Sparse Regression Codes

Сенин Игорь

Московский Физико-Технический Институт

senin.ia@phystech.edu

2 апреля 2024 г.

Outline

- Основные понятия теории кодирования
- Конструкция SPARC'ов
- Оптимальный декодировщик
- План работы на оставшийся семестр

Модель канала с шумом

Пусть задано множество слов $\mathcal{A}$, которое хотим кодировать; $|\mathcal{A}| = N$. $\mathcal{B}$, Y - некоторые множества; $\mathcal{B}$ - передаваемые символы; Y - принимаемые символы. Пара (E, D) - *кодировщик* и *декодировщик*. $\alpha \xrightarrow{\text{E}} \beta \rightsquigarrow y \xrightarrow{\text{D}} \alpha'$, где $\beta \in \mathcal{B}^n$; $y \in Y^n$; $\alpha, \alpha' \in \mathcal{A}$.

Определение

- Шум вида $y_i = \beta_i + w_i$, $w_i \sim \mathcal{N}(0, \sigma^2)$ – i.i.d., $i = 1, \ldots, n$, будем называть **аддитивным белым гауссовским**, или просто **AWGN**.
- На канал накладывается **power constraint**: для *кодового слова* $\alpha = \alpha_1 \ldots \alpha_n$ должно выполняться $\frac{1}{n} \sum\limits_{i=1}^{n} \alpha_i^2 \leqslant P$. P будем называть **мощностью** канала.
- Отношение сигнал-шум (**signal-to-noise ratio**): $snr = \frac{P}{\sigma^2}$.

Пропускная способность канала

Определение

Пропускной способностью гауссовского канала мощности P называется величина $\mathcal{C} = \sup\limits_{\xi : \mathbb{E}(\xi^2) \leqslant P} I(\xi; \eta)$, где супремум берётся по всем распределениям символов на входе; ξ - входное распределение, $\eta = \eta(\xi)$ - распределение на выходе. $I(\xi; \eta)$ - *взаимная информация* случайных величин (распределений) ξ, η.

Утверждение

Для модели AWGN $\mathcal{C} = \frac{1}{2}\log(1 + snr)$.

Эффективность кода

Определение

Пропускной способностью (эффективностью, *rate*) пары (E_N, D_N) называется $\lim\limits_{N\to\infty} \frac{\log N}{n(N)}$. Будем обозначать $R = \frac{\log N}{n}$.

$\lambda_i = P(D(y) \neq a_i | \beta = E(a_i))$, где $a_i \in \mathcal{A} = \{a_1, ..., a_N\}$

$\lambda^{(n)} = \max\limits_{i=1,...,N} \lambda_i$

$P_e^{(n)} = \frac{1}{N} \sum\limits_{i=1}^{N} \lambda_i$

Эффективность R называется **достижимой**, если $\exists$ посл-ть $(N = 2^{Rn}, n)$-кодов, что $\lambda^{(n)} \to 0$.

Теорема Шеннона-Хартли

Теорема

$\mathcal{C}$ - точная верхняя грань всех достижимых R.

Для AWGN это означает, что

1. $\forall R < \mathcal{C} \ \exists (E, D)$ с эффективностью R, для которых $\lambda^{(n)} \to 0$ и $P_e^{(n)} \to 0$.
2. Если же $R > \mathcal{C}$, то $\forall (E, D)$ с эффективностью R выполняется $\lambda^{(n)} \geqslant const$ при всех достаточно больших N.

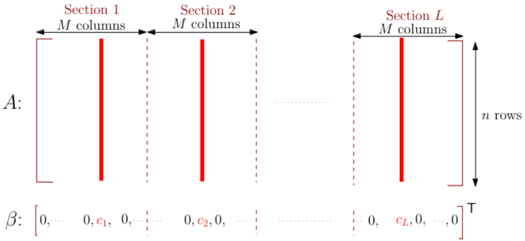
Достижимость $\mathcal{C}$

Цель - придумать код с быстрыми алгоритмами кодирования и декодирования, который (асимптотически) достигал бы пропускной способности гауссовского канала $\frac{1}{2}\log(1 + snr)$.

Хотя примеры таких кодов существуют (например, **гауссовские коды** - Cover, Thomas), они не используются на практике в том числе из-за экспоненциальной сложности декодирования.

SPARCs

В SPARC'ах кодовые слова - вектора длины n вида $A\beta$, где A - матрица плана n x ML, β - разреженный вектор длины ML.
A и β разбиты на L секций длины M, причём в каждой из L секций β ровно одна ненулевая координата c_l.



Параметры M, L, элементы A, c_1, ..., c_L фиксируются и считаются известными.

Выбор параметров

- Элементы матрицы A выбираются i.i.d. $\mathcal{N}(0, \frac{1}{n})$.
- L выбирается $\Theta(n/\log n)$.
- $M = L^a$ для некоторой константы $a > 0$.

Благодаря такому выбору M и L размер матрицы A растёт полиномиально с n.

Общее число кодовых слов - M^L.
$M^L = 2^{Rn} \Leftrightarrow L \log M = Rn$

Вероятность ошибки

Как сравнивать перформанс декодировщиков в теории?

Пусть $\hat{\beta}$ - декодированное сообщение; β - истинное сообщение.

Введём $\mathcal{E}_{sec} := \frac{1}{L} \sum_{i=1}^{L} I\{\beta_i \neq \hat{\beta}_i\}$ - доля неправильно декодированных секций.

Цель: для заданного декодировщика D ограничить $P(\mathcal{E}_{sec} > \epsilon)$.

Оптимальный декодировщик

Обозначим через $\mathcal{B}_{M,L}$ множество всех кодовых слов. Пусть на $\mathcal{B}_{M,L}$ задано априорное равномерное распределение.
Декодировщик максимального правдоподобия минимизирует вероятность ошибки декодирования. О.Н.К. даёт оптимальную оценку: $\hat{\beta}_{opt} = argmin_{\beta \in \mathcal{B}_{M,L}} ||y - A\beta||^2$.

Поскольку *argmin* берётся по дискретному множеству, аналитически решение не выражается. Поэтому лучше полного перебора β метода не существует.

Перформанс оптимального декодировщика

Фиксируются $c_l = \sqrt{\frac{nP}{L}}$.

Теорема

Пусть $M = L^a$, где $a \geqslant a_L^*(snr)$, $a_L^* = \max\limits_{\gamma \in \{\frac{1}{L}, \dots, 1-\frac{1}{L}\}} \frac{R \log \binom{L}{L\gamma}}{D_1(C_\gamma - \gamma C, \frac{\gamma(1-\gamma)snr}{1+\gamma snr}) L \log L}$; $\Delta = \mathcal{C} - R$ - строго положительный. Тогда для любой константы $\epsilon > 0$ выполняется $P(\mathcal{E}_{sec} \geqslant \epsilon) = e^{-nE(\epsilon, R)}$, где $E(\epsilon, R) = h(\epsilon, \Delta) - \frac{\log 2L}{n}$, $h(\epsilon, \Delta) = \min\{\epsilon \Delta w(snr), \frac{1}{4} g(\frac{\Delta}{2\sqrt{snr}})\}$.

Проще говоря, если правильно подобрать константы, то вероятность $\{\mathcal{E}_{sec} \geqslant \epsilon\}$ будет экспоненциально мала.

Конкатенация с кодом Рида-Соломона

От вероятности $P(\mathcal{E}_{sec} \geqslant \epsilon)$ можно перейти к вероятности *ошибки декодирования*: $P(\hat{\beta} \neq \beta)$.
Пусть $M = 2^m$. Рассмотрим $(n_{out}, k_{out}, n_{out} - k_{out} + 1)_M$ Р. С. код, где $n_{out} = M$; $k_{out} = \lceil (1 - \epsilon) M \rceil$.

Утверждение

Пусть SPARC - внутренний код с эффективностью R - построен так, что $R = \mathcal{C} - \Delta_n$, где Δ_n стремится к нулю медленнее $\frac{1}{\sqrt{n}}$ $\left(\frac{1}{\log n}, \frac{1}{n^{\frac{1}{4}}}\right)$.
Тогда композиция с кодом Р.С. даёт код эффективности $(\mathcal{C} - \Delta_n)(1 - \Delta_n)$ с $P(\hat{\beta} \neq \beta) \sim e^{(-kn\Delta_n^2)}$, где k - положительная константа, не зависящая от n.

План работы на оставшийся семестр

- Lossy Compression с SPARC и AMP декодирование
- Подобраться ближе к $\mathcal{C}$ на ограниченных n
- Исследовать SPARC на основе матриц Адамара