Теория кодирования на кафедре теории вероятностей

Полянский Никита Андреевич

МГУ им. М.В. Ломоносова, Механико-математический факультет, Кафедра теории вероятностей

22.03.2017

Содержание доклада

1. Групповые проверки, задачи поиска и двоичные свободные от перекрытий (СП) коды
2. Недвоичные сигнатурные коды
3. Процедуры многошагового поиска. Адаптивный поиск скрытого гиперграфа.

Раздел 1

Групповые проверки, задачи поиска и двоичные свободные от перекрытий (СП) коды

- Комбинаторная модель неадаптивного поиска и границы скорости для СП-кодов. Границы скорости при списочном декодировании.
- Вероятностные модели неадаптивного поиска. Границы пропускной способности при декодировании перебором и при дизъюнктивном декодировании. Проверка гипотезы об объеме дефектного множества.

Групповое тестирование

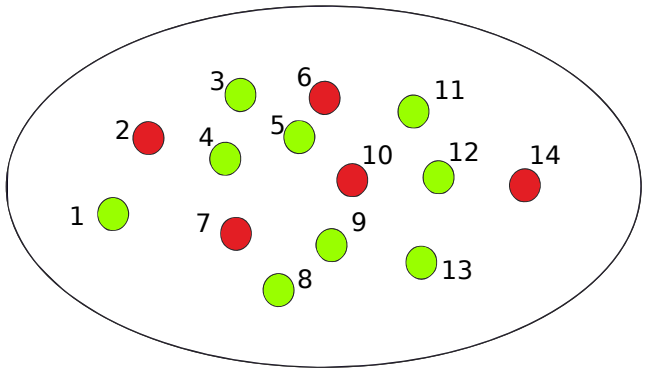
Постановка задачи

Дано множество $T = \{1, 2, \ldots, t\}$ из t элементов.
$S_{un} \subset T$ – подмножество дефектных элементов, $|S_{un}| \leqslant s$.
Тест – подмножество $S \subset T$.
Результат теста положителен, если $S \cap S_{un} \neq \varnothing$.
Задача: найти дефекты, используя минимальное число тестов.

Групповое тестирование

Постановка задачи

Дано множество $T = \{1, 2, \ldots, t\}$ из t элементов.
$S_{un} \subset T$ – подмножество дефектных элементов, $|S_{un}| \leqslant s$.
Тест – подмножество $S \subset T$.
Результат теста положителен, если $S \cap S_{un} \neq \varnothing$.
Задача: найти дефекты, используя минимальное число тестов.

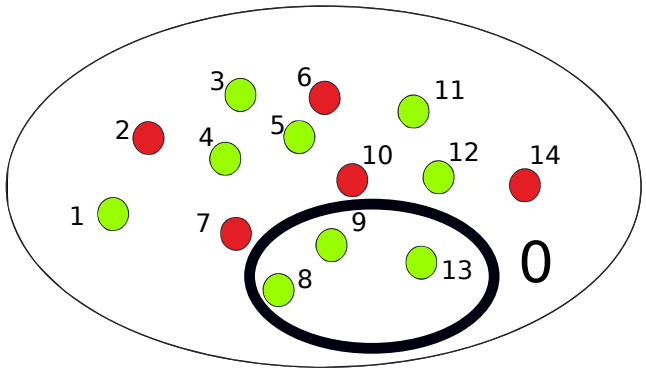


Пример:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.

Групповое тестирование

Постановка задачи

Дано множество $T = \{1, 2, \ldots, t\}$ из t элементов.
$S_{un} \subset T$ – подмножество дефектных элементов, $|S_{un}| \leqslant s$.
Тест – подмножество $S \subset T$.
Результат теста положителен, если $S \cap S_{un} \neq \varnothing$.
Задача: найти дефекты, используя минимальное число тестов.

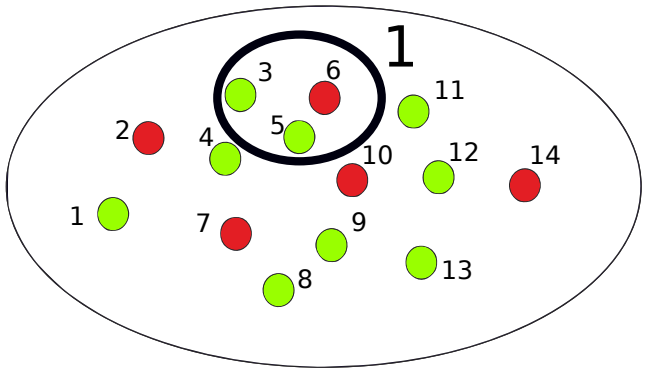


Пример:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.
$S = \{8, 9, 13\}$.
$S \cap S_{un} = \varnothing$, результат теста отрицательный (0).

Групповое тестирование

Постановка задачи

Дано множество $T = \{1, 2, \ldots, t\}$ из t элементов.
$S_{un} \subset T$ – подмножество дефектных элементов, $|S_{un}| \leqslant s$.
Тест – подмножество $S \subset T$.
Результат теста положителен, если $S \cap S_{un} \neq \varnothing$.
Задача: найти дефекты, используя минимальное число тестов.



Пример:
$t = 14$,
$S_{un} = \{2, 6, 7, 10, 14\}$.
$S = \{3, 5, 6\}$.
$S \cap S_{un} = \{6\}$, результат теста положительный (1).

Групповое тестирование

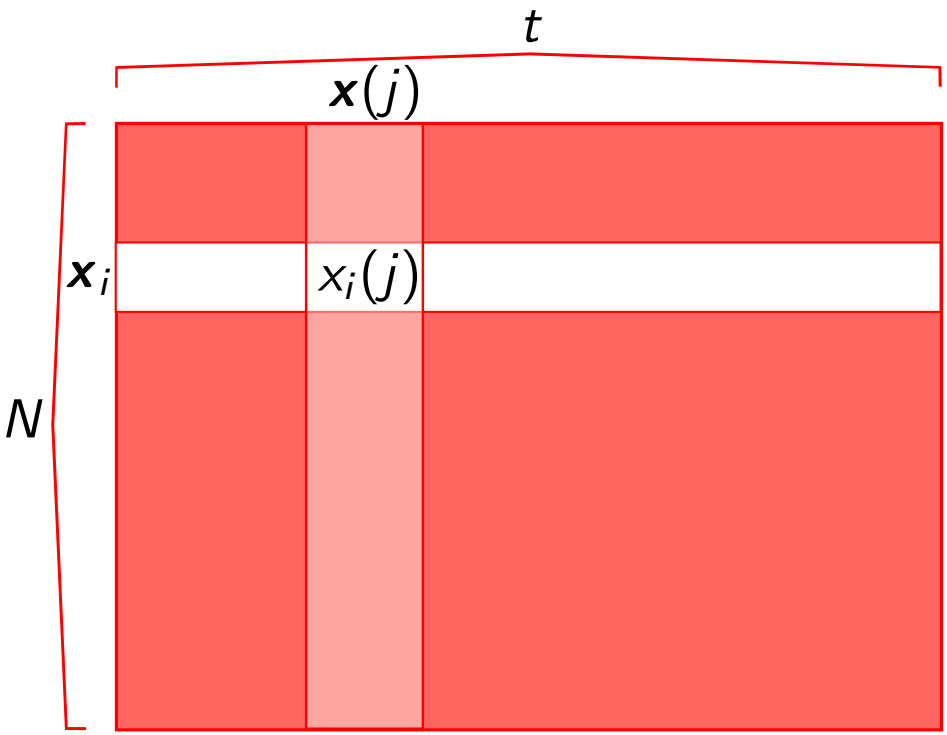
Существуют две классические модели тестирования — адаптивная и неадаптивная:

1. Если все тесты проводятся одновременно, то такое тестирование называется неадаптивным.
2. Если при выборе следующего тестируемого множества можно использовать результаты предыдущих тестов, то тестирование называется адаптивным или последовательным.

Основные обозначения

$[n] = \{1, 2, \ldots, n\}$

Семейство из t двоичных векторов $\{\boldsymbol{x}(j) \in \{0,1\}^N, j \in [t]\}$ будем называть двоичным кодом длины N и объема (мощности) t. Двоичный код можно представить как двоичную $(N \times t)$-таблицу $X \triangleq \|x_i(j)\|$, $x_i(j) \in \{0,1\}$, $i \in [N]$, $j \in [t]$.

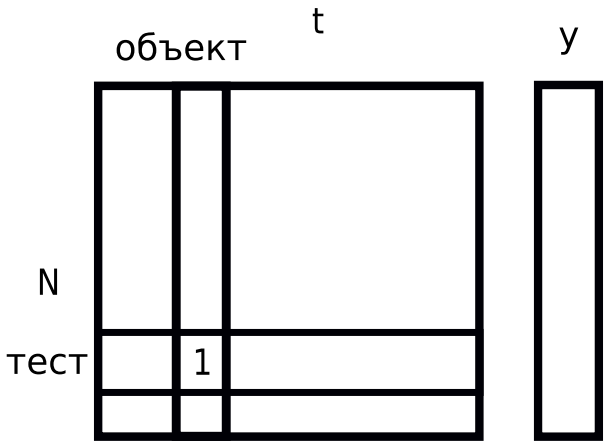


$\bigvee$ – (покомпонентная) дизъюнктивная сумма.
$\bigwedge$ – (покомпонентная) конъюнкция.

Будем говорить, что двоичный столбец $\mathbf{u}$ покрывает двоичный столбец $\mathbf{v}$, если $u_i \geqslant v_i, \forall\, i \in [N]$.

Групповое тестирование

Неадаптивную модель тестирования можно задать с помощью двоичной таблицы.



Строчкам соответствуют тесты, столбцам – объекты. На пересечении стоит единица, если объект включен в тест.

Вектор $\boldsymbol{y}$, содержащий результаты тестов, называется откликом:

$$\boldsymbol{y} = \boldsymbol{y}\left(X, \mathcal{S}_{un}\right) = \begin{cases} \bigvee\limits_{i \in \mathcal{S}_{un}} \boldsymbol{x}(i), & \mathcal{S}_{un} \neq \varnothing, \\ \boldsymbol{0}, & \mathcal{S}_{un} = \varnothing \end{cases}$$

План поиска и дизъюнктивный код

Пусть известно, что $|\mathcal{S}_{un}| \leqslant s$.

Определение 1

X — план s-поиска, если для любых $\mathcal{S}_1, \mathcal{S}_2 \subset [t]$ мощности не более s выполнено условие

$$\boldsymbol{y}(X, \mathcal{S}_1) \neq \boldsymbol{y}(X, \mathcal{S}_2).$$

Определение 2

X —дизъюнктивный[a] s-код, если для любых $\mathcal{S} \subset [t]$ мощности не более s и $j \in [t] \setminus \mathcal{S}$ отклик $\boldsymbol{y}(X, \mathcal{S})$ не покрывает $\boldsymbol{x}(j)$, т.е.

$$\boldsymbol{x}(j) \bigvee \boldsymbol{y}(X, \mathcal{S}) \neq \boldsymbol{y}(X, \mathcal{S})$$

[a] Kautz W.H., Singleton R.C., 1964.

План поиска и дизъюнктивный код

План (код) поиска для переборного декодирования

X — план s-поиска, если для любых $\mathcal{S}_1, \mathcal{S}_2 \subset [t]$ мощности не более s выполнено условие

$$\boldsymbol{y}(X, \mathcal{S}_1) \neq \boldsymbol{y}(X, \mathcal{S}_2).$$

Дизъюнктивный код

X —дизъюнктивный s-код, если для любых $\mathcal{S} \subset [t]$ мощности не более s и $j \in [t] \setminus \mathcal{S}$ результат $\boldsymbol{y}(X, \mathcal{S})$ не покрывает $\boldsymbol{x}(j)$, т.е.

$$\boldsymbol{x}(j) \bigvee \boldsymbol{y}(X, \mathcal{S}) \neq \boldsymbol{y}(X, \mathcal{S})$$

Связь плана и кода

Если X — дизъюнктивный s-код, то X — план s-поиска.

Если X — план s-поиска, то X — дизъюнктивный $(s-1)$-код.

Поиск по признаку в библиотеке ДНК

Пусть имеются 4 копии ДНК последовательности. Последовательности $\{C_1, C_2, C_3, C_4, C_5\}$ называются клонами и образуют библиотеку ДНК. Необходимо найти все клоны, содержащий признак, например AAA или TAA.

$$\overbrace{\boxed{AAA}GCGTCT\boxed{TAA}}^{C_1}\overbrace{CCGATAGGCAAC}^{C_3}TTG,$$

$$\boxed{AAA}GC\overbrace{GTCT\boxed{TAA}CCGA}^{C_2}TAGGCAACTTG,$$

$$\boxed{AAA}GCGT\overbrace{CT\boxed{TAA}CCGATAGGC}^{C_4}AACTTG,$$

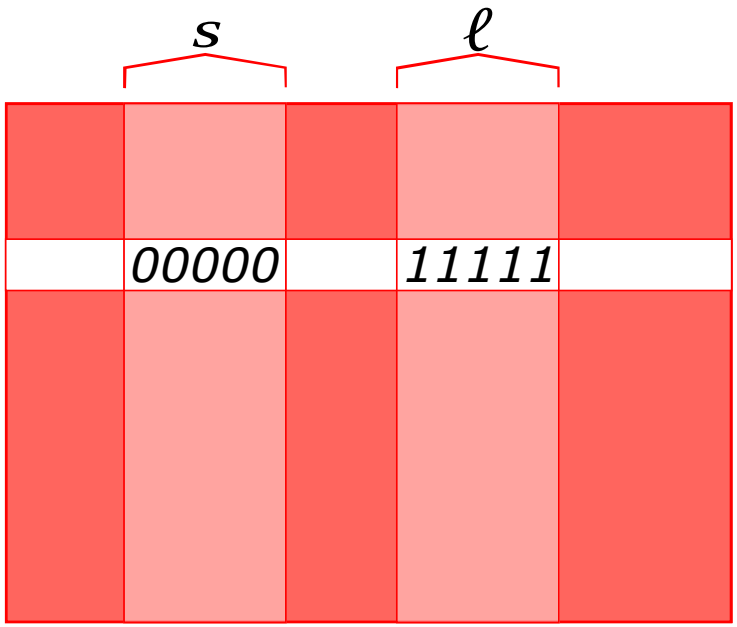
$$\boxed{AAA}GCGTCT\boxed{TAA}CCGAT\overbrace{AGGCAACTTG}^{C_5}.$$

СП (s, ℓ)-коды

Определение 3

Двоичный код X называется свободным от перекрытий (s, ℓ)-кодом (СП (s, ℓ)-кодом[a]), если дизъюнктивная сумма любых s столбцов кода X не покрывает конъюнкцию ℓ любых других столбцов кода X.

[a]1988, Mitchell C.J, Piper F.C.



Обозначим через $t(N, s, \ell)$ максимальный объем СП (s, ℓ)-кодов длины N, и определим скорость СП (s, ℓ)-кодов:

$$R(s, \ell) \triangleq \varlimsup_{N \to \infty} \frac{\log_2 t(N, s, \ell)}{N}.$$

Границы СП (s, ℓ)-кодов

Верхние границы скорости СП кодов

Для скорости дизъюнктивных кодов верна следующая граница[a]

$$R(s,1) \leqslant \frac{2\log_2 s}{s^2}(1+o(1)), \quad s \to \infty.$$

Для скорости СП (s, ℓ)-кодов верна следующая граница[b]

$$R(s,\ell) \leqslant \frac{(\ell+1)^{\ell+1}}{2\,e^{\ell-1}}\frac{\log_2 s}{s^{\ell+1}}(1+o(1)), \quad \ell \text{ фикс.}, \ s \to \infty$$

[a] 1982, Дьячков А.Г., Рыков В.В.
[b] 2002, Виленкин П.А., Еханин С.М., Дьячков А.Г.

Границы СП (s, ℓ)-кодов

Нижние границы скорости СП кодов

Для скорости дизъюнктивных кодов верна следующая граница[a]

$$R(s,1) \geqslant \frac{1}{s^2 \log_2 e}(1+o(1)), \quad s \to \infty.$$

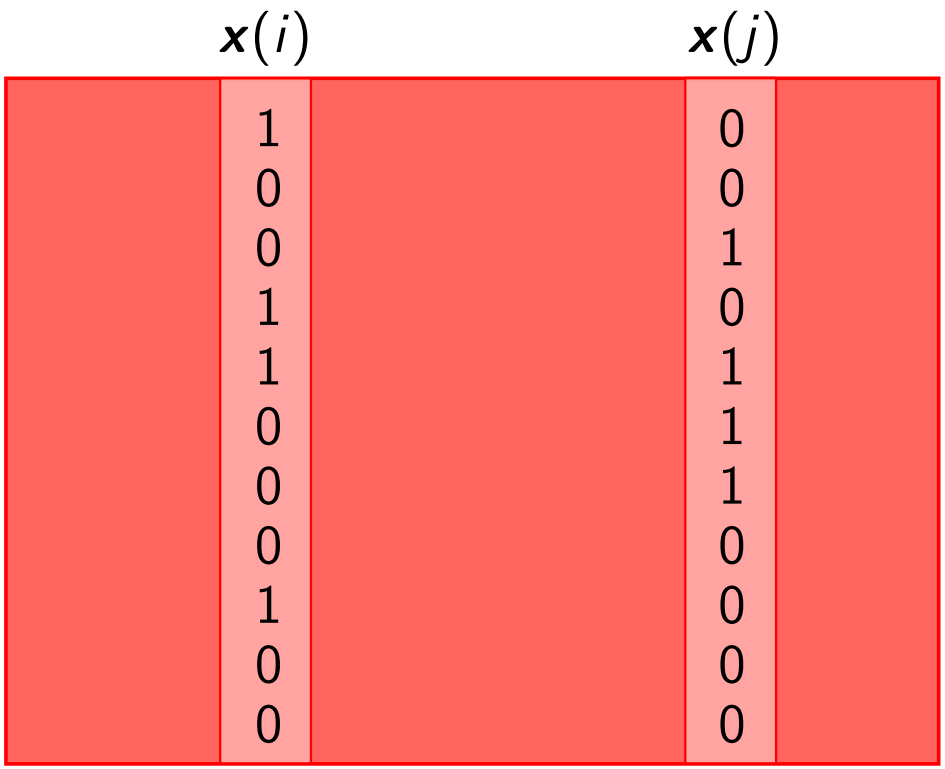
Для скорости СП (s, ℓ)-кодов верна следующая граница[b]

$$R(s,\ell) \geqslant \frac{\ell^\ell}{e^\ell}\frac{\log_2 e}{s^{\ell+1}}(1+o(1)), \quad \ell \text{ фикс.}, s \to \infty$$

[a] 1989, Дьячков А.Г., Рашад А.М., Рыков В.В.
[b] 2014, Воробьев И.В., Дьячков А.Г., Полянский Н.А., Щукин В.Ю. [ВДПЩ]

Схема доказательства нижних границ

Доказательство основано на методе случайного кодирования на ансамбле двоичных равновесных кодов.
Зафиксируем параметры N, t и Q, $0 < Q < 1$. Столбцы случайной матрицы выбираются независимо и равновероятно из множества всех $\binom{N}{\lfloor QN \rfloor}$ двоичных столбцов веса $\lfloor QN \rfloor$.



$\boldsymbol{x}(i)$ и $\boldsymbol{x}(j)$ независимы

$\#1 = \lfloor QN \rfloor$

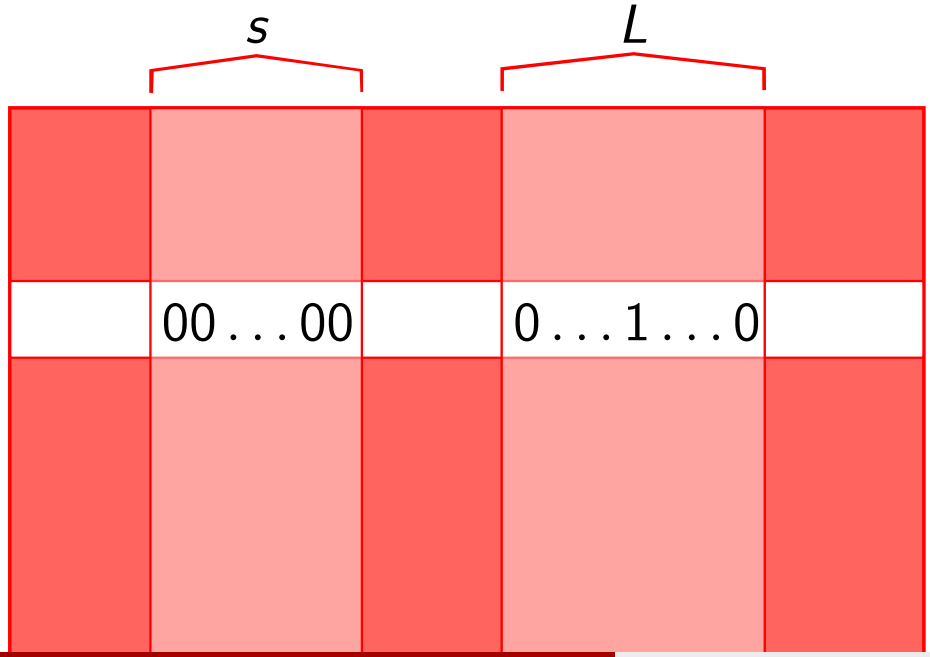
$\#0 = N - \lfloor QN \rfloor$

СД s_L-коды

Определение 4

Двоичный код X называется дизъюнктивным кодом со списочным декодированием силы s с объемом списка L (СД s_L-кодом[a]), если дизъюнктивная сумма любых s столбцов кода X покрывает не более $L-1$ других столбцов кода X, не входящих в эту сумму.

[a][1981, А.Г. Дьячков, В.В. Рыков]



Обозначим через $t(N,s,L)$ максимальный объем СД s_L-кодов длины N, и определим скорость СД s_L-кодов:

$$R_L(s) \triangleq \varlimsup_{N\to\infty} \frac{\log_2 t(N,s,L)}{N}.$$

Границы скорости СД s_L-кодов

Нижняя граница скорости СД s_L-кодов

Для скорости СД s_L-кодов верна следующая граница[a]

$$R_L(s) \geqslant \frac{L}{s^2 \log_2 e}(1 + o(1)), \quad L \text{ фикс.}, \ s \to \infty.$$

[a] 2014, [ВДПЩ]

Верхняя граница скорости СД s_L-кодов

Для скорости СД s_L-кодов верна следующая граница[a]

$$R_L(s) \leqslant \frac{2L \log_2 s}{s^2}(1 + o(1)), \quad L \text{ фикс.}, \ s \to \infty.$$

[a] 2014, [ВДПЩ]

Вероятностная модель неадаптивного поиска

Определение 5

X — план s-поиска с вероятностью ошибки $\leqslant \varepsilon$, если для не менее $(1-\varepsilon)\binom{t}{s}$ множеств $\mathcal{S} \subset [t]$ мощности s выполнено условие

$$\boldsymbol{y}(X, \mathcal{S}) \neq \boldsymbol{y}(X, \mathcal{S}'), \quad \text{для } \forall S' \subset [t], S \neq S', |S'| = s$$

Определение 6

X — почти дизъюнктивный (s, ε)-код, если для не менее $(1-\varepsilon)\binom{t}{s}$ множеств $\mathcal{S} \subset [t]$ мощности s выполнено условие

$$\boldsymbol{y}(X, \mathcal{S}) \neq \boldsymbol{y}(X, \mathcal{S}) \bigvee \boldsymbol{x}(j), \quad \text{для } \forall j \subset [t] \setminus \mathcal{S}$$

План поиска с вероятностью ошибки

Определение 7

Множество $\mathcal{S}$, $\mathcal{S} \subset [t]$, $|\mathcal{S}| = s$, назовем s-плохим для кода X, если существует хотя бы одно множество $\mathcal{S}' \subset [t]$, $S' \neq S, |\mathcal{S}'| = s$ такое, что

$$\boldsymbol{y}(X, \mathcal{S}) = \boldsymbol{y}(X, \mathcal{S}').$$

Через $B(s, X)$ обозначим количество s-плохих множеств для кода X. Зафиксируем скорость R, $R > 0$. Определим наименьшую ошибку для планов s-поиска с вероятностью ошибки:

$$\varepsilon(s, R, N) \triangleq \min_{X : t = \lfloor 2^{RN} \rfloor} \left\{ \frac{B(s, X)}{\binom{t}{s}} \right\}.$$

Экспонента ошибки:

$$\mathbf{E}(s, R) \triangleq \varlimsup_{N \to \infty} \frac{-\log_2 \varepsilon(s, R, N)}{N}.$$

План поиска с вероятностью ошибки

Пропускная способность планов поиска с вероятностью ошибки:

$$C(s) \triangleq \sup\{ R \,:\, \mathbf{E}(s,R) > 0\}.$$

Равенство для пропускной способности

Пропускная способность $C(s)$ удовлетворяет равенству[a]

$$C(s) = \frac{1}{s}.$$

[a] 1973, Малютов М.Б., Фрейдлина В.Л.

Почти дизъюнктивные s-коды

Определение 8

Множество $\mathcal{S}$, $\mathcal{S} \subset [t]$, $|\mathcal{S}| = s$, назовем $(s,1)$-плохим для кода X, если дизъюнктивная сумма соответствующих s кодовых слов покрывает постороннее слово $x(j)$, т.е. существует такое $j \in [t] \setminus \mathcal{S}$, что

$$\boldsymbol{y}(X,\mathcal{S}) = \boldsymbol{y}(X,\mathcal{S}) \bigvee \boldsymbol{x}(j).$$

Через $B(s,1,X)$ обозначим количество $(s,1)$-плохих множеств для кода X. Зафиксируем скорость R, $R > 0$. Определим наименьшую ошибку для почти дизъюнктивных (s,ε)-кодов:

$$\varepsilon(s,1,R,N) \triangleq \min_{X:t=\lfloor 2^{RN} \rfloor} \left\{ \frac{B(s,1,X)}{\binom{t}{s}} \right\}.$$

Экспонента ошибки:

$$\mathbf{E}(s,1,R) \triangleq \varlimsup_{N\to\infty} \frac{-\log_2 \varepsilon(s,1,R,N)}{N}.$$

Почти дизъюнктивные s-коды

Пропускная способность:

$$C(s,1) \triangleq \sup\{\, R \,:\, \mathbf{E}(s,1,R) > 0\}.$$

Верхняя границы для пропускной способности

Для пропускной способности почти дизъюнктивных кодов верна следующая граница

$$C(s,1) \leqslant \frac{1}{s}$$

Нижняя границы для пропускной способности

Для пропускной способности почти дизъюнктивных кодов верна следующая граница[a]

$$C(s,1) \geqslant \frac{\ln 2}{s}(1+o(1)), \quad s \to \infty.$$

[a] 2015, [ВДПЩ]

Проверка гипотез

Пусть заранее неизвестна мощность дефектного множества.

Основная гипотеза
$$H_0 \ : \ |\mathcal{S}_{un}| \leqslant s$$

Альтернатива
$$H_1 \ : \ |\mathcal{S}_{un}| \geqslant s+1$$

Безошибочная проверка гипотезы H_0 против H_1

Результаты неадаптивных групповых тестов, заданных кодом X, позволяют безошибочно проверить гипотезу H_0 против альтернативы H_1 в том и только том случае, если код X является дизъюнктивным s-кодом.

Правила принятия решения

Основная гипотеза
$H_0 : |\mathcal{S}_{un}| \leqslant s$

Альтернатива
$H_1 : |\mathcal{S}_{un}| \geqslant s+1$

Дизъюнктивный критерий

$$\begin{cases} \text{принять } H_0, & \text{если } \boldsymbol{y}(X, \mathcal{S}_{un}) \text{ покрывает } \leqslant s \text{ столбцов кода } X, \\ \text{принять } H_1, & \text{если } \boldsymbol{y}(X, \mathcal{S}_{un}) \text{ покрывает } \geqslant s+1 \text{ столбцов кода } X. \end{cases}$$

Пороговый критерий

$$\begin{cases} \text{принять } \{H_0 : |\mathcal{S}_{un}| \leqslant s\}, & \text{если } |\boldsymbol{y}(X, \mathcal{S}_{un})| \leqslant \lfloor \tau N \rfloor, \\ \text{принять } \{H_1 : |\mathcal{S}_{un}| \geqslant s+1\}, & \text{если } |\boldsymbol{y}(X, \mathcal{S}_{un})| \geqslant \lfloor \tau N \rfloor + 1, \end{cases}$$

где τ, $0 < \tau < 1$, – заданная константа.

Экспонента ошибки

Распределение вероятностей случайного множества $\mathcal{S}_{un}$, $\mathcal{S}_{un} \subset [t]$, задано вектором $\psi \triangleq (p_0, p_1, \ldots, p_t)$, $p_k \geqslant 0$, $\sum_{k=0}^{t} p_k = 1$:

$$\Pr\{\mathcal{S}_{un} = \mathcal{S}\} \triangleq \frac{p_{|\mathcal{S}|}}{\binom{t}{|\mathcal{S}|}} \quad \text{для любого подмножества} \quad \mathcal{S} \subseteq [t].$$

Максимальная вероятность ошибки:

$$\mathbf{e}_s(\tau, \psi, X) \triangleq \max\left\{\Pr\{\text{принять } H_1 \big| H_0\}, \Pr\{\text{принять } H_0 \big| H_1\}\right\}.$$

Универсальная вероятность ошибки:

$$\mathbf{e}_s^N(\tau, R) \triangleq \max_{\psi}\left\{\min_{X: t=\lfloor 2^{RN} \rfloor} \mathbf{e}_s(\tau, \psi, X)\right\}.$$

Экспонента ошибки порогового критерия:

$$\mathbf{E}_s(\tau, R) \triangleq \varlimsup_{N\to\infty} \frac{-\log_2 \mathbf{e}_s^N(\tau, R)}{N}.$$

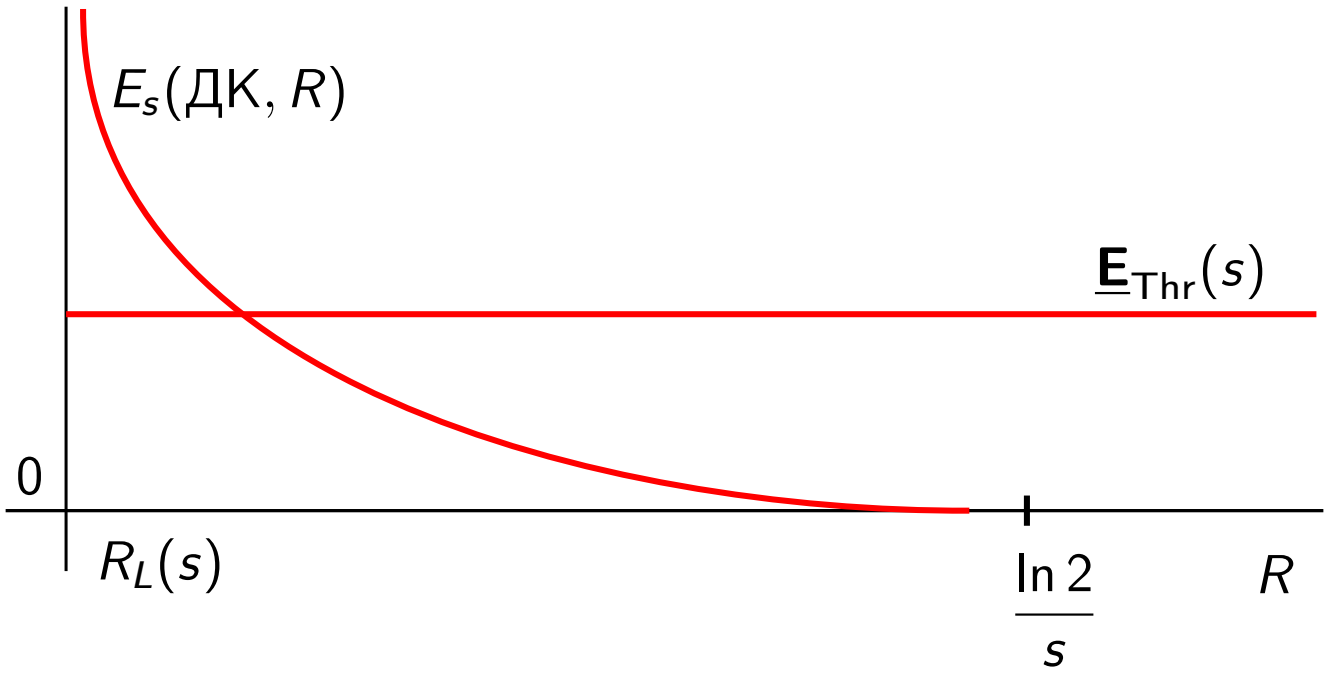
Сравнение экспонент ошибки

Дизъюнктивный критерий

$$E_s(\text{ДК}, R) = 0 \quad \text{если} \quad R \geqslant \frac{\ln 2}{s}.$$

Пороговый критерий

$$E_s(\text{ПК}, R) \geqslant \underline{\mathbf{E}}_{\text{Thr}}(s) \geqslant \frac{\log_2 e}{4s^2}.$$



Раздел 2

Недвоичные сигнатурные коды

- Границы скорости разделяющих кодов
- Границы скорости при списочном декодировании для объединяющего канала множественного доступа (КМД)
- Границы скорости для композиционного КМД

Канал множественного доступа

Канал множественного доступа (КМД) имеет s входов и один выход.
Пусть алфавиты для каждого из s входов одинаковы и совпадают с алфавитом $\mathcal{A}_q \triangleq \{0, 1, \ldots, q-1\}$ – стандартным q-ичным алфавитом.
Через Z обозначим алфавит на выходе КМД.
Тогда дискретный канал множественного доступа без шума описывается

$$z = f(a_1, \ldots, a_s) \triangleq f(a_1^s), \quad z \in Z,\ a_1^s \in \mathcal{A}_q^s.$$

Мы будем рассматривать только симметричные КМД, т.е.

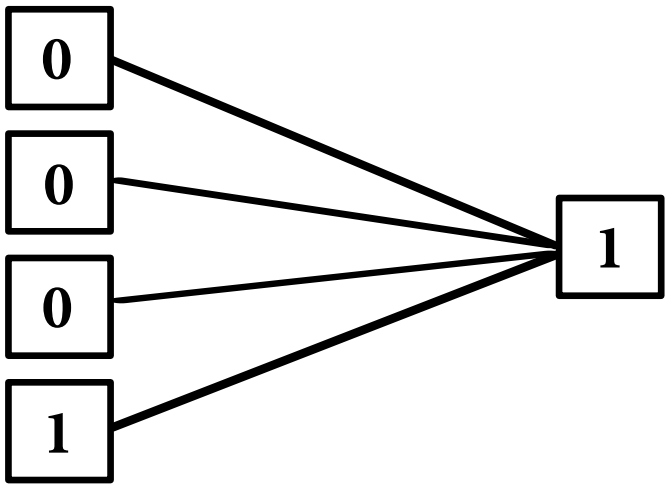
$$f(\pi(a_1, \ldots, a_s)) = f(a_1, \ldots, a_s) \quad \forall \pi \in S_s$$

Канал множественного доступа

Канал множественного доступа (КМД) имеет s входов и один выход.
Пусть алфавиты для каждого из s входов одинаковы и совпадают с алфавитом $\mathcal{A}_q \triangleq \{0, 1, \ldots, q-1\}$ – стандартным q-ичным алфавитом.
Через Z обозначим алфавит на выходе КМД.
Тогда дискретный канал множественного доступа без шума описывается

$$z = f(a_1, \ldots, a_s) \triangleq f(a_1^s), \quad z \in Z,\ a_1^s \in \mathcal{A}_q^s.$$

Пример: дизъюнктивный КМД



Алфавит на входе $\mathcal{A}_2 = \{0, 1\}$,
алфавит на выходе $Z = \{0, 1\}$

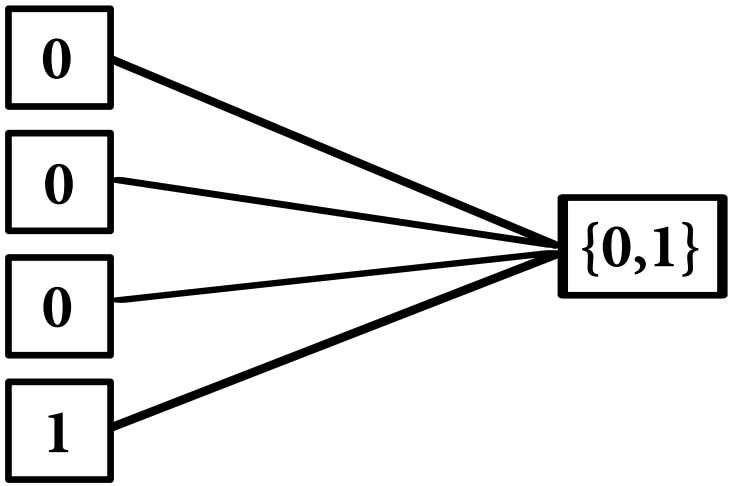
$$z = \bigvee_{i \in [s]} a_i$$

Канал множественного доступа

Канал множественного доступа (КМД) имеет s входов и один выход.
Пусть алфавиты для каждого из s входов одинаковы и совпадают с алфавитом $\mathcal{A}_q \triangleq \{0, 1, \ldots, q-1\}$ – стандартным q-ичным алфавитом.
Через Z обозначим алфавит на выходе КМД.
Тогда дискретный канал множественного доступа без шума описывается

$$z = f(a_1, \ldots, a_s) \triangleq f(a_1^s), \quad z \in Z, \ a_1^s \in \mathcal{A}_q^s.$$

Пример: объединяющий КМД



Алфавит $\mathcal{A}_q = \{0, 1, \ldots, q-1\}$,
$Z = 2^{\mathcal{A}_q} \setminus \varnothing$

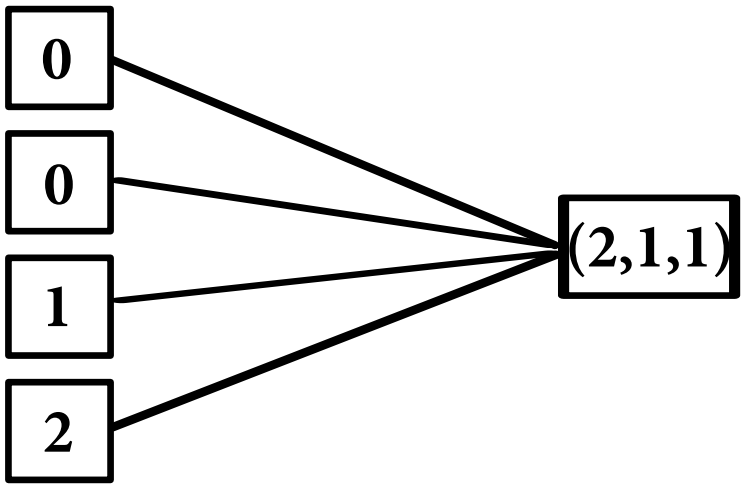
$$z = \bigcup_{i \in [s]} a_i$$

Канал множественного доступа

Канал множественного доступа (КМД) имеет s входов и один выход.
Пусть алфавиты для каждого из s входов одинаковы и совпадают с алфавитом $\mathcal{A}_q \triangleq \{0, 1, \ldots, q-1\}$ – стандартным q-ичным алфавитом.
Через Z обозначим алфавит на выходе КМД.
Тогда дискретный канал множественного доступа без шума описывается

$$z = f(a_1, \ldots, a_s) \triangleq f(a_1^s), \quad z \in Z,\ a_1^s \in \mathcal{A}_q^s.$$

Пример: композиционный КМД



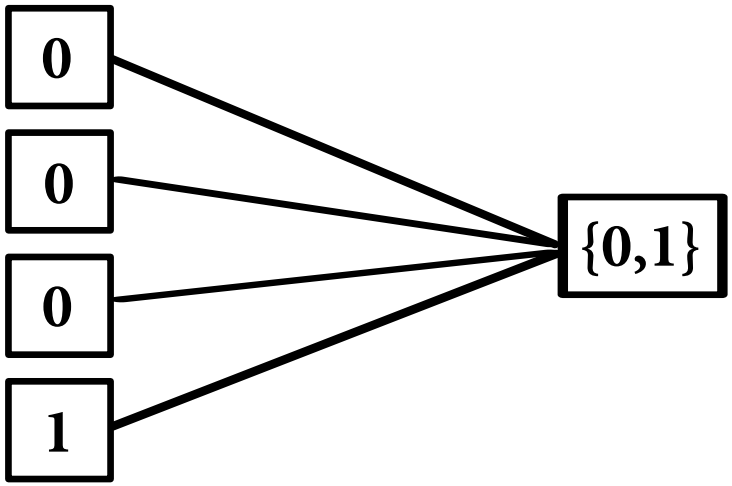
Алфавит $\mathcal{A}_q = \{0, 1, \ldots, q-1\}$,
$Z = \{(t_0, t_1, \ldots, t_{q-1}) :$
$t_i \geqslant 0, \sum_i t_i = s\}$
z — это тип (композиция),
соответствующий $(a_1, a_2, \ldots, a_s)$

q-ичный код

Столбец $\mathbb{Q} = (Q_1, Q_2, \ldots, Q_N)$, $Q_i \subset \mathcal{A}_q$, $i \in [N]$, будем называть гиперсловом.
Определим гиперсумму q-ичных символов $a_1, a_2, \ldots, a_s \in \mathcal{A}_q$:

$$\langle a_1, \ldots, a_s \rangle \triangleq \bigcup_{i=1}^{s} \{a_i\}.$$

Объединяющий КМД



Алфавит $\mathcal{A}_q = \{0, 1, \ldots, q-1\}$,
$Z = 2^{\mathcal{A}_q} \setminus \varnothing$

$$z = \bigcup_{i \in [s]} a_i$$

q-ичный код

Столбец $\mathbb{Q} = (Q_1, Q_2, \ldots, Q_N)$, $Q_i \subset \mathcal{A}_q$, $i \in [N]$, будем называть гиперсловом. Определим гиперсумму q-ичных символов $a_1, a_2, \ldots, a_s \in \mathcal{A}_q$:

$$\langle a_1, \ldots, a_s \rangle \triangleq \bigcup_{i=1}^{s} \{a_i\}.$$

Тогда определим гиперсумму q-ичных векторов $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(s)$

$$\langle \boldsymbol{x}(1), \ldots, \boldsymbol{x}(s) \rangle \triangleq \begin{pmatrix} \langle x_1(1), x_1(2), \ldots, x_1(s) \rangle \\ \vdots \\ \langle x_N(1), x_N(2), \ldots, x_N(s) \rangle \end{pmatrix}$$

q-ичный код

Столбец $\mathbb{Q} = (Q_1, Q_2, \ldots, Q_N)$, $Q_i \subset \mathcal{A}_q$, $i \in [N]$, будем называть гиперсловом. Определим гиперсумму q-ичных символов $a_1, a_2, \ldots, a_s \in \mathcal{A}_q$:

$$\langle a_1, \ldots, a_s \rangle \triangleq \bigcup_{i=1}^{s} \{a_i\}.$$

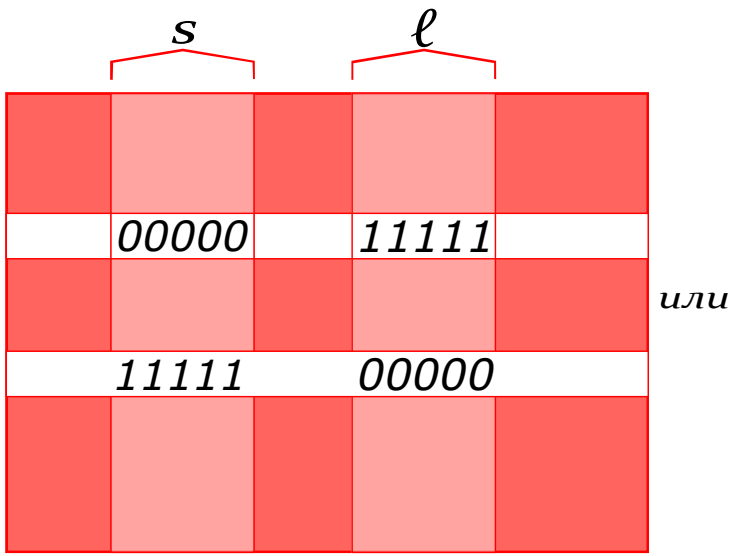
Будем говорить, что гиперслово $\mathbb{Q}$ подчиняет q-ичный столбец $\mathbf{a}$, если $a_i \in Q_i, \forall\, i \in [N]$.

Разделяющий код

q-ичный разделяющий (s, ℓ)-код

Код X называется q-**ичным разделяющим** (s, ℓ)-**кодом**[a,b], если в соответствующей ему матрице для для любых двух непересекающихся множеств столбцов $\mathcal{S}$ и $\mathcal{L}$, $|\mathcal{S}| \leqslant s$, $|\mathcal{L}| \leqslant \ell$, существует такая строка i, что в ней множества символов из столбцов $\mathcal{S}$ и $\mathcal{L}$ не пересекаются.

[a] Friedman A.D., Graham R.L., Ullman J.D., 1969.
[b] Сагалович Ю.Л., 1965. (случай $q = 2$)



Обозначим через $N_{sep}^{(q)}(t, s, \ell)$ минимальную длину q-ичного разделяющего (s, ℓ)-кода мощности t и определим скорость разделяющих (s, ℓ)-кодов:

$$R_{sep}^{(q)}(s, \ell) = \overline{\lim_{t \to \infty}} \frac{\log_q t}{N_{sep}^{(q)}(t, s, \ell)}.$$

Границы для скорости разделяющих кодов

Нижняя граница для скорости разделяющих кодов

При фиксированных $q \geqslant 2$, $\ell \geqslant 1$ и $s \to \infty$ скорость $R_{sep}^{(q)}(s, \ell)$ q-ичных разделяющих кодов удовлетворяет неравенству[a]

$$R_{sep}^{(q)}(s, \ell) \geqslant \frac{(q-1)^\ell}{e^\ell \ln q} \frac{1}{s^{\ell+1}}(1 + o(1)), \qquad s \to \infty.$$

[a] 2015, [ВДПЩ]

Верхняя граница для скорости разделяющих кодов

При фиксированных $q \geqslant 2$, $\ell \geqslant 1$ и $s \to \infty$ скорость $R_{sep}^{(q)}(s, \ell)$ q-ичных разделяющих кодов удовлетворяет неравенству[a]

$$R_{sep}^{(q)}(s, \ell) \leqslant \frac{\sum\limits_{k=1}^{\ell} \binom{q-1}{k} (\ell+1)^{\ell+1}}{2e^{\ell-1} \log_2 q} \cdot \frac{\log_2 s}{s^{\ell+1}} (1 + o(1)).$$

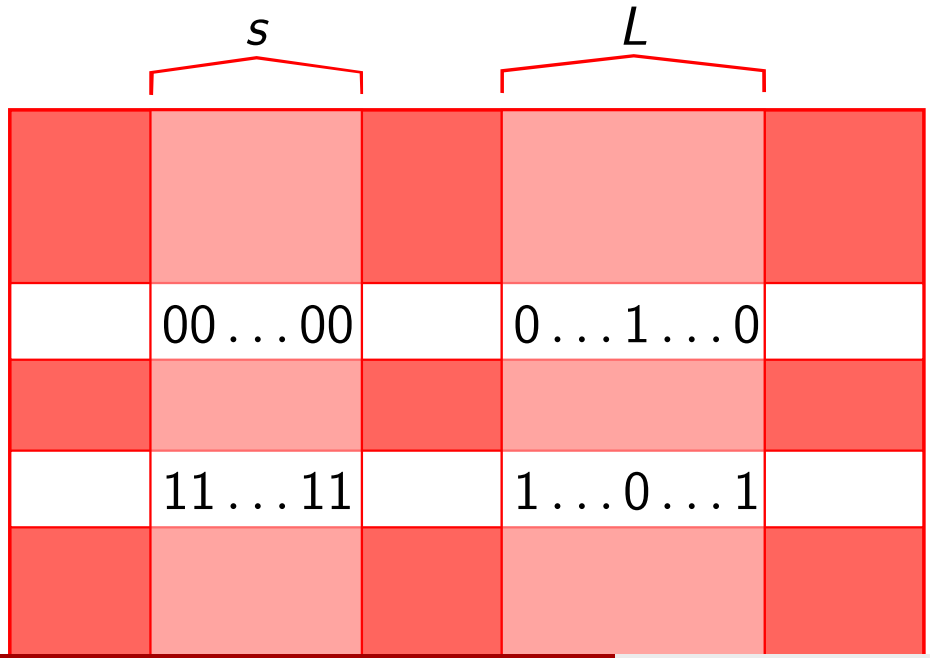
[a] 2016, [ВДПЩ]

СД s_L-гиперкоды

Определение 9

Код X_q называется гиперкодом со списочным декодированием силы s с объемом списка L (СД s_L-гиперкодом[a]), если гиперсумма любых s столбцов кода X_q подчиняет не более $L-1$ других столбцов кода X_q, не входящих в эту гиперсумму.

[a] [1998, D. Boneh, J. Shaw]



Обозначим через $t^{(q)}(N,s,L)$ максимальный объем СД s_L-гиперкодов длины N, и определим скорость СД s_L-гиперкодов:

$$R_L^{(q)}(s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_q t^{(q)}(N,s,L)}{N}$$

Границы для скорости СД гиперкодов

Верхняя граница для скорости СД гиперкодов

При любых фиксированных $q \geqslant 2$, $L \geqslant 1$ и $s \to \infty$ скорость q-ичных СД s_L-гиперкодов удовлетворяет неравенству[a]

$$R_L^{(q)}(s) \leqslant \frac{2L(q-1)\log_q s}{s^2}(1+o(1)), \quad s \to \infty.$$

[a] 2016, [ВДПЩ]

Нижняя граница для скорости СД гиперкодов

При любых фиксированных $q \geqslant 2$, $L \geqslant 1$ и $s \to \infty$ скорость q-ичных СД s_L-гиперкодов удовлетворяет неравенству[a]

$$R_L^{(q)}(s) \geqslant \frac{L(q-1)\log_q e}{s^2(\log_2 e)^2}(1+o(1)), \quad s \to \infty.$$

[a] 2016, [ВДПЩ]

Границы для скорости СД гиперкодов

Равенство для предела скорости СД гиперкодов

При любых фиксированных $s \geqslant 2$, $L \geqslant 1$ и $q \to \infty$ скорость q-ичных СД s_L-гиперкодов удовлетворяет равенству[a]

$$\lim_{q\to\infty} R_L^{(q)}(s) = \frac{L}{s+L-1}.$$

[a] 2017, [ВДПЩ]

Определим композицию от символов $a_1, \ldots, a_s \in \mathcal{A}_q$

$$\boldsymbol{comp}(a_1, \ldots, a_s) \triangleq (t_0, \ldots, t_{q-1}),$$

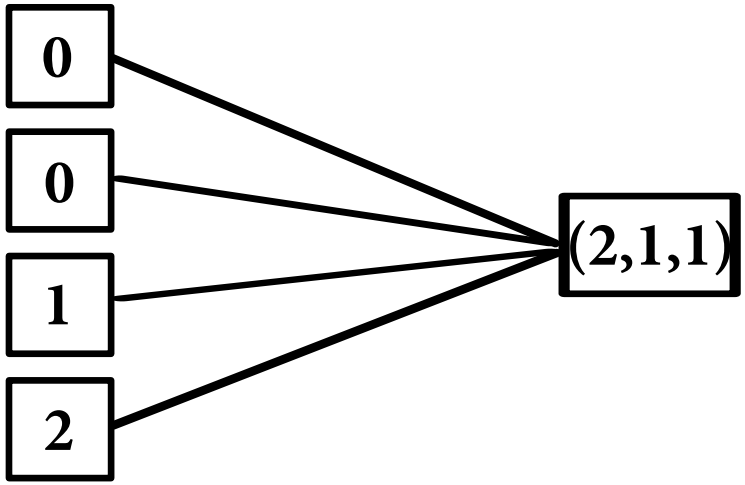
где $t_i \triangleq |\{j \in [s] : a_j = i\}|$.

Определим композицию от символов $a_1, \ldots, a_s \in \mathcal{A}_q$

$$\boldsymbol{comp}(a_1, \ldots, a_s) \triangleq (t_0, \ldots, t_{q-1}),$$

где $t_i \triangleq |\{j \in [s] : a_j = i\}|$.

Композиционный КМД



[a]1981, Ченг С., Вульф Д.

Определим композицию от символов $a_1, \ldots, a_s \in \mathcal{A}_q$

$$\boldsymbol{comp}(a_1, \ldots, a_s) \triangleq (t_0, \ldots, t_{q-1}),$$

где $t_i \triangleq |\{j \in [s] : a_j = i\}|$.

Тогда композиция от q-ичных векторов $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(s) \in \mathcal{A}_q^N$

$$\boldsymbol{comp}(\boldsymbol{x}(1), \ldots, \boldsymbol{x}(s)) \triangleq \begin{pmatrix} \boldsymbol{comp}(x_1(1), x_1(2), \ldots, x_1(s)) \\ \vdots \\ \boldsymbol{comp}(x_N(1), x_N(2), \ldots, x_N(s)) \end{pmatrix}$$

Композиционные коды

Определение 10

Код X_q называется композиционным s-кодом, если композиция любых $\leqslant s$ столбцов не совпадает с композицией любых $\leqslant s$ других столбцов.

Обозначим через $t_{comp}^{(q)}(N, s)$ максимальный объем композиционных s-кодов длины N и определим скорость композиционных s-кодов:

$$R_{comp}^{(q)}(s) \triangleq \varlimsup_{N\to\infty} \frac{\log_q t_{comp}^{(q)}(N, s)}{N}.$$

Границы композиционных кодов

Верхняя граница для скорости композиционных кодов

При фиксированном $q \geqslant 2$ и $s \to \infty$ скорость q-ичных композиционных s-кодов удовлетворяет неравенству[a,b]

$$R_{comp}^{(q)}(s) \leqslant \frac{q-1}{2} \frac{\log_q s}{s}(1 + o(1)), \quad s \to \infty.$$

[a] 2016, Егорова Е.Е., Потапова В.С.
[b] 1981, Дьячков А.Г., Рыков В.В. (случай $q = 2$)

Нижняя граница для скорости композиционных кодов

При фиксированном $q \geqslant 2$ и $s \to \infty$ скорость q-ичных композиционных s-кодов удовлетворяет неравенству[a,b]

$$R_{comp}^{(q)}(s) \geqslant \frac{q-1}{4} \frac{\log_q s}{s}(1 + o(1)), \quad s \to \infty.$$

[a] 2016, Егорова Е.Е., Потапова В.С.
[b] 1975, Дьячков А.Г. (случай $q = 2$)

Границы композиционных кодов

Верхняя граница для скорости композиционных кодов

При фиксированном $s \geqslant 2$ и $q \to \infty$ предел скорости q-ичных композиционных s-кодов удовлетворяет неравенству[a]

$$\overline{\lim_{q\to\infty}} R^{(q)}(s) \leqslant \frac{s+1}{2s}.$$

[a] 2017, [ВДПЩ]

Нижняя граница для скорости композиционных кодов

При фиксированном $s \geqslant 2$ и $q \to \infty$ предел скорости q-ичных композиционных s-кодов удовлетворяет неравенству[a]

$$\underline{\lim_{q\to\infty}} R^{(q)}(s) \geqslant \frac{s}{2s-1}.$$

[a] 2017, [ВДПЩ]

Раздел 3

Процедуры многошагового поиска. Адаптивный поиск скрытого гиперграфа.

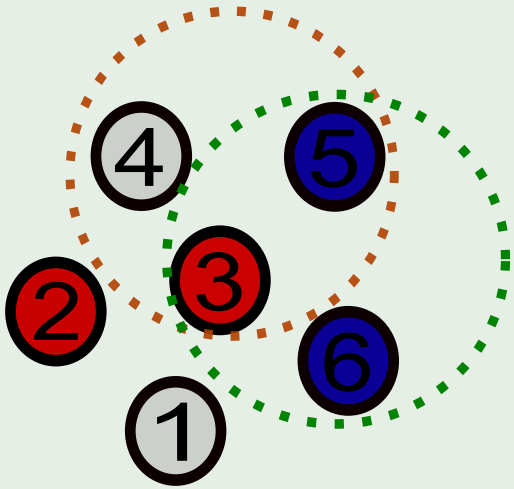
- Задача поиска скрытого гиперграфа.
- Границы скорости оптимальных алгоритмов поиска.
- Границы пропускной способности при поиске скрытого гиперграфа.

Задача поиска скрытого гиперграфа (супермножества)

Гиперграф — пара (V, E), где V — множество объектов или вершин гиперграфа, а E — семейство непустых подмножеств множества V, называемых рёбрами гиперграфа.

Пусть $|V| = t$ и фиксировано множество ребер E_{un}. С помощью группового тестирования необходимо обнаружить все ребра множества E_{un}. Тест — подмножество $T \subset V$. Результат теста положительный, если $\exists e \in E_{un}$ и $e \subset T$.

Пример 1



$x(1)$	$x(2)$	$x(3)$	$x(4)$	$x(5)$	$x(6)$	y
0	0	1	0	1	1	1
0	0	1	1	1	0	0

Задача адаптивного поиска скрытого гиперграфа

Пусть $|E_{un}| \leqslant s$ и $\forall e \in E_{un}$, $|e| \leqslant \ell$. Тогда (V, E_{un}) будем называть локализованным (s, ℓ)-гиперграфом. Для простоты будем считать $V = \{1, 2, \ldots, t\}$. Обозначим множество всех локализованных (s, ℓ)-гиперграфов через $H(t, s, \ell)$.

Алгоритм адаптивного поиска скрытого гиперграфа

Определим $N^a(t, s, \ell)$ — минимальное число тестов, необходимых для поиска гиперграфа в худшем случае, среди всех адаптивных алгоритмов поиска скрытого локализованного (s, ℓ)-гиперграфа.

Алгоритм многошагового поиска скрытого гиперграфа

Определим $N^{p\text{-st}}(t, s, \ell)$ — минимальное число тестов, необходимых для поиска гиперграфа в худшем случае, среди всех p-шаговых алгоритмов поиска скрытого локализованного (s, ℓ)-гиперграфа.

Задача адаптивного поиска скрытого гиперграфа

Скорость оптимальных алгоритмов адаптивного и многошагового поиска скрытого гиперграфа

Определим скорость оптимальных алгоритмов адаптивного поиска

$$R^{a}(s,\ell) \triangleq \varlimsup_{t\to\infty} \frac{\log_2 t}{N^{a}(t,s,\ell)}$$

и скорость оптимальных алгоритмов p-шагового поиска

$$R^{p\text{-st}}(s,\ell) \triangleq \varlimsup_{t\to\infty} \frac{\log_2 t}{N^{p\text{-st}}(t,s,\ell)}$$

Задача адаптивного поиска скрытого гиперграфа

Границы для скорости оптимальных алгоритмов

Для скорости оптимальных алгоритмов адаптивного (s, ℓ)-поиска выполнено равенство[a]

$$R^a(s, \ell) = \frac{1}{s\ell}.$$

Для скорости оптимальных многошаговых алгоритмов $(s, 1)$-поиска выполнены неравенства[a]

$$R^{p\text{-st}}(s, 1) \leqslant \frac{1}{s},$$

$$R^{(2s-1)\text{-st}}(s, 1) \geqslant \frac{1}{2s-1}.$$

Уточнение для $s = 2$

$$R^{4\text{-st}}(2, 1) = \frac{1}{2}.$$

[a] 2016, [ВДПЩ]

Задача неадаптивного поиска скрытого гиперграфа

Определение 11

X — план (s,ℓ)-поиска, если для любых локализованных (s,ℓ)-гиперграфов (V,E_1) и (V,E_2) выполнено условие

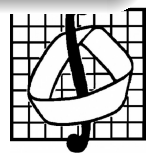
$$\bigvee_{e\in E_1}\bigwedge_{v\in e}\boldsymbol{x}(v)\neq\bigvee_{e\in E_2}\bigwedge_{v\in e}\boldsymbol{x}(v).$$

Связь неадаптивных планов поиска и СП кодов

Если X — дизъюнктивный СП (s,ℓ)-код, то X — план (s,ℓ)-поиска.

Если X — план (s,ℓ)-поиска, то X — дизъюнктивный СП $(s-1,\ell)$-код и дизъюнктивный СП $(s,\ell-1)$-код. [a]

[a] 2002, Виленкин П.А., Дьячков А.Г., Макула Э., Торни Д.



План (s, ℓ)-поиска с вероятностью ошибки

Определение 12

Гиперграф $(V, E) \in H(t, s, \ell)$, назовем (s, ℓ)-плохим для кода X, если существует хотя бы один другой гиперграф $(V, E') \in H(t, s, \ell)$ такой, что

$$\bigvee_{e \in E} \bigwedge_{v \in e} \boldsymbol{x}(v) \neq \bigvee_{e \in E'} \bigwedge_{v \in e} \boldsymbol{x}(v).$$

Через $B(s, \ell, X)$ обозначим количество (s, ℓ)-плохих гиперграфов для кода X. Зафиксируем скорость R, $R > 0$. Наименьшая ошибка планов (s, ℓ)-поиска с вероятностью ошибки:

$$\varepsilon(s, \ell, R, N) \triangleq \min_{X : t = \lfloor 2^{RN} \rfloor} \left\{ \frac{B(s, \ell, X)}{|H(t, s, \ell)|} \right\}.$$

Экспонента ошибки:

$$\mathbf{E}(s, \ell, R) \triangleq \varlimsup_{N \to \infty} \frac{-\log_2 \varepsilon(s, \ell, R, N)}{N}.$$

План (s, ℓ)-поиска с вероятностью ошибки

Пропускная способность планов (s, ℓ)-поиска с вероятностью ошибки:

$$C(s, \ell) \triangleq \sup\{ R \,:\, \mathbf{E}(s, \ell, R) > 0\}.$$

Граница для пропускной способности

Пропускная способность планов (s, ℓ)-поиска с вероятностью ошибки $C(s, \ell)$ удовлетворяет неравенству[a]

$$C(s, \ell) \geqslant \frac{(\ell+1)^{\ell+1}}{e^{\ell+1}} \frac{\log_2 e}{s^{\ell}}, \quad \ell \text{ фикс.}, \; s \to \infty.$$

[a] 2015, [ВДПЩ]

План (s, ℓ)-поиска с вероятностью ошибки

Аналогичным образом можно определить пропускная способность двухшаговых стратегий (s, ℓ)-поиска с вероятностью ошибки $C^{2\text{-st}}(s, \ell)$

Равенство для пропускной способности

Пропускная способность двухшаговых стратегий (s, ℓ)-поиска с вероятностью ошибки $C^{2\text{-st}}(s, \ell)$ удовлетворяет равенству[a]

$$C^{2\text{-st}}(s, \ell) = \frac{1}{s\ell}.$$

[a] 2016, [ВДПЩ]

Спасибо за внимание!

Публикации в регулярных журналах I

1. *Дьячков А.Г.*, *Воробьев И.В.*, *Полянский Н.А.*, *Щукин В.Ю.*, Границы скорости дизъюнктивных кодов, *Проблемы передачи информации*, Т. 50, № 1, С. 27-56, 2014.
2. *Дьячков А.Г.*, *Воробьев И.В.*, *Полянский Н.А.*, *Щукин В.Ю*, Почти дизъюнктивные коды со списочным декодированием, *Проблемы передачи информации*, Т. 51, № 2, С. 27-49, 2015.
3. *Полянский Н.А.*, Почти свободные от перекрытий коды, *Проблемы передачи информации*, Т. 52, № 2, С. 45-59, 2016.
4. *Щукин В.Ю*, Списочное декодирование для гиперканала множественного доступа, *Проблемы передачи информации*, Т. 52, № 4, С. 14-30, 2016.

Публикации в регулярных журналах II

5. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Symmetric Disjunctive List-Decoding Codes, *Designs, Codes, and Cryptography*, **82**:1 211-229, 2017.
6. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Almost Cover-Free Codes and Designs, *Designs, Codes, and Cryptography*, **82**:1 231-247, 2017.
7. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Cover-Free Codes and Separating System Codes, *Designs, Codes, and Cryptography*, **82**:1 197-209, 2017.
8. *Воробьев И.В.*, Границы разделяющих кодов, *Проблемы передачи информации*, Т. 53, № 1, С. 33-45, 2017.

Тезисы конференций I

1. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Almost disjunctive list-decoding codes (two talks), *Proc. 14th Int. Workshop on Algebraic and Combinatorial Coding Theory*, Svetlogorsk, pp. 115-126, 2014.
2. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Bounds on the Rate of Superimposed Codes, *2014 IEEE International Symposium on Information Theory*, Honolulu, pp. 2341-2345, 2014.
3. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Symmetric Disjunctive List-Decoding Codes, *2015 IEEE International Symposium on Information Theory*, Hong Kong, pp. 2236-2240, 2015.
4. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Almost Cover-Free Codes and Designs, *2015 IEEE International Symposium on Information Theory*, Hong Kong, pp. 2899-2903, 2015.

Тезисы конференций II

5. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Cover-Free Codes and Separating System Codes, *2015 IEEE International Symposium on Information Theory*, Hong Kong, pp. 2894-2898, 2015.
6. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, On a Hypergraph Approach to Multistage Group Testing Problems, *2016 IEEE International Symposium on Information Theory*, Barcelona, pp. 1183-1187, 2016.
7. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, On Multistage Learning a Hidden Hypergraph, *2016 IEEE International Symposium on Information Theory*, Barcelona, pp. 1178-1182, 2016.
8. *D'yachkov A.G.*, *Vorobyev I.V.*, *Polyanskii N.A.*, *Shchukin V.Yu.*, Threshold Decoding for Disjunctive Group Testing, *Proc. 15th Int. Workshop on Algebraic and Combinatorial Coding Theory*, Albena, pp. 145-150, 2016.