Логический анализ данных в распознавании. Лекция № 9 Логические корректоры

д.ф.-м.н. Елена Всеволодовна Дюкова
edjukova@mail.ru
к.ф.-м.н. Пётр Александрович Прокофьев
p_prok@mail.ru, https://t.me/pprok

МГУ, Москва

24 апреля 2024 г.

Содержание лекции

Свойства элементарных классификаторов

Элементарный классификатор

Определение

Элементарным классификатором (эл.кл.) ранга r, $1 \leqslant r \leqslant n$, *называется пара* (H, σ), *где* $H = (x_{j_1}, \ldots, x_{j_r})$ — *набор различных признаков и* $\sigma = (\sigma_1, \ldots, \sigma_r)$ — *вектор, в котором* σ_t — *допустимое значение признака* x_{j_t}, $t \in \{1, \ldots, r\}$.

- $H(S) = (x_{j_1}(S), \ldots, x_{j_r}(S))$ — признаковое подописание объекта S
- эл.кл. (H, σ), выделяет объект, если $H(S) = \sigma$, т.е. фактически задаёт конъюнкцию $[[x_{j_1}(S) = \sigma_1 \wedge \ldots \wedge x_{j_r}(S) = \sigma_r]]$

Элементарный классификатор

Определение

Элементарным классификатором (эл.кл.) ранга r, $1 \leqslant r \leqslant n$, *называется пара* (H, σ), *где* $H = (x_{j_1}, \ldots, x_{j_r})$ — *набор различных признаков и* $\sigma = (\sigma_1, \ldots, \sigma_r)$ — *вектор, в котором* σ_t — *допустимое значение признака* x_{j_t}, $t \in \{1, \ldots, r\}$.

- $H(S) = (x_{j_1}(S), \ldots, x_{j_r}(S))$ — признаковое подописание объекта S
- эл.кл. (H, σ), выделяет объект, если $H(S) = \sigma$, т.е. фактически задаёт конъюнкцию $[[x_{j_1}(S) = \sigma_1 \wedge \ldots \wedge x_{j_r}(S) = \sigma_r]]$

Корректный элементарный классификатор

Определение

Эл.кл. (H, σ) называется корректным для класса K, если не существует двух прецедентов S_i и S_t, выделяемых эл.кл. (H, σ), таких, что $S_i \in K$ и $S_t \notin K$.

Определение

Корректный для класса K эл.кл. (H, σ) ранга r называется тупиковым, если для любого $k \in \{1, \ldots, r\}$ и $H' = (x_{j_1}, \ldots, x_{j_{k-1}}, x_{j_{k+1}}, \ldots, x_{j_r})$, $\sigma' = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_r)$ эл.кл. (H', σ') не является корректным для K.

Корректный элементарный классификатор

Определение

Эл.кл. (H, σ) *называется корректным для класса* K, *если не существует двух прецедентов* S_i *и* S_t, *выделяемых эл.кл.* (H, σ), *таких, что* $S_i \in K$ *и* $S_t \notin K$.

Определение

Корректный для класса K *эл.кл.* (H, σ) *ранга* r *называется тупиковым, если для любого* $k \in \{1, \ldots, r\}$ *и* $H' = (x_{j_1}, \ldots, x_{j_{k-1}}, x_{j_{k+1}}, \ldots, x_{j_r})$, $\sigma' = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_r)$ *эл.кл.* (H', σ') *не является корректным для* K.

Представительные наборы

Определение

Представительным набором класса K называется эл.кл. (H, σ), который

1. *выделяет хотя бы один прецедент из K и*
2. *не выделяет ни одного прецедента не из K.*

Представительные наборы — один из видов корректных эл.кл.

Определение

Представительный для класса K набор (H, σ) ранга r называется тупиковым, если для любого $k \in \{1, \ldots, r\}$ и $H' = (x_{j_1}, \ldots, x_{j_{k-1}}, x_{j_{k+1}}, \ldots, x_{j_r})$, $\sigma' = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_r)$ эл.кл. (H', σ') не является представительным для K набором.

Представительные наборы

Определение

Представительным набором класса K называется эл.кл. (H, σ), который

1. *выделяет хотя бы один прецедент из K и*
2. *не выделяет ни одного прецедента не из K.*

Представительные наборы — один из видов корректных эл.кл.

Определение

Представительный для класса K набор (H, σ) ранга r называется тупиковым, если для любого $k \in \{1, \ldots, r\}$ и $H' = (x_{j_1}, \ldots, x_{j_{k-1}}, x_{j_{k+1}}, \ldots, x_{j_r})$, $\sigma' = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_r)$ эл.кл. (H', σ') не является представительным для K набором.

Алгоритм голосования по представительным наборам

Построение алгоритма голосования по представительным наборам включает два этапа:

Голосование по представительным наборам

- На этапе обучения для каждого класса $K \in \{K_1, \ldots, K_l\}$ строится семейство C_K представительных наборов
- При распознавании объекта S для каждого класса $K \in \{K_1, \ldots, K_l\}$ вычисляется оценка принадлежности S к K

$$\Gamma(S, K) = \sum_{(H,\sigma) \in C_K} w_{(H,\sigma)} [[H(S) = \sigma]],$$

Объект S относится к классу с номером

$$y^* = \arg\max_y \Gamma(S, K_y)$$

Алгоритм голосования по представительным наборам

Построение алгоритма голосования по представительным наборам включает два этапа:

Голосование по представительным наборам

- На этапе обучения для каждого класса $K \in \{K_1, \ldots, K_l\}$ строится семейство C_K представительных наборов
- При распознавании объекта S для каждого класса $K \in \{K_1, \ldots, K_l\}$ вычисляется оценка принадлежности S к K

$$\Gamma(S, K) = \sum_{(H,\sigma) \in C_K} w_{(H,\sigma)} [[H(S) = \sigma]],$$

Объект S относится к классу с номером

$$y^* = \arg\max_y \Gamma(S, K_y)$$

Преимущества и недостатки голосования по представительным наборам

- **Преимущества**
  - \+ Корректность алгоритма обеспечивается тем, представительные наборы голосуют только за объекты «своего» класса
  - \+ Хорошая интерпретируемость распознающего алгоритма:
    - голосование по конъюнкциям
    - вычисление оценки близости объекта к прецедентам
  - \+ Ограничение ранга эл.кл. позволяет ускорить обучение (например, в алгоритме «Кора» не больше 3 признаков)
- **Недостатки**
  - Среди эк.кл. небольшого ранга может не оказаться корректных
  - В общем случае этап обучения связан с решением вычислительно сложных задач

Преимущества и недостатки голосования по представительным наборам

- **Преимущества**
  - \+ Корректность алгоритма обеспечивается тем, представительные наборы голосуют только за объекты «своего» класса
  - \+ Хорошая интерпретируемость распознающего алгоритма:
    - голосование по конъюнкциям
    - вычисление оценки близости объекта к прецедентам
  - \+ Ограничение ранга эл.кл. позволяет ускорить обучение (например, в алгоритме «Кора» не больше 3 признаков)
- **Недостатки**
  - \- Среди эк.кл. небольшого ранга может не оказаться корректных
  - \- В общем случае этап обучения связан с решением вычислительно сложных задач

Предикаты на множестве объектов

- $B : M \to \{0, 1\}$ — предикат на множестве объектов (множестве векторов значений признаков)
- Если $B(S) = 1$, то говорят, что предикат B выделяет объект S
- Например, эл.кл. (H, σ) определяет предикат

$$B(S) = [[H(S) = \sigma]]$$

- Для класса K обозначим:

$$P_K = |T \cap K|, \qquad N_K = |T \setminus K|,$$

$$p_K(B) = \sum_{S_i \in T \cap K} B(S_i), \qquad n_K(B) = \sum_{S_i \in T \setminus K} B(S_i)$$

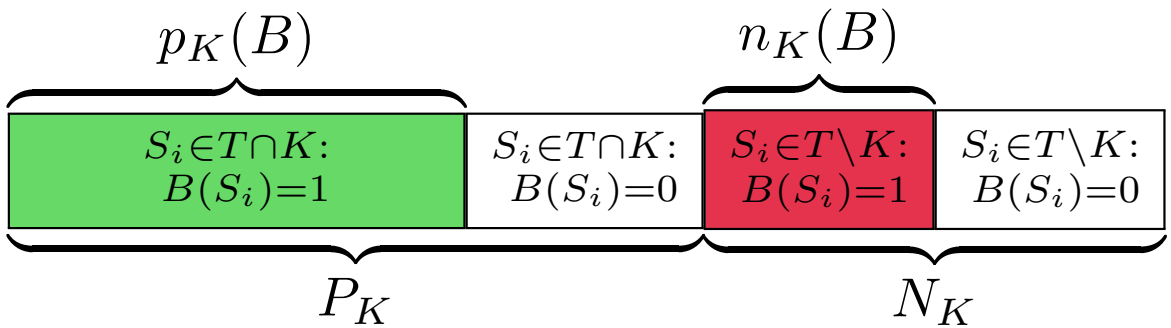
Предикаты на множестве объектов

- $B : M \to \{0, 1\}$ — предикат на множестве объектов (множестве векторов значений признаков)
- Если $B(S) = 1$, то говорят, что предикат B выделяет объект S
- Например, эл.кл. (H, σ) определяет предикат

$$B(S) = [[H(S) = \sigma]]$$

- Для класса K обозначим:

$$P_K = |T \cap K|, \qquad N_K = |T \setminus K|,$$
$$p_K(B) = \sum_{S_i \in T \cap K} B(S_i), \qquad n_K(B) = \sum_{S_i \in T \setminus K} B(S_i)$$



Логические закономерности

- Предикат называется *логической закономерностью*, если он:
  - записывается на естественном языке и обычно зависит от небольшого числа признаков (*интерпретируемость*)
  - выделяет преимущественно объекты одного класса (*информативность*):

$$p_K(B) \to \max, \quad n_K(B) \to \min, \quad \frac{p_K(B)}{P_K} \gg \frac{n_K(B)}{N_K}$$

- В логическом подходе часто от закономерностей требуется *корректность* (непротиворечивость): $n_K(B) = 0$
- Также применяются «антипредставительные» предикаты со свойством $p_K(B) = 0$, $n_K(B) > 0$, которые также называются корректными

Логические закономерности

- Предикат называется *логической закономерностью*, если он:
  - записывается на естественном языке и обычно зависит от небольшого числа признаков (*интерпретируемость*)
  - выделяет преимущественно объекты одного класса (*информативность*):

$$p_K(B) \to \max, \quad n_K(B) \to \min, \quad \frac{p_K(B)}{P_K} \gg \frac{n_K(B)}{N_K}$$

- В логическом подходе часто от закономерностей требуется *корректность* (непротиворечивость): $n_K(B) = 0$
- Также применяются «антипредставительные» предикаты со свойством $p_K(B) = 0$, $n_K(B) > 0$, которые также называются корректными

Метрики информативности предикатов

Оптимизацию информативности предиката удобнее осуществлять, когда есть одна метрика $I_K(B)$ вместо двух:

$$p_K(B) \to \max, \quad n_K(B) \to \min$$

Примеры хороших метрик информативности:

- **Статистическая информативность**:

$$I_K(B) = -\ln h_{P_K,N_K}(p_K(B), n_K(B)), \text{ где } h_{P,N}(p,n) = \frac{C_P^p C_N^n}{C_{P+N}^{p+n}},$$

согласно *точному тесту Фишера*, при небольших значениях вероятности $h_{P_K,N_K}(p_K(B), n_K(B))$ отвергается гипотеза независимости случайных индикаторов $[[S_i \in K]]$ и $[[B(S_i) = 1]]$ при равномерном выборе S_i из обучающей выборки T.

Метрики информативности предикатов

Оптимизацию информативности предиката удобнее осуществлять, когда есть одна метрика $I_K(B)$ вместо двух:

$$p_K(B) \to \max, \quad n_K(B) \to \min$$

Примеры хороших метрик информативности:

- **Статистическая информативность**:

$$I_K(B) = -\ln h_{P_K,N_K}(p_K(B), n_K(B)), \text{ где } h_{P,N}(p,n) = \frac{C_P^p C_N^n}{C_{P+N}^{p+n}},$$

согласно *точному тесту Фишера*, при небольших значениях вероятности $h_{P_K,N_K}(p_K(B), n_K(B))$ отвергается гипотеза независимости случайных индикаторов $[[S_i \in K]]$ и $[[B(S_i) = 1]]$ при равномерном выборе S_i из обучающей выборки T.

Метрики информативности предикатов (2)

- **Энтропийная информативность (IGain)**:

$$I_K(B) = \hat{H}(P_K, N_K) - \hat{H}_B(P_K, N_K, p_K(B), n_K(B)),$$

где $\hat{H}(P, N) = H\left(\frac{P}{P+N}, \frac{N}{P+N}\right)$ — энтропия выборки,

$$H(p, q) = -p \log p - q \log q,$$

$$\hat{H}_B(P, N, p, n) = \frac{p+n}{P+N}\hat{H}(p, n) + \frac{P+N-p-n}{P+N}\hat{H}(P-p, N-n)$$

— энтропия с информацией о том, какие объекты выделяет $B(S)$

- **Информативность для *бустинга***

$$I_K(B) = \sqrt{p_K(B)} - \sqrt{n_K(B)}$$

Метрики информативности предикатов (2)

- **Энтропийная информативность (IGain)**:

$$I_K(B) = \hat{H}(P_K, N_K) - \hat{H}_B(P_K, N_K, p_K(B), n_K(B)),$$

где $\hat{H}(P, N) = H\left(\frac{P}{P+N}, \frac{N}{P+N}\right)$ — энтропия выборки,

$$H(p, q) = -p \log p - q \log q,$$

$$\hat{H}_B(P, N, p, n) = \frac{p+n}{P+N}\hat{H}(p, n) + \frac{P+N-p-n}{P+N}\hat{H}(P-p, N-n)$$

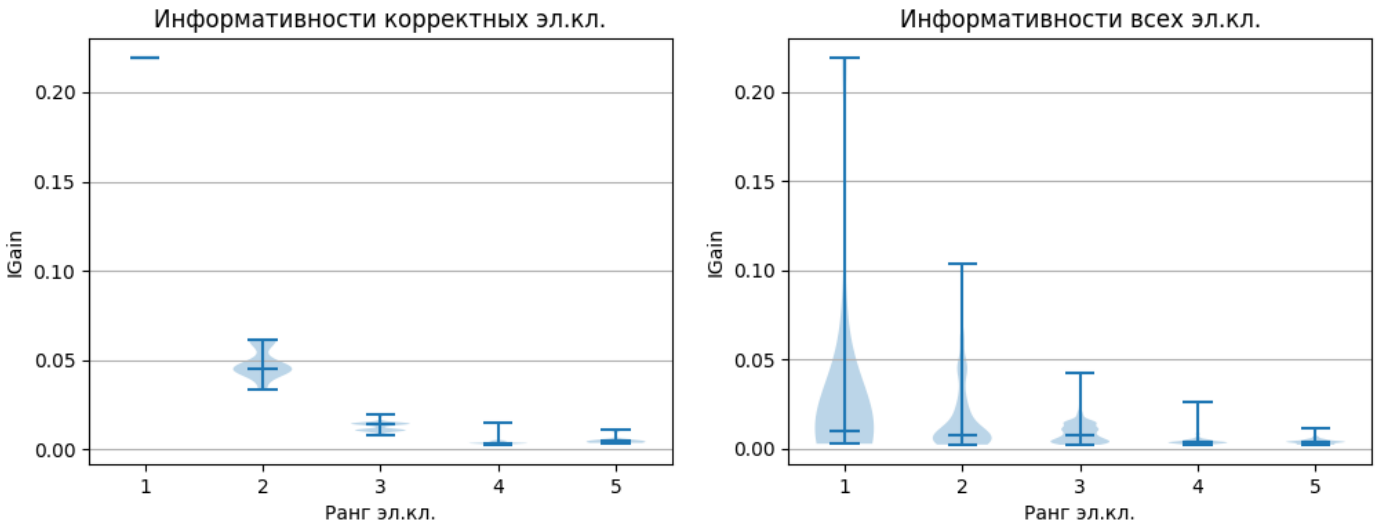
— энтропия с информацией о том, какие объекты выделяет $B(S)$

- **Информативность для *бустинга***

$$I_K(B) = \sqrt{p_K(B)} - \sqrt{n_K(B)}$$

Связь ранга, корректности и информативности эл.кл.

- Рассмотрим прикладную задачу **Car Evaluation** из репозитория **UCI** (6 признаков, 1728 объектов, 4 класса).
- Отберем эл.кл. с информативностью IGain выше порога $I_0 = 0,002$ (всего: 3950 эл.кл., из них 1631 корректных эл.кл.)



Корректные vs некорректные эл.кл.

- В прикладных задачах существенная часть корректных эл.кл. имеет большой ранг, что усложняет алгоритм и приводит к *переобучению* (плохо распознает те объекты, которые не встречаются в выборке)
- Среди некорректных эл.кл. существенная доля достаточно информативных, согласно общепринятым и теоретически обоснованнным метрикам информативности
- Корректные эл.кл., как правило, малоинформативны в случае *большой значности признаков*, посколько каждый эл.кл. встречается в небольшой доле прецедентов
- Попробуем отказаться от непротиворечивости эл.кл. и скорректировать выделяемые ими объекты в совокупности

Корректные vs некорректные эл.кл.

- В прикладных задачах существенная часть корректных эл.кл. имеет большой ранг, что усложняет алгоритм и приводит к *переобучению* (плохо распознает те объекты, которые не встречаются в выборке)
- Среди некорректных эл.кл. существенная доля достаточно информативных, согласно общепринятым и теоретически обоснованнным метрикам информативности
- Корректные эл.кл., как правило, малоинформативны в случае *большой значности признаков*, посколько каждый эл.кл. встречается в небольшой доле прецедентов
- Попробуем отказаться от непротиворечивости эл.кл. и скорректировать выделяемые ими объекты в совокупности

Корректные vs некорректные эл.кл.

- В прикладных задачах существенная часть корректных эл.кл. имеет большой ранг, что усложняет алгоритм и приводит к *переобучению* (плохо распознает те объекты, которые не встречаются в выборке)
- Среди некорректных эл.кл. существенная доля достаточно информативных, согласно общепринятым и теоретически обоснованнным метрикам информативности
- Корректные эл.кл., как правило, малоинформативны в случае *большой значности признаков*, посколько каждый эл.кл. встречается в небольшой доле прецедентов
- Попробуем отказаться от непротиворечивости эл.кл. и скорректировать выделяемые ими объекты в совокупности

Корректные vs некорректные эл.кл.

- В прикладных задачах существенная часть корректных эл.кл. имеет большой ранг, что усложняет алгоритм и приводит к *переобучению* (плохо распознает те объекты, которые не встречаются в выборке)
- Среди некорректных эл.кл. существенная доля достаточно информативных, согласно общепринятым и теоретически обоснованнным метрикам информативности
- Корректные эл.кл., как правило, малоинформативны в случае *большой значности признаков*, посколько каждый эл.кл. встречается в небольшой доле прецедентов
- Попробуем отказаться от непротиворечивости эл.кл. и скорректировать выделяемые ими объекты в совокупности

Алгебро-логическая коррекция

Алгебраический подход

- Решается задача распознавания объектов множества M на классы с номерами из Y. На этапе обучения по выборке T строится алгоритм $A_T : M \to Y$
- В *алгебраическом* подходе (*Журавлев, 1978*) строится набор операций $\{B_1, \ldots, B_p\}$, для которого можно найти «хорошую» *корректирующую функцию* $F : R^p \to R$ и «простое» *решающее правило* $C : R \to Y$, R — множество оценок :

$$\begin{array}{ccc} M & \xrightarrow{A_T} & Y \\ {\scriptstyle B_1,\ldots,B_p}\downarrow & & \uparrow{\scriptstyle C} \\ R^p & \xrightarrow[F(B_1,\ldots,B_p)]{} & R \end{array}$$

- Алгоритм распознавания — это композиция

$$A_T(S) = C(F(B_1(S), \ldots, B_p(S)))$$

Алгебраический подход

- Решается задача распознавания объектов множества M на классы с номерами из Y. На этапе обучения по выборке T строится алгоритм $A_T : M \to Y$
- В *алгебраическом* подходе (*Журавлев, 1978*) строится набор операций $\{B_1, \ldots, B_p\}$, для которого можно найти «хорошую» *корректирующую функцию* $F : R^p \to R$ и «простое» *решающее правило* $C : R \to Y$, R — множество оценок :

$$\begin{array}{ccc} M & \xrightarrow{A_T} & Y \\ {\scriptstyle B_1,\ldots,B_p}\downarrow & & \uparrow{\scriptstyle C} \\ R^p & \xrightarrow[F(B_1,\ldots,B_p)]{} & R \end{array}$$

- Алгоритм распознавания — это композиция

$$A_T(S) = C(F(B_1(S), \ldots, B_p(S)))$$

Алгебраический подход

- Решается задача распознавания объектов множества M на классы с номерами из Y. На этапе обучения по выборке T строится алгоритм $A_T : M \to Y$
- В *алгебраическом* подходе (*Журавлев, 1978*) строится набор операций $\{B_1, \ldots, B_p\}$, для которого можно найти «хорошую» *корректирующую функцию* $F : R^p \to R$ и «простое» *решающее правило* $C : R \to Y$, R — множество оценок :

$$\begin{array}{ccc} M & \xrightarrow{A_T} & Y \\ {\scriptstyle B_1,\ldots,B_p}\downarrow & & \uparrow{\scriptstyle C} \\ R^p & \xrightarrow[F(B_1,\ldots,B_p)]{} & R \end{array}$$

- Алгоритм распознавания — это композиция

$$A_T(S) = C(F(B_1(S), \ldots, B_p(S)))$$

Алгебро-логический подход

- Рассмотрим в качестве операций предикаты $\{B_1, \ldots, B_p\}$, заданные эл.кл., $B_k : M \to \{0, 1\}$
- Корректирующие функции есть возможность выбирать среди булевых функций
$$F : \{0, 1\}^p \to \{0, 1\}.$$
- Распознающие процедуры, построенные путем коррекции эл.кл., называются *логическими корректорами*
- Корректное распознавание на базе произвольных, не обязательно корректных эл.кл., основано на идее *алгебро-логического подхода*, которая предложена в (*Дюкова, Журавлёв, Рудаков, 1996*).
Зарубежный аналог — Logical Analysis of Data (LAD) (*Boros, Hammer et al., 2000*).

Алгебро-логический подход

- Рассмотрим в качестве операций предикаты $\{B_1, \ldots, B_p\}$, заданные эл.кл., $B_k : M \to \{0, 1\}$
- Корректирующие функции есть возможность выбирать среди булевых функций
$$F : \{0, 1\}^p \to \{0, 1\}.$$
- Распознающие процедуры, построенные путем коррекции эл.кл., называются *логическими корректорами*
- Корректное распознавание на базе произвольных, не обязательно корректных эл.кл., основано на идее *алгебро-логического подхода*, которая предложена в (*Дюкова, Журавлёв, Рудаков, 1996*).
Зарубежный аналог — Logical Analysis of Data (LAD) (*Boros, Hammer et al., 2000*).

Алгебро-логический подход

- Рассмотрим в качестве операций предикаты $\{B_1, \ldots, B_p\}$, заданные эл.кл., $B_k : M \to \{0, 1\}$
- Корректирующие функции есть возможность выбирать среди булевых функций
$$F : \{0, 1\}^p \to \{0, 1\}.$$
- Распознающие процедуры, построенные путем коррекции эл.кл., называются *логическими корректорами*
- Корректное распознавание на базе произвольных, не обязательно корректных эл.кл., основано на идее *алгебро-логического подхода*, которая предложена в (*Дюкова, Журавлёв, Рудаков, 1996*).
Зарубежный аналог — Logical Analysis of Data (LAD) (*Boros, Hammer et al., 2000*).

Основные понятия алгебро-логического подхода

$U = \big((H_1, \sigma_1), \ldots, (H_d, \sigma_d)\big)$ — набор эл.кл.
$U(S) = \big([[H_1(S) = \sigma_1]], \ldots, [[H_d(S) = \sigma_d]]\big)$ — отклик U на объекте S

Определение

Набор эл.кл. U называется корректным для класса K, если для любой пары прецедентов $S_i \in K$ и $S_t \notin K$ выполняется $U(S_i) \neq U(S_t)$.

Определение

Булева функция $F(t_1, \ldots, t_d)$, для которой $F(U(S_i)) \neq F(U(S_t))$, $S_i \in T \cap K$, $S_t \in T \setminus K$, называется корректирующей.

Определение

Корректный набор эл.кл. U, имеющий монотонную корректирующую функцию, называется монотонным.

Основные понятия алгебро-логического подхода

$U = \big((H_1, \sigma_1), \ldots, (H_d, \sigma_d)\big)$ — набор эл.кл.
$U(S) = \big([[H_1(S) = \sigma_1]], \ldots, [[H_d(S) = \sigma_d]]\big)$ — отклик U на объекте S

Определение

Набор эл.кл. U называется корректным для класса K, если для любой пары прецедентов $S_i \in K$ и $S_t \notin K$ выполняется $U(S_i) \neq U(S_t)$.

Определение

Булева функция $F(t_1, \ldots, t_d)$, для которой $F(U(S_i)) \neq F(U(S_t))$, $S_i \in T \cap K$, $S_t \in T \setminus K$, называется корректирующей.

Определение

Корректный набор эл.кл. U, имеющий монотонную корректирующую функцию, называется монотонным.

Основные понятия алгебро-логического подхода

$U = \big((H_1, \sigma_1), \ldots, (H_d, \sigma_d)\big)$ — набор эл.кл.
$U(S) = \big([[H_1(S) = \sigma_1]], \ldots, [[H_d(S) = \sigma_d]]\big)$ — отклик U на объекте S

Определение

Набор эл.кл. U называется корректным для класса K, если для любой пары прецедентов $S_i \in K$ и $S_t \notin K$ выполняется $U(S_i) \neq U(S_t)$.

Определение

Булева функция $F(t_1, \ldots, t_d)$, для которой $F(U(S_i)) \neq F(U(S_t))$, $S_i \in T \cap K$, $S_t \in T \setminus K$, называется корректирующей.

Определение

Корректный набор эл.кл. U, имеющий монотонную корректирующую функцию, называется монотонным.

Основные понятия алгебро-логического подхода

$U = \big((H_1, \sigma_1), \ldots, (H_d, \sigma_d)\big)$ — набор эл.кл.
$U(S) = \big([[H_1(S) = \sigma_1]], \ldots, [[H_d(S) = \sigma_d]]\big)$ — отклик U на объекте S

Определение

Набор эл.кл. U называется корректным для класса K, если для любой пары прецедентов $S_i \in K$ и $S_t \notin K$ выполняется $U(S_i) \neq U(S_t)$.

Определение

Булева функция $F(t_1, \ldots, t_d)$, для которой $F(U(S_i)) \neq F(U(S_t))$, $S_i \in T \cap K$, $S_t \in T \setminus K$, называется корректирующей.

Определение

Корректный набор эл.кл. U, имеющий монотонную корректирующую функцию, называется монотонным.

Алгоритмы голосования по корректным наборам эл.кл.

Первое экспериментальное исследование логических корректоров проведено в (*Дюкова, Журавлёв, Сотнезов, 2010*).

Простейший логический корректор

- На этапе обучения для каждого класса $K \in \{K_1, \ldots, K_l\}$ строится семейство W_K корректных для K наборов эл.кл.
- При распознавании объекта S для каждого класса $K \in \{K_1, \ldots, K_l\}$ вычисляется оценка принадлежности S к K

$$\Gamma(S, K) = \frac{1}{|W_K|} \sum_{U \in W_K} \frac{1}{|T \cap K|} \sum_{S_i \in T \cap K} [[U(S_i) = U(S)]] \qquad (1)$$

Объект S относится к классу с номером $y^* = \arg\max_y \Gamma(S, K_y)$

В случае голосования по монотонным корректным наборам эл.кл. в формуле (1) вместо $[[U(S_i) = U(S)]]$ берётся $[[U(S_i) \preccurlyeq U(S)]]$, где $(a_1, \ldots, a_n) \preccurlyeq (b_1, \ldots, b_n) \Leftrightarrow a_j \leqslant b_j, \forall j \in \{1, \ldots, n\}$

Алгоритмы голосования по корректным наборам эл.кл.

Первое экспериментальное исследование логических корректоров проведено в (*Дюкова, Журавлёв, Сотнезов, 2010*).

Простейший логический корректор

- На этапе обучения для каждого класса $K \in \{K_1, \ldots, K_l\}$ строится семейство W_K корректных для K наборов эл.кл.
- При распознавании объекта S для каждого класса $K \in \{K_1, \ldots, K_l\}$ вычисляется оценка принадлежности S к K

$$\Gamma(S, K) = \frac{1}{|W_K|} \sum_{U \in W_K} \frac{1}{|T \cap K|} \sum_{S_i \in T \cap K} [[U(S_i) = U(S)]] \qquad (1)$$

Объект S относится к классу с номером $y^* = \arg\max_y \Gamma(S, K_y)$

В случае голосования по монотонным корректным наборам эл.кл. в формуле (1) вместо $[[U(S_i) = U(S)]]$ берётся $[[U(S_i) \preccurlyeq U(S)]]$, где $(a_1, \ldots, a_n) \preccurlyeq (b_1, \ldots, b_n) \Leftrightarrow a_j \leqslant b_j, \forall j \in \{1, \ldots, n\}$

Алгоритмы голосования по корректным наборам эл.кл.

Первое экспериментальное исследование логических корректоров проведено в (*Дюкова, Журавлёв, Сотнезов, 2010*).

Простейший логический корректор

- На этапе обучения для каждого класса $K \in \{K_1, \ldots, K_l\}$ строится семейство W_K корректных для K наборов эл.кл.
- При распознавании объекта S для каждого класса $K \in \{K_1, \ldots, K_l\}$ вычисляется оценка принадлежности S к K

$$\Gamma(S, K) = \frac{1}{|W_K|} \sum_{U \in W_K} \frac{1}{|T \cap K|} \sum_{S_i \in T \cap K} [[U(S_i) = U(S)]] \qquad (1)$$

Объект S относится к классу с номером $y^* = \arg\max_y \Gamma(S, K_y)$

В случае голосования по монотонным корректным наборам эл.кл. в формуле (1) вместо $[[U(S_i) = U(S)]]$ берётся $[[U(S_i) \preccurlyeq U(S)]]$, где $(a_1, \ldots, a_n) \preccurlyeq (b_1, \ldots, b_n) \Leftrightarrow a_j \leqslant b_j, \forall j \in \{1, \ldots, n\}$.

Компоненты алгоритма голосования по корректным наборам эл.кл.

- Каждый набор $U \in W_K$ длины d имеет корректирующую булеву функцию

$$F(t_1, \ldots, t_d) = \bigvee_{S_i \in T \cap K} [[U(S_i) = (t_1, \ldots, t_d)]]$$

- Для семейства W_K корректных наборов эл.кл. функция вычисления оценки $\Gamma(S, K)$ обладает свойством

$$\Gamma(S_i, K) > 0, \qquad S_i \in K,$$
$$\Gamma(S_i, K) = 0, \qquad S_i \notin K,$$

поэтому может выступать в роли корректирующей в терминах алгебраического подхода

- Стандартное решающее правило для задачи с l классами над вычисленными оценками:

$$C(\Gamma_1, \ldots, \Gamma_l) = \arg\max_y \Gamma_y$$

Компоненты алгоритма голосования по корректным наборам эл.кл.

- Каждый набор $U \in W_K$ длины d имеет корректирующую булеву функцию

$$F(t_1, \ldots, t_d) = \bigvee_{S_i \in T \cap K} [[U(S_i) = (t_1, \ldots, t_d)]]$$

- Для семейства W_K корректных наборов эл.кл. функция вычисления оценки $\Gamma(S, K)$ обладает свойством

$$\Gamma(S_i, K) > 0, \qquad S_i \in K,$$
$$\Gamma(S_i, K) = 0, \qquad S_i \notin K,$$

поэтому может выступать в роли корректирующей в терминах алгебраического подхода

- Стандартное решающее правило для задачи с l классами над вычисленными оценками:

$$C(\Gamma_1, \ldots, \Gamma_l) = \arg\max_y \Gamma_y$$

Компоненты алгоритма голосования по корректным наборам эл.кл.

- Каждый набор $U \in W_K$ длины d имеет корректирующую булеву функцию

$$F(t_1, \ldots, t_d) = \bigvee_{S_i \in T \cap K} [[U(S_i) = (t_1, \ldots, t_d)]]$$

- Для семейства W_K корректных наборов эл.кл. функция вычисления оценки $\Gamma(S, K)$ обладает свойством

$$\Gamma(S_i, K) > 0, \qquad S_i \in K,$$
$$\Gamma(S_i, K) = 0, \qquad S_i \notin K,$$

поэтому может выступать в роли корректирующей в терминах алгебраического подхода

- Стандартное решающее правило для задачи с l классами над вычисленными оценками:

$$C(\Gamma_1, \ldots, \Gamma_l) = \arg\max_y \Gamma_y$$

Решение основной задачи этапа обучения

Поиск корректных для K наборов эл.кл. сводится к поиску покрытий булевой матрицы (матрицы «сравнения»)

$$L_K = \begin{array}{c|c|c} & (H,\sigma) & \\ & \vdots & \\ \cdots & [[H(S_i)=\sigma]] \neq [[H(S_t)=\sigma]] & \cdots \\ & \vdots & \end{array} \begin{array}{l} \\ \\ (S_i, S_t), \\ S_i \in T \cap K, \\ S_t \in T \setminus K \end{array}$$

Утверждение

Набор эл.кл. $U = ((H_1,\sigma_1),\ldots,(H_d,\sigma_d))$ является (тупиковым) корректным для K тогда и только тогда, когда набор столбцов матрицы L_K, соответствующий эл.кл. $(H_1,\sigma_1),\ldots,(H_d,\sigma_d)$, является (неприводимым) покрытием матрицу L_K.

Решение основной задачи этапа обучения

Поиск корректных для K наборов эл.кл. сводится к поиску покрытий булевой матрицы (матрицы «сравнения»)

$$L_K = \left\langle \begin{array}{c} (H,\sigma) \\ \vdots \\ \cdots \quad [[H(S_i)=\sigma]] \neq [[H(S_t)=\sigma]] \quad \cdots \\ \vdots \end{array} \right\rangle \begin{array}{l} (S_i, S_t), \\ S_i \in T \cap K, \\ S_t \in T \setminus K \end{array}$$

Утверждение

Набор эл.кл. $U = ((H_1,\sigma_1),\ldots,(H_d,\sigma_d))$ является (тупиковым) корректным для K тогда и только тогда, когда набор столбцов матрицы L_K, соответствующий эл.кл. $(H_1,\sigma_1),\ldots,(H_d,\sigma_d)$, является (неприводимым) покрытием матрицу L_K.

Проблема вычислительного характера

Матрица L_K, как правило, имеет большой размер и много покрытий. Для снижения вычислительной сложности применяются следующие подходы:

- Рассматривались только эл.кл. ранга не больше r_{max} и/или информативностью не меньше I_{min} (уменьшение числа столбцов). Чаще всего $r_{\max} = 1$
- Матрица L_K строится по случайной подвыборке и случайному набору признаков (*стохастический алгоритм, bagging*)
- Предварительно отбирается *локальный базис* наиболее информативных эл.кл. Применяется *boosting* для итеративного формирования локальных базисов с хорошей *диверсификацией*
- Отбирается не более N_{max} различных неприводимых покрытий матрицы L_K с *распознающей способностью* близкой к максимальной. Поиск покрытий осуществляется *генетическим* алгоритмом.

Проблема вычислительного характера

Матрица L_K, как правило, имеет большой размер и много покрытий. Для снижения вычислительной сложности применяются следующие подходы:

- Рассматривались только эл.кл. ранга не больше r_{max} и/или информативностью не меньше I_{min} (уменьшение числа столбцов). Чаще всего $r_{\max} = 1$
- Матрица L_K строится по случайной подвыборке и случайному набору признаков (*стохастический алгоритм, bagging*)
- Предварительно отбирается *локальный базис* наиболее информативных эл.кл. Применяется *boosting* для итеративного формирования локальных базисов с хорошей *диверсификацией*
- Отбирается не более N_{max} различных неприводимых покрытий матрицы L_K с *распознающей способностью* близкой к максимальной. Поиск покрытий осуществляется *генетическим* алгоритмом.

Проблема вычислительного характера

Матрица L_K, как правило, имеет большой размер и много покрытий. Для снижения вычислительной сложности применяются следующие подходы:

- Рассматривались только эл.кл. ранга не больше r_{max} и/или информативностью не меньше I_{min} (уменьшение числа столбцов). Чаще всего $r_{\max} = 1$
- Матрица L_K строится по случайной подвыборке и случайному набору признаков (*стохастический алгоритм, bagging*)
- Предварительно отбирается *локальный базис* наиболее информативных эл.кл. Применяется *boosting* для итеративного формирования локальных базисов с хорошей *диверсификацией*
- Отбирается не более N_{max} различных неприводимых покрытий матрицы L_K с *распознающей способностью* близкой к максимальной. Поиск покрытий осуществляется *генетическим* алгоритмом.

Проблема вычислительного характера

Матрица L_K, как правило, имеет большой размер и много покрытий. Для снижения вычислительной сложности применяются следующие подходы:

- Рассматривались только эл.кл. ранга не больше r_{max} и/или информативностью не меньше I_{min} (уменьшение числа столбцов). Чаще всего $r_{\max} = 1$
- Матрица L_K строится по случайной подвыборке и случайному набору признаков (*стохастический алгоритм, bagging*)
- Предварительно отбирается *локальный базис* наиболее информативных эл.кл. Применяется *boosting* для итеративного формирования локальных базисов с хорошей *диверсификацией*
- Отбирается не более N_{max} различных неприводимых покрытий матрицы L_K с *распознающей способностью* близкой к максимальной. Поиск покрытий осуществляется *генетическим* алгоритмом.

Проблема вычислительного характера

Матрица L_K, как правило, имеет большой размер и много покрытий. Для снижения вычислительной сложности применяются следующие подходы:

- Рассматривались только эл.кл. ранга не больше r_{max} и/или информативностью не меньше I_{min} (уменьшение числа столбцов). Чаще всего $r_{\max} = 1$
- Матрица L_K строится по случайной подвыборке и случайному набору признаков (*стохастический алгоритм, bagging*)
- Предварительно отбирается *локальный базис* наиболее информативных эл.кл. Применяется *boosting* для итеративного формирования локальных базисов с хорошей *диверсификацией*
- Отбирается не более N_{max} различных неприводимых покрытий матрицы L_K с *распознающей способностью* близкой к максимальной. Поиск покрытий осуществляется *генетическим* алгоритмом.

Матрица сравнения для поиска монотонных корректных наборов эл.кл.

Поиск монотонных корректных для K наборов эл.кл. сводится к поиску покрытий булевой матрицы (матрицы «сравнения»)

$$\tilde{L}_K = \left\langle \begin{array}{c} (H,\sigma) \\ \vdots \\ \cdots \quad [[H(S_i)=\sigma]] > [[H(S_t)=\sigma]] \quad \cdots \\ \vdots \end{array} \right\rangle \begin{array}{l} (S_i, S_t), \\ S_i \in T \cap K, \\ S_t \in T \setminus K \end{array}$$

Утверждение

Набор эл.кл. $U = ((H_1,\sigma_1),\ldots,(H_d,\sigma_d))$ является (тупиковым) монотонным корректным для K тогда и только тогда, когда набор столбцов матрицы $\tilde{L}_K$, соответствующий эл.кл. $(H_1,\sigma_1),\ldots,(H_d,\sigma_d)$, является (неприводимым) покрытием матрицу $\tilde{L}_K$.

Матрица сравнения для поиска монотонных корректных наборов эл.кл.

Поиск монотонных корректных для K наборов эл.кл. сводится к поиску покрытий булевой матрицы (матрицы «сравнения»)

$$\tilde{L}_K = \begin{array}{c} (H,\sigma) \\ \vdots \\ \cdots \quad [[H(S_i)=\sigma]] > [[H(S_t)=\sigma]] \quad \cdots \\ \vdots \end{array} \begin{array}{l} (S_i, S_t), \\ S_i \in T \cap K, \\ S_t \in T \setminus K \end{array}$$

Утверждение

Набор эл.кл. $U = ((H_1,\sigma_1),\ldots,(H_d,\sigma_d))$ является (тупиковым) монотонным корректным для K тогда и только тогда, когда набор столбцов матрицы $\tilde{L}_K$, соответствующий эл.кл. $(H_1,\sigma_1),\ldots,(H_d,\sigma_d)$, является (неприводимым) покрытием матрицу $\tilde{L}_K$.

Тестовые задачи

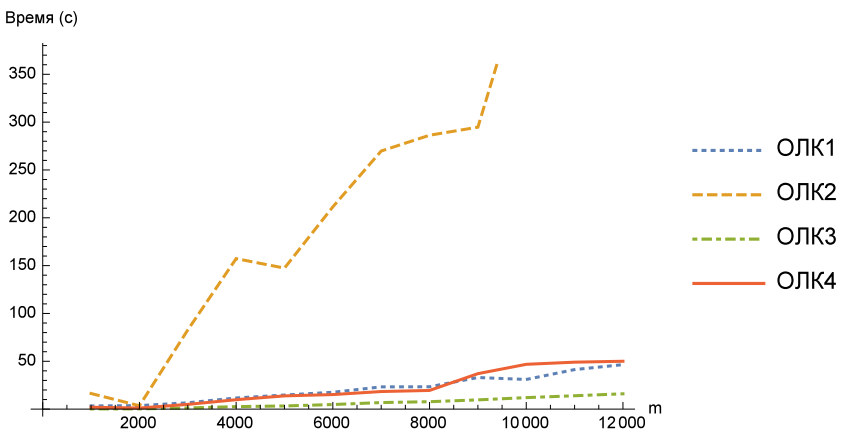
	Название	l	m	n	z
1.	audiology	24	226	69	161
2.	balance scale	3	625	4	20
3.	breast cancer	2	699	9	90
4.	car	4	1728	6	21
5.	dermatology	4	366	34	192
6.	house votes	2	435	16	48
7.	kr vs kp	2	3196	36	73
8.	monks-2	2	601	6	17
9.	nursery	5	12960	8	27
10.	soybean large	19	307	35	132
11.	tic tac toe	2	958	9	27
12.	optdigits	10	5620	64	914
13.	letter recognition	26	20000	16	256
14.	lenses	3	24	4	9
15.	soybean small	4	47	35	72

Результаты экспериментов

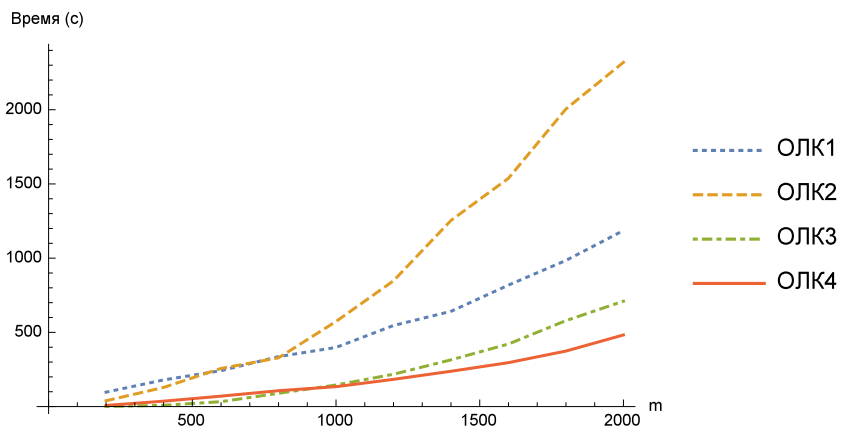
	Задача	T*	ПН*	МОН*	ОЛК1	ОЛК2	ОЛК3	ОЛК4
1.	audiology	0.03	0.03	0.03	0.03	0.03	**0.02**	0.03
2.	b. scale	0.25	0.2	0.19	**0.18**	0.23	0.23	0.25
3.	b. cancer	0.046	**0.044**	0.057	0.061	0.059	0.065	0.059
4.	car	0.061	0.032	0.033	0.013	0.027	0.022	**0.011**
5.	dermat.	0.41	0.4	0.4	**0.39**	0.42	0.44	0.43
6.	h. votes	0.07	**0.05**	0.07	**0.05**	0.06	0.07	0.08
7.	kr-vs-kp	0.008	0.004	**0.003**	0.008	0.007	0.004	**0.003**
8.	monks-2	0.37	0.55	0.42	**0.04**	0.44	0.56	0.36
9.	nursery	0.027	0.003	0.005	0.002	—	**0.0019**	0.004
10.	soybean l.	0.075	**0.064**	0.072	0.078	0.106	0.083	0.075
11.	tic-tac-toe	0.011	0.002	0.005	0.028	0.002	**0.001**	0.007
12.	letter r.	0.21	**0.16**	0.25	—	—	0.23	0.25
13.	optdigits	0.25	0.23	0.17	0.15	—	0.27	**0.14**
14.	lenses	0.42	**0.25**	0.29	0.33	0.29	0.38	**0.25**
15.	soybean s.	**0**	**0**	**0**	**0**	0.02	0.04	**0**

Время обучения логических корректоров с различными стратегиями построения локальных базисов

- ОЛК1 — один локальный базис из одноранговых эл.кл.
- ОЛК2 — один локальный базис строится бустингом над представительными наборами.
- ОЛК3 — локальный базис модифицируется на каждой итерации классическим голосованием по представительным наборам.
- ОЛК4 — локальный базис модифицируется на каждой итерации бустингом над представительными наборами.



nursery

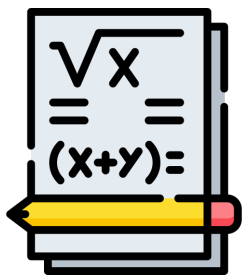


optdigits

Задачи на самостоятельную работу

1. Построить матрицу
сравнения для поиска корректных эл.кл. выборки

$$K_1 = \{(0,1,0,0),(0,0,1,0),(1,0,0,0)\}$$
$$K_2 = \{(1,1,0,0),(0,1,1,0),(0,1,1,1)\}.$$



Рассмотреть варианты с различными ограничениями
на ранг эл.кл. и информативность (IGain)
2. Выписать формулу для корректирующей булево функции
монотонного корректного набора эл.кл.

Выводы

- Корректность эл.кл. является существенным ограничением в прикладных задачах: мешает искать информативные логические закономерности, плохо сказывается на обобщающей способности и интерпретируемости обученных алгоритмов
- Идеи алгебраического подхода позволили развить новое направление — алгебро-логический подход. Логические корректоры на практике имеют распознающую способность лучше, чем процедуры голосования по корректным эл.кл.
- Обучение логических корректоров сводится к задаче монотонной дуализации. Однако из-за больших входных матриц решать эту задачу «в лоб» не эффективно, поэтому применяются разнообразные эвристики для уменьшения времени счета.

Выводы

- Корректность эл.кл. является существенным ограничением в прикладных задачах: мешает искать информативные логические закономерности, плохо сказывается на обобщающей способности и интерпретируемости обученных алгоритмов
- Идеи алгебраического подхода позволили развить новое направление — алгебро-логический подход. Логические корректоры на практике имеют распознающую способность лучше, чем процедуры голосования по корректным эл.кл.
- Обучение логических корректоров сводится к задаче монотонной дуализации. Однако из-за больших входных матриц решать эту задачу «в лоб» не эффективно, поэтому применяются разнообразные эвристики для уменьшения времени счета.

Выводы

- Корректность эл.кл. является существенным ограничением в прикладных задачах: мешает искать информативные логические закономерности, плохо сказывается на обобщающей способности и интерпретируемости обученных алгоритмов
- Идеи алгебраического подхода позволили развить новое направление — алгебро-логический подход. Логические корректоры на практике имеют распознающую способность лучше, чем процедуры голосования по корректным эл.кл.
- Обучение логических корректоров сводится к задаче монотонной дуализации. Однако из-за больших входных матриц решать эту задачу «в лоб» не эффективно, поэтому применяются разнообразные эвристики для уменьшения времени счета.