафиксированными в данном понятии признаками

Примеры:

искусство: литература, живопись, архитектура, ...

наука: биология, физика, химия...

Понятие: философское отступление...

Понятие — это ...

... целостная совокупность суждений,
утверждающих об отличительных признаках вещи

Примеры:

искусство, наука, ...

Объём понятия — это ...

... *совокупность всех вещей*, обладающих
зафиксированными в данном понятии признаками

Примеры:

искусство: *литература, живопись, архитектура, ...*

наука: *биология, физика, химия...*

Понятие: философское отступление...

Понятие — это ...

... целостная совокупность суждений,
утверждающих об отличительных признаках вещи

Примеры:

искусство, наука, ...

Объём понятия — это ...

... *совокупность всех вещей*, обладающих
зафиксированными в данном понятии признаками

Примеры:

искусство: *литература, живопись, архитектура, ...*

наука: *биология, физика, химия...*

Понятие: философское отступление...

Понятие — это ...

... целостная совокупность суждений,
утверждающих об отличительных признаках вещи

Примеры:

искусство, наука, ...

Объём понятия — это ...

... *совокупность всех вещей*, обладающих
зафиксированными в данном понятии признаками

Примеры:

искусство: литература, живопись, архитектура, ...

наука: биология, физика, химия...

Понятие: философское отступление...

Понятие — это ...

... целостная совокупность суждений,
утверждающих об отличительных признаках вещи

Примеры:

искусство, наука, ...

Объём понятия — это ...

... *совокупность всех вещей*, обладающих
зафиксированными в данном понятии признаками

Примеры:

искусство: литература, живопись, архитектура, ...

наука: биология, физика, химия...

Понятие: философское отступление...

Содержание понятия — это ...

... *совокупность свойств*, присущих всем объектам данного понятия

Примеры:

искусство: *результат отражения действительности в форме чувственных образов, создание выразительных форм, ...*

наука: *познавательная деятельность, объективность, систематичность, ...*

Понятие: философское отступление...

Содержание понятия — это ...

... *совокупность свойств*, присущих всем объектам данного понятия

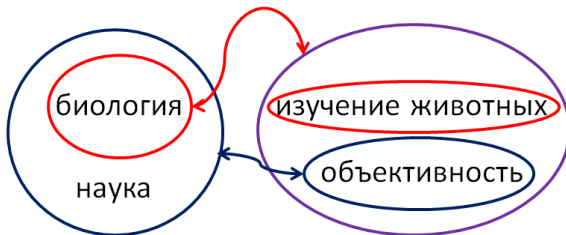
Примеры:

искусство: *результат отражения действительности в форме чувственных образов, создание выразительных форм, ...*

наука: *познавательная деятельность, объективность, систематичность, ...*

Закон обратного отношения между содержанием и объёмом понятия:

Большее по объёму понятие имеет меньшее содержание



Антимонотонность соответствий Галуа отражает этот закон

Классификация: положительные и отрицательные примеры

Рассматриваются задачи, в которых множество \mathcal{X} разбито на два непересекающихся класса \mathcal{X}^+ (*положительный*) и \mathcal{X}^- (*отрицательный*) относительно обладания/необладания их объектами некоторым *целевым признаком* $z \notin M$.

Прецеденты из данных классов называются, соответственно, *положительными* и *отрицательными примерами*.

Имеем 2 класса и $z = "x \in \mathcal{X}^+"$

АФП: формальный контекст

Пусть G и M — множества, называемые соответственно *множествами объектов* и *признаков*, а I — соответствие между G и M *отношением инцидентности*.

gIt означает, что объект $g \in G$ обладает признаком $t \in M$.

Определение

Тройка $K = (G, M, I)$ называется *формальным контекстом*.

В конечном случае контекст может быть задан виде объектно-признаковой $(0, 1)$ -матрицы.

Соответствия Галуа в АФП: нотация

Утверждение

Если для произвольных $A \subseteq G$ и $B \subseteq M$ ввести отображения $\varphi : 2^G \rightarrow 2^M$ и $\psi : 2^M \rightarrow 2^G$ такие, что

$$A\varphi = \{ m \subseteq M \mid \forall g \in A (gIm) \},$$

$$B\psi = \{ g \subseteq G \mid \forall m \in B (gIm) \},$$

то пара отображений (φ, ψ) является *соответствием Галуа между ч.у. множествами 2^G и 2^M* , упорядоченными по включению.

Множества $A\varphi$ и $B\psi$ традиционно записываются как A' и B' соответственно.

Формальные объём и содержание

Определение

Пусть дан контекст $K = (G, M, I)$. Пара подмножеств (A, B) , где $A \subseteq G$, а $B \subseteq M$, и таких, что $A' = B$ и $B' = A$, называется *формальным понятием* данного контекста с *формальным объёмом* A и *формальным содержанием* B .

Если контекст K представлен в виде объектно-признаковой $(0, 1)$ -матрицы, то формальному понятию соответствует **максимальная её подматрица, заполненная единицами.**

Решётка формальных понятий

Формальные объём и содержание — замкнутые, соответственно, относительно $\varphi\psi$ и $\psi\varphi$ множества.

Теорема

Множество всех формальных понятий данного контекста K образует **решётку**, обозначаемую $\mathfrak{B}(K)$, относительно операций \vee (объединение) и \wedge (пересечение):

$$\begin{aligned}(A_1, B_1) \vee (A_2, B_2) &= ((B_1 \cap B_2)', B_1 \cap B_2), \\ (A_1, B_1) \wedge (A_2, B_2) &= (A_1 \cap A_2, (A_1 \cap A_2)').\end{aligned}$$

$\mathfrak{B}(K)$ называется **решёткой формальных понятий** или **решёткой Галуа**.

Решётка формальных понятий...

У решётки $\mathfrak{B}(K)$ формального контекста
 $K = (G, M, I)$

- 1 — формальное понятие (G, G')
(возможен случай $G' = \emptyset$);
- атомы — формальные понятия вида (g, g') с
одноэлементными объёмами (однако,
не все $g \in G$ порождают атомы);
- 0 — формальное понятие (\emptyset, M) с пустым
объёмом.

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \leq A_2 \Leftrightarrow B_1 \geq B_2$$

Раздел

- 1 Введение. Задача классификации и методы её решения
- 2 Частично упорядоченные множества, решётки и отображения
- 3 Решётка формальных понятий
- 4 Классификация на основе гипотез**

Три контекста

Данные для обучения описываются положительным $K_+ = (G_+, M, I_+)$, отрицательным $K_- = (G_-, M, I_-)$ и недоопределённым $K_\tau = (G_\tau, M, I_\tau)$ контекстами.

Операторы Галуа в этих контекстах обозначаются соответствующими **верхними индексами**: например, A^+ , A^- , A^τ и т.д.

Положительные и отрицательные объём и содержание

Определение

Формальное понятие $(A_+, B_+) \in K_+$ называется *положительным*.

A_+ — *положительный формальный объём*,

B_+ — *положительное формальное содержание*.

Аналогично определяются *отрицательные* и *недоопределённые* понятия, *формальные объём* и *содержание* для контекстов K_- и K_τ .

Гипотезы

Определение

Положительное формальное содержание B_+ положительного понятия (A_+, B_+) называется:

- *положительной предгипотезой*, если $\forall (A_-, B_-) \in K_- (B_+ \neq B_-)$, т.е. оно не является формальным содержанием ни одного отрицательного понятия;
- *положительной гипотезой*, если $\forall (g, g^-) \in K_- (B_+ \not\subseteq g^-)$, т.е. оно не является подмножеством содержания понятия какого-либо отрицательного примера g ;
- *фальсифицированной положительной гипотезой*, если $\exists (g, g^-) \in K_- (B_+ \subseteq g^-)$.

Гипотезы

Определение

Положительное формальное содержание B_+ положительного понятия (A_+, B_+) называется:

- *положительной предгипотезой*, если $\forall (A_-, B_-) \in K_- (B_+ \neq B_-)$, т.е. оно не является формальным содержанием ни одного отрицательного понятия;
- *положительной гипотезой*, если $\forall (g, g^-) \in K_- (B_+ \not\subseteq g^-)$, т.е. оно не является подмножеством содержания понятия какого-либо отрицательного примера g ;
- *фальсифицированной положительной гипотезой*, если $\exists (g, g^-) \in K_- (B_+ \subseteq g^-)$.

Гипотезы

Определение

Положительное формальное содержание B_+ положительного понятия (A_+, B_+) называется:

- *положительной предгипотезой*, если $\forall (A_-, B_-) \in K_- (B_+ \neq B_-)$, т.е. оно не является формальным содержанием ни одного отрицательного понятия;
- *положительной гипотезой*, если $\forall (g, g^-) \in K_- (B_+ \not\subseteq g^-)$, т.е. оно не является подмножеством содержания понятия какого-либо отрицательного примера g ;
- *фальсифицированной положительной гипотезой*, если $\exists (g, g^-) \in K_- (B_+ \subseteq g^-)$.

Гипотезы

Определение

Положительное формальное содержание B_+ положительного понятия (A_+, B_+) называется:

- *положительной предгипотезой*, если $\forall (A_-, B_-) \in K_- (B_+ \neq B_-)$, т.е. оно не является формальным содержанием ни одного отрицательного понятия;
- *положительной гипотезой*, если $\forall (g, g^-) \in K_- (B_+ \not\subseteq g^-)$, т.е. оно не является подмножеством содержания понятия какого-либо отрицательного примера g ;
- *фальсифицированной положительной гипотезой*, если $\exists (g, g^-) \in K_- (B_+ \subseteq g^-)$.

Гипотезы...

Отрицательные (или $(-)$ -)

- *предгипотезы, гипотезы,*
- *фальсифицированные гипотезы,*

определяются аналогично.

- Гипотеза является также и предгипотезой.
- Гипотезы используются для **классификации недоопределённых примеров** из множества G_T .

Простейшее решающее правило

Неопределённый объект $g \in G_T$ относится к **положительному** классу, если g^T содержит **хотя бы одну положительную гипотезу и не содержит ни одной из отрицательных гипотез** (к отрицательному классу — аналогично).

Отказ от классификации происходит, если формальное содержание g^T :

- **не включает** в себя ни положительных, ни отрицательных гипотез (недостаток обобщений);
- **содержит** как положительные, так и отрицательные гипотезы (противоречие в гипотезах).

Простейшее решающее правило

Неопределённый объект $g \in G_T$ относится к **положительному** классу, если g^T содержит **хотя бы одну положительную гипотезу и не содержит ни одной из отрицательных гипотез** (к отрицательному классу — аналогично). Отказ от классификации происходит, если формальное содержание g^T :

- **не включает** в себя ни положительных, ни отрицательных гипотез (недостаток обобщений);
- **содержит** как положительные, так и отрицательные гипотезы (противоречие в гипотезах).

Многозначные контексты

В АФП предполагается **двоичной** информации о признаках.

Для её получения из количественных и качественных признаков используется процедура **шкалирования**.

Многозначный контекст — четвёрка (G, M, Z, I) , где

- G, M, Z — множества объектов, признаков и значений признаков соответственно,
- I — тернарное отношение $I \subseteq G \times M \times Z$, задающее значение $z \in Z$ признака $m \in M$ объекта $g \in G$,

причем отображение $G \times M \rightarrow Z$ **функционально**.

Шкалирование — это...

... представление многозначных контекстов
двузначными

Примеры: шкалы номинальная, порядковая,
дихотомичная (булева), ...

Пример «Фрукты»: постановка задачи

Задача:

построить классификатор по целевому свойству $z = \text{«являться фруктом»}$ и следующей объектно-признаковой таблице примеров:

№	G \ M	цвет	жёсткий	гладкий	форма	фрукт
1	яблоко	жёлтое	нет	да	круглое	+
2	грейпфрут	жёлтый	нет	нет	круглый	+
3	киви	зелёное	нет	нет	овальное	+
4	слива	синяя	нет	да	овальная	+
5	кубик	зелёный	да	да	кубический	—
6	яйцо	белое	да	да	овальное	—
7	теннисный мяч	белый	нет	нет	круглый	—

Пример «Фрукты»: результат шкалирования

G \ M	w	y	g	b	f	\bar{f}	s	\bar{s}	r	\bar{r}	Фрукт
1		x				x	x		x	\bar{r}	+
2		x				x		x	x		+
3			x			x		x		x	+
4				x		x	x			x	+
5			x		x		x			x	-
6	x				x		x			x	-
7	x					x		x			-

$$G_+ = \{1, 2, 3, 4\}, \quad G_- = \{5, 6, 7\}.$$

Отношение I_+ [I_-] представлено верхней [нижней] частью таблицы.

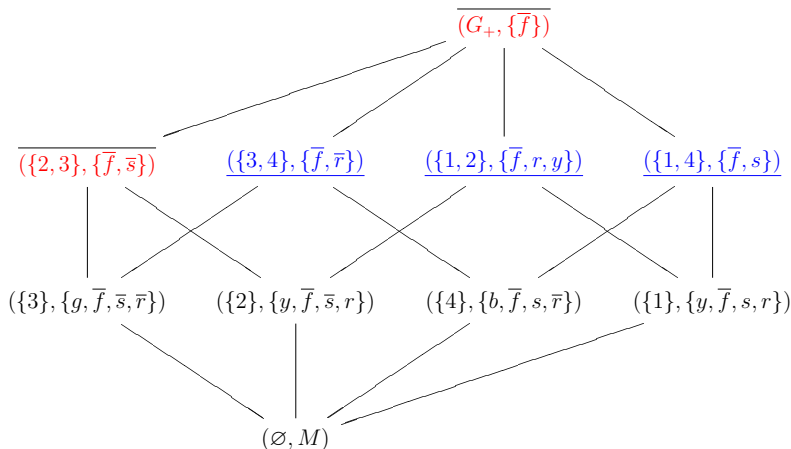
Признаки означают: w — белый, y — жёлтый,

g — зелёный, b — синий,

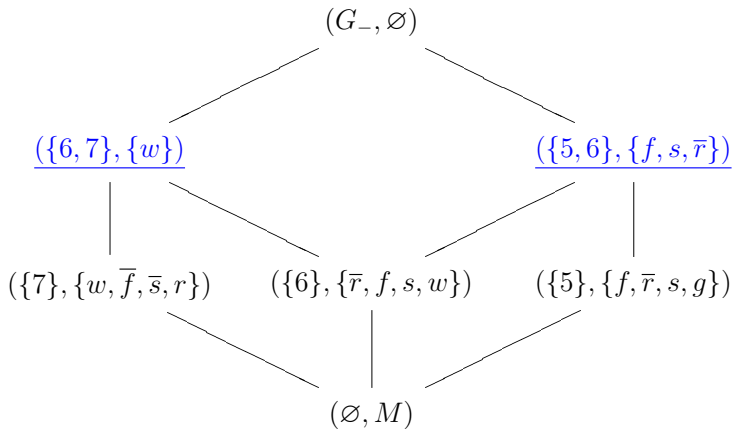
f — твёрдый, \bar{f} — мягкий, s — гладкий,

\bar{s} — шероховатый, r — круглый, \bar{r} — некруглый.

Пример «Фрукты»: решётка $\mathfrak{B}(K_+)$ положительного контекста



Пример «Фрукты»: решётка $\mathfrak{B}(K_-)$ отрицательного контекста



Пример «Фрукты»: формирование гипотез

Формальные содержания

- $\{\bar{f}, \bar{r}\}$ (мягкий, некруглый),
 $\{\bar{f}, r, y\}$ (мягкий, круглый, жёлтый) и
 $\{\bar{f}, s\}$ (мягкий, гладкий) являются
+-гипотезами;
- $\{\bar{f}, \bar{s}\}$ (мягкий, шероховатый) является
фальсифицированной +-гипотезой, т.к.
она — часть содержания $\{w, \bar{f}, \bar{s}, r\}$
отрицательного примера 7 (теннисный мяч);
- $\{w\}$ (белый) и
 $\{f, s, \bar{r}\}$ (твёрдый, гладкий, некруглый)
являются --гипотезами.

Пример «Фрукты»: формирование гипотез

Формальные содержания

- $\{\bar{f}, \bar{r}\}$ (мягкий, некруглый),
 $\{\bar{f}, r, y\}$ (мягкий, круглый, жёлтый) и
 $\{\bar{f}, s\}$ (мягкий, гладкий) являются
+-гипотезами;
- $\{\bar{f}, \bar{s}\}$ (мягкий, шероховатый) является
фальсифицированной +-гипотезой, т.к.
она — часть содержания $\{w, \bar{f}, \bar{s}, r\}$
отрицательного примера 7 (теннисный мяч);
- $\{w\}$ (белый) и
 $\{f, s, \bar{r}\}$ (твёрдый, гладкий, некруглый)
являются --гипотезами.

Пример «Фрукты»: формирование гипотез

Формальные содержания

- $\{\bar{f}, \bar{r}\}$ (мягкий, некруглый),
 $\{\bar{f}, r, y\}$ (мягкий, круглый, жёлтый) и
 $\{\bar{f}, s\}$ (мягкий, гладкий) являются
+-гипотезами;
- $\{\bar{f}, \bar{s}\}$ (мягкий, шероховатый) является
фальсифицированной +-гипотезой, т.к.
она — часть содержания $\{w, \bar{f}, \bar{s}, r\}$
отрицательного примера 7 (теннисный мяч);
- $\{w\}$ (белый) и
 $\{f, s, \bar{r}\}$ (твёрдый, гладкий, некруглый)
являются --гипотезами.

Пример «Фрукты»: формирование гипотез

Формальные содержания

- $\{\bar{f}, \bar{r}\}$ (мягкий, некруглый),
 $\{\bar{f}, r, y\}$ (мягкий, круглый, жёлтый) и
 $\{\bar{f}, s\}$ (мягкий, гладкий) являются
 +-гипотезами;
- $\{\bar{f}, \bar{s}\}$ (мягкий, шероховатый) является
 фальсифицированной +-гипотезой, т.к.
 она — часть содержания $\{w, \bar{f}, \bar{s}, r\}$
 отрицательного примера 7 (теннисный мяч);
- $\{w\}$ (белый) и
 $\{f, s, \bar{r}\}$ (твёрдый, гладкий, некруглый)
 являются --гипотезами.

Пример «Фрукты»: классификация

Неопределённый объект g

- **мирабель** будет классифицирован как *фрукт*, т.к. его формальное содержание *жёлтый, мягкий, гладкий* ($\{y, \bar{f}, s\}$) содержит *положительную гипотезу* $\{\bar{f}, s\}$ и не содержит ни одной из отрицательных гипотез;
- **кусочек сахара** со свойствами *белый, некруглый, твёрдый* будет классифицирован как *не-фрукт*;
- **брикет пломбира** со свойствами *белый, мягкий, некруглый* вызовет отказ от классификации, поскольку $g^T = \{w, \bar{f}, \bar{r}\}$ содержит как *положительную гипотезу* $\{\bar{f}, \bar{r}\}$, так и *отрицательную гипотезу* $\{w\}$.

Пример «Фрукты»: классификация

Неопределённый объект g

- **мирабель** будет классифицирован как **фрукт**, т.к. его формальное содержание **жёлтый, мягкий, гладкий** ($\{y, \bar{f}, s\}$) содержит **положительную гипотезу** $\{\bar{f}, s\}$ и не содержит ни одной из отрицательных гипотез;
- **кусочек сахара** со свойствами **белый, некруглый, твёрдый** будет классифицирован как **не-фрукт**;
- **брикет пломбира** со свойствами **белый, мягкий, некруглый** вызовет отказ от классификации, поскольку $g^T = \{w, \bar{f}, \bar{r}\}$ содержит как **положительную гипотезу** $\{\bar{f}, \bar{r}\}$, так и **отрицательную гипотезу** $\{w\}$.

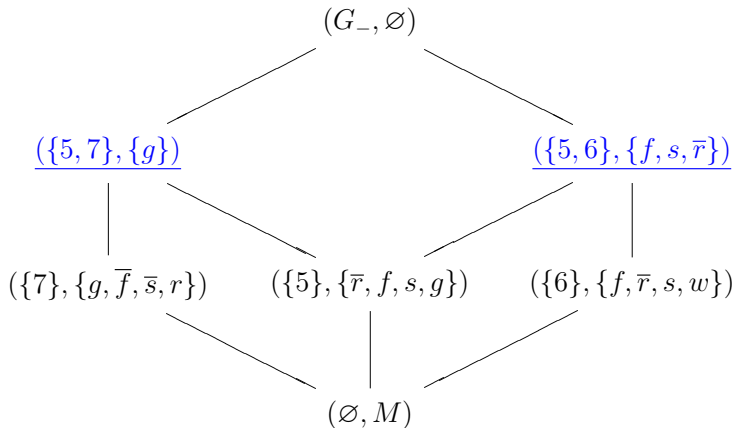
Пример «Фрукты»: классификация

Неопределённый объект g

- **мирабель** будет классифицирован как *фрукт*, т.к. его формальное содержание *жёлтый, мягкий, гладкий* ($\{y, \bar{f}, s\}$) содержит положительную гипотезу $\{\bar{f}, s\}$ и не содержит ни одной из отрицательных гипотез;
- **кусочек сахара** со свойствами *белый, некруглый, твёрдый* будет классифицирован как *не-фрукт*;
- **брикет пломбира** со свойствами *белый, мягкий, некруглый* вызовет отказ от классификации, поскольку $g^T = \{w, \bar{f}, \bar{r}\}$ содержит как положительную гипотезу $\{\bar{f}, \bar{r}\}$, так и отрицательную гипотезу $\{w\}$.

Пример «Фрукты»: дополнение

Если считать, что теннисный мяч — зелёный, то $\mathfrak{B}(K_-)$:



Пример «Фрукты»: дополнение...

При таком изменении свойств объекта № 7 изменятся только **отрицательные гипотезы**; они будут иметь формальные содержания $\{g\} = \{5, 7\}'$ и $\{f, s, \bar{r}\} = \{5, 6\}'$.

В силу этого,

- объекты со свойствами *жёлтый, мягкий, гладкий* и *белый, мягкий, некруглый* будет классифицированы как *фрукт*;
- на объекте с единственным свойством *белый* произойдёт **отказ от классификации**.

И в заключение:

Следует иметь в виду, что рассмотренный выше весьма простой классифицирующий алгоритм обычно даёт

значительную долю отказов от классификации,
как правило, неприемлемую на практике.

В то же время в литературе можно найти удачные примеры применения АФП для решения некоторых задач распознавания.

Направления развития АФП:

- методы формирования гипотез;
- бикластеризация;
- ...

И в заключение:

Следует иметь в виду, что рассмотренный выше весьма простой классифицирующий алгоритм обычно даёт

значительную долю отказов от классификации,
как правило, неприемлемую на практике.

В то же время в литературе можно найти удачные примеры применения АФП для решения некоторых задач распознавания.

Направления развития АФП:

- методы формирования гипотез;
- бикластеризация;
- ...

Что в сухом остатке?

- АФП — эффективный метод решения задач обработки данных, основанный на теории решёток и соответствиях Галуа.
- АФП — направление в реляционном подходе к задаче классификации.
- АФП допускает развитие в разных направлениях.

Спасибо за внимание

Вопросы?