

Выбор функций потерь в задачах неотрицательного матричного разложения

Рябенко Евгений Алексеевич, ВМК МГУ
научный руководитель д.ф.-м.н., К.В. Воронцов

Семинар отдела прикладных проблем оптимизации
ВЦ РАН • 18 июня 2014 г.

Примеры прикладных задач неотрицательного матричного разложения

1 Коллаборативная фильтрация

$$R_{iu} = \sum_t p_{it} q_{tu}$$

дано: R_{iu} — рейтинги товаров i , поставленные пользователем u ;

найти: p_{it} — профиль интересов товара i ;

q_{tu} — профиль интересов пользователя u .

2 Тематическое моделирование текстовых коллекций

$$f_{wd} = \sum_t \phi_{wt} \theta_{td}$$

дано: f_{wd} — частоты слов w в документах d ;

найти: ϕ_{wt} — распределения слов w в темах t ;

θ_{td} — распределения тем t в документах d .

3 Анализ данных ДНК-микрочипов

$$I_{pk} = \sum_g a_{pg} c_{gk}$$

дано: I_{pk} — интенсивность флуоресценции p -й пробы на k -м чипе;

найти: a_{pg} — коэффициент сродства p -й пробы g -му гену;

c_{gk} — уровень экспрессии g -го гена на k -м чипе.

Дивергенции

Дивергенция — неотрицательная сепарабельная функция потерь с единственным нулём:

$$D(P, Q) = \sum_{i=1}^m \sum_{j=1}^n d(p_{ij}, q_{ij}),$$

$$d(p, q) \geq 0,$$

$$d(p, q) = 0 \Leftrightarrow p = q.$$

Название	$d(p, q)$
норма l_1	$d_1(p, q) = p - q $
квадрат нормы Фробениуса	$d_F(p, q) = (p - q)^2$
дивергенция Кульбака-Лейблера	$d_{KL}(p, q) = p \ln \frac{p}{q} - p + q$
дивергенция Итакура-Саито	$d_{IS}(p, q) = \ln \frac{q}{p} + \frac{p}{q} - 1$
расстояние Хеллингера	$d_H(p, q) = (\sqrt{p} - \sqrt{q})^2$
χ^2 Пирсона	$d_P(p, q) = \frac{(p-q)^2}{q}$
χ^2 Неймана	$d_N(p, q) = d_P(q, p) = \frac{(p-q)^2}{p}$

α - и β -дивергенции

α -дивергенция [Amari, 1985]:

$$d_A^\alpha(p, q) = \begin{cases} \frac{1}{\alpha(\alpha-1)} (p^\alpha q^{1-\alpha} - \alpha p + (\alpha-1)q), & \alpha \neq 0, \alpha \neq 1, \\ p \ln \frac{p}{q} - p + q, & \alpha = 1, \\ q \ln \frac{q}{p} - q + p, & \alpha = 0. \end{cases}$$

Непрерывно соединяет хи-квадрат Пирсона, дивергенцию Кульбака-Лейблера, расстояние Хеллингера и хи-квадрат Неймана.

β -дивергенция [Minami, Eguchi, 2002]:

$$d_B^\beta(p, q) = \begin{cases} \frac{1}{\beta(\beta+1)} (p^{\beta+1} - q^{\beta+1} - (\beta+1)q^\beta(p-q)), & \beta \neq 0, \beta \neq -1, \\ p \ln \frac{p}{q} - p + q, & \beta = 0, \\ \ln \frac{q}{p} + \frac{p}{q} - 1, & \beta = -1. \end{cases}$$

Непрерывно соединяет дивергенцию Кульбака-Лейблера, дивергенцию Итакура-Саито и норму Фробениуса.

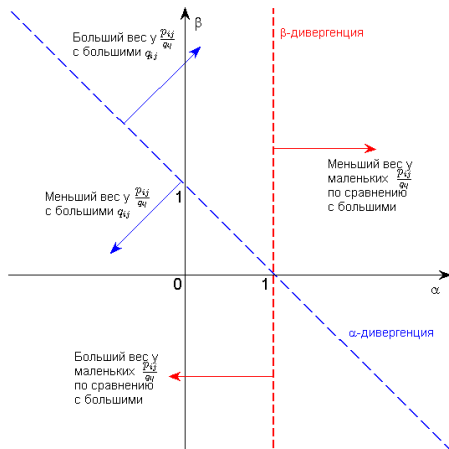
АБ-дивергенция

АБ-дивергенция [Cichocki et al., 2011]:

$$d_{AB}^{(\alpha, \beta)}(p, q) = \begin{cases} \frac{1}{\alpha\beta} \left(\frac{\alpha}{\alpha+\beta} p^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} q^{\alpha+\beta} - p^\alpha q^\beta \right), & \alpha, \beta, \alpha + \beta \neq 0, \\ \frac{1}{\alpha^2} \left(p^\alpha \ln \frac{p^\alpha}{q^\alpha} - p^\alpha + q^\alpha \right), & \alpha \neq 0, \beta = 0, \\ \frac{1}{\alpha^2} \left(\ln \frac{q^\alpha}{p^\alpha} + \left(\frac{q^\alpha}{p^\alpha} \right)^{-1} - 1 \right), & \alpha = -\beta \neq 0, \\ \frac{1}{\beta^2} \left(q^\beta \ln \frac{q^\beta}{p^\beta} - q^\beta + p^\beta \right), & \alpha = 0, \beta \neq 0, \\ \frac{1}{2} (\ln p - \ln q)^2, & \alpha = \beta = 0. \end{cases}$$

Объединяет α - и β -дивергенции, пересекающиеся в дивергенции Кульбака-Лейблера.

АБ-дивергенция



α регулирует разреженность модели [Minka, 2005],
 β определяет соотношение между эффективностью и устойчивостью получаемых оценок [Basu et al., 1998, Jones et al., 1998].

Выбор функции потерь

Различные функции потерь оптимальны для разных моделей шума.

Минимизация некоторых функций потерь эквивалентна максимизации правдоподобия в известных параметрических моделях:

Дивергенция	Модель шума	$p(P)$
Фробениуса	аддитивная гауссовская	$\prod_{ij} N(p_{ij}, \sigma^2)$
Кульбака-Лейблера	пуассоновская	$\prod_{ij} P(p_{ij})$
Итакура-Саито	мультипликативная гамма	$\prod_{ij} G(p_{ij}, \alpha/q_{ij})$

Идея: когда модель шума не задана, задачу выбора оптимальной функции потерь можно свести к выбору параметров в семействе АБ-дивергенций.

Задача оптимизации гиперпараметров α и β

Пусть существует семейство плотностей $p(P, \alpha, \beta)$, задаваемое в следующем виде:

$$p(P, \alpha, \beta) = \frac{1}{Z(\alpha, \beta)} p_0(P, \alpha, \beta),$$

$$p_0(P, \alpha, \beta) = e^{-D_{AB}^{(\alpha, \beta)}(P, Q)}, \quad (1)$$

$$Z(\alpha, \beta) = \int_X p_0(X, \alpha, \beta) dX.$$

Тогда оценка максимального правдоподобия для α и β может быть найдена как

$$(\alpha^*, \beta^*) = \operatorname{argmax}_{\alpha, \beta} \sum_{i, j} \ln p(p_{ij}, \alpha, \beta).$$

Проблема: $Z(\alpha, \beta) - ?$

Метод согласования вклада (score matching)

Поскольку нормировочный множитель $Z(\alpha, \beta)$ неизвестен, вместо метода максимизации правдоподобия можно использовать метод согласования вклада [Hyvärinen, 2006, 2007].

Пусть $p_T(x)$ — истинная плотность распределения данных,
 $p(x, \theta)$ — модельное семейство плотностей.

ОМП (оценка максимума правдоподобия):

$$\theta^* = \operatorname{argmin}_{\theta} \int_x p_T(x) \ln \frac{p_T(x)}{p(x, \theta)} dx.$$

ОСВ (оценка согласования вклада):

$$\theta^* = \operatorname{argmin}_{\theta} \int_x p_T(x) \left\| \nabla_x \ln \frac{p_T(x)}{p(x, \theta)} \right\|^2 dx.$$

$$\nabla_x \ln p(x, \theta) = \nabla_x \ln p_0(x, \theta) \implies$$

согласование вклада можно использовать, не зная $Z(\alpha, \beta)$.

Более того, не обязательно даже, чтобы он существовал [Hyvärinen, 2008].

Метод оптимизации гиперпараметров α и β

Теорема

В модели (1) ОСВ принимает следующий вид:

$$(\alpha^*, \beta^*) = \underset{\alpha, \beta}{\operatorname{argmin}} J(P, \alpha, \beta),$$

$$J(P, \alpha, \beta) = \begin{cases} \frac{1}{\beta} \sum_{i,j} p_{ij}^{\alpha} \left(\frac{1}{2\beta} p_{ij}^{\alpha} (p_{ij}^{\beta} - q_{ij}^{\beta})^2 - p_{ij}^{\beta} (\alpha + \beta + 1) + q_{ij}^{\beta} (\alpha + 1) \right), & \beta \neq 0, \\ \sum_{i,j} p_{ij}^{\alpha} \left(\ln \frac{q_{ij}}{p_{ij}} \left(\frac{p_{ij}^{\alpha}}{2} \ln \frac{q_{ij}}{p_{ij}} + \alpha + 1 \right) - 1 \right), & \beta = 0. \end{cases}$$

Оптимальные значения α и β предлагается находить, численно решая приведённую задачу минимизации.

Задача поиска неотрицательного матричного разложения при фиксированных α и β

Оптимизационная задача в общем виде:

$$(A^*, X^*) = \underset{A \geq 0, X \geq 0}{\operatorname{argmin}} D(P, AX).$$

$D(P, AX)$ не выпуклы по совокупности аргументов (A, X) , поэтому используются блочно-покоординатные методы минимизации:

Вход: матрица P , ранг разложения r , гиперпараметры α, β ;

Выход: матрицы-множители A и X ;

- 1 инициализация $A^0 \geq 0, X^0 \geq 0$;
- 2 **для всех** итераций $t = 1, 2, \dots$
- 3 $X^t = f(P, A^{t-1}, X^{t-1});$
- 4 $(A^t)^T = f(P^T, (X^t)^T, (A^{t-1})^T).$

В силу симметрии задачи для обновлений A и X можно использовать одну и ту же функцию.

Мультипликативный алгоритм

Идея: выбрать шаги градиентного спуска так, чтобы обновления стали мультипликативными, тогда будет сохраняться неотрицательность элементов матриц [Lee, Seung, 1999, 2001].

$$\begin{aligned}
 [\nabla_X]_{kj} &= \frac{\partial D(P, AX)}{\partial x_{kj}} = [\nabla_X^+]_{kj} - [\nabla_X^-]_{kj}, \\
 \nu_{kj} &= \frac{x_{kj}}{[\nabla_X^+]_{kj}}, \\
 x_{kj} &\leftarrow x_{kj} - \frac{x_{kj}}{[\nabla_X^+]_{kj}} \left([\nabla_X^+]_{kj} - [\nabla_X^-]_{kj} \right) = \\
 &= x_{kj} \frac{[\nabla_X^-]_{kj}}{[\nabla_X^+]_{kj}};
 \end{aligned}$$

в матричном виде:

$$X \leftarrow X \otimes \nabla_X^- \oslash \nabla_X^+.$$

\otimes — поэлементное умножение матриц, \oslash — поэлементное деление.

Мультипликативный алгоритм для нормы Фробениуса

Оптимизационная задача для нормы Фробениуса:

$$(A^*, X^*) = \underset{A \geq 0, X \geq 0}{\operatorname{argmin}} D_F(P, AX). \quad (2)$$

Мультипликативный алгоритм для нормы Фробениуса [Lee, Seung, 1999]:

$$\begin{aligned} \nabla_X D_F &= A^T AX - A^T P, \\ \nabla_A D_F &= AX X^T - P X^T; \\ X &\leftarrow X \otimes (A^T P) \oslash (A^T AX), \\ A &\leftarrow A \otimes (P X^T) \oslash (AX X^T). \end{aligned} \quad (3)$$

Правая часть — глобальный минимум квадратичной функции, мажорирующей D_F на текущей итерации [Lee, Seung, 2001]. Следовательно, функционал D_F монотонно невозрастает.

Сходимость мультипликативного алгоритма для нормы Фробениуса

Поскольку задача не является выпуклой, лучшее, что можно гарантировать — сходимость к стационарной точке, задаваемой условиями Каруша-Куна-Таккера:

$$\begin{aligned} X^* &\geq 0, & A^* &\geq 0, \\ \nabla_X D(P, A^* X^*) &\geq 0, & \nabla_A D(P, A^* X^*) &\geq 0, \\ X^* \otimes \nabla_X D(P, A^* X^*) &= 0, & A^* \otimes \nabla_A D(P, A^* X^*) &= 0. \end{aligned} \tag{4}$$

Проблема: обновления (3) могут останавливаться в нестационарных точках вблизи нулей: если $x_{kj} = 0$, то он останется равным нулю, даже если $[\nabla_X D_F]_{kj} < 0$.

ε -модификация мультипликативного алгоритма для нормы Фробениуса

Можно отделять элементы A и X от нуля небольшой положительной константой ε [Gillis, Glineur, 2012, Hibi, Takashi, 2011]:

$$\begin{aligned} X &\leftarrow \max \left(\varepsilon, X \otimes \left(A^T P \right) \oslash \left(A^T A X \right) \right), \\ A &\leftarrow \max \left(\varepsilon, A \otimes \left(P X^T \right) \oslash \left(A X X^T \right) \right). \end{aligned} \quad (5)$$

В [Gillis, Glineur, 2012] показано, что в алгоритме с такими обновлениями $D_F(P, AX)$ монотонно не возрастает, а любая предельная точка порождаемой последовательности (A^t, X^t) является стационарной точкой модифицированной задачи:

$$(A_\varepsilon^*, X_\varepsilon^*) = \underset{A \geq \varepsilon, X \geq \varepsilon}{\operatorname{argmin}} D_F(P, AX). \quad (6)$$

Утверждение о сходимости можно усилить.

Лемма

Мультипликативный алгоритм с обновлениями (5) сходится к стационарной точке отделённой от нуля задачи (6).

ε -прореживание для нормы Фробениуса

Получив решение задачи (6), проредим его, обнулив элементы, равные ε :

$$X = X_\varepsilon^* \otimes [X_\varepsilon^* > \varepsilon],$$

$$A = A_\varepsilon^* \otimes [A_\varepsilon^* > \varepsilon].$$

В [Gillis, Glineur, 2012] показано, что для A и X условия стационарности (4) исходной задачи (2) выполняются с точностью до $\mathcal{O}(\varepsilon)$: $\forall i, k, j$

$$\left\{ \begin{array}{l} \left[\begin{array}{l} a_{ik} = 0, \quad [\nabla_A D_F]_{ik} \geq -\mathcal{O}(\varepsilon), \\ a_{ik} > 0, \quad |[\nabla_A D_F]_{ik}| \leq \mathcal{O}(\varepsilon), \end{array} \right. \\ \left[\begin{array}{l} x_{kj} = 0, \quad [\nabla_X D_F]_{kj} \geq -\mathcal{O}(\varepsilon), \\ x_{kj} > 0, \quad |[\nabla_X D_F]_{kj}| \leq \mathcal{O}(\varepsilon), \end{array} \right. \end{array} \right.$$

Таким образом, ε -модификация мультипликативного алгоритма для нормы Фробениуса в сочетании с ε -прореживанием позволяют гарантированно получить решение, достаточно близкое к стационарной точке.

Сходимость для АБ-дивергенции

Оптимизационная задача для АБ-дивергенции:

$$(A^*, X^*) = \underset{A \geq 0, X \geq 0}{\operatorname{argmin}} D_{AB}^{(\alpha, \beta)}(P, AX). \quad (7)$$

Мультипликативный алгоритм для АБ-дивергенции [Cichocki et al., 2011]:

$$X \leftarrow X \otimes \left(\left(A^T \left(P^{[\alpha]} \otimes Q^{[\beta-1]} \right) \right) \oslash \left(A^T Q^{[\alpha+\beta-1]} \right) \right)^{[\omega(\alpha, \beta)]},$$

$$A \leftarrow A \otimes \left(\left(\left(P^{[\alpha]} \otimes Q^{[\beta-1]} \right) X^T \right) \oslash \left(Q^{[\alpha+\beta-1]} X^T \right) \right)^{[\omega(\alpha, \beta)]},$$

$$\omega(\alpha, \beta) = \begin{cases} \frac{1}{1-\beta}, & \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1, \\ \frac{1}{\alpha}, & \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right], \\ \frac{1}{\alpha+\beta-1}, & \frac{\beta}{\alpha} > \frac{1}{\alpha}. \end{cases}$$

$Z^{[z]}$ — поэлементное возведение матрицы Z в степень z .

При $\alpha = \beta = 1$ совпадает с мультипликативным алгоритмом для нормы Фробениуса.

В [Cichocki et al., 2011] показано, что в ходе обновлений функция потерь монотонно невозрастает. Однако, как и в предыдущем случае, обновления могут останавливаться в нестационарных точках вблизи нулей.

ε -модификация мультипликативного алгоритма для АБ-дивергенции

Отделим A и X от нуля небольшой положительной константой ε :

$$\begin{aligned}
 X &\leftarrow \max \left(\varepsilon, X \otimes \left(\left(A^T \left(P^{[\alpha]} \otimes Q^{[\beta-1]} \right) \right) \otimes \left(A^T Q^{[\alpha+\beta-1]} \right) \right)^{[\omega(\alpha, \beta)]} \right), \\
 A &\leftarrow \max \left(\varepsilon, A \otimes \left(\left(\left(P^{[\alpha]} \otimes Q^{[\beta-1]} \right) X^T \right) \otimes \left(Q^{[\alpha+\beta-1]} X^T \right) \right)^{[\omega(\alpha, \beta)]} \right).
 \end{aligned} \tag{8}$$

Теорема

При любом $\varepsilon > 0$ функция $D_{AB}^{(\alpha, \beta)}(P, AX)$ монотонно не возрастает при обновлениях (8) для любого начального приближения $A^0 \geq \varepsilon$, $X^0 \geq \varepsilon$.

Теорема

Алгоритм с обновлениями (8) для любого начального приближения $A^0 \geq \varepsilon$, $X^0 \geq \varepsilon$ сходится к стационарной точке отделённой от нуля задачи

$$(A_\varepsilon^*, X_\varepsilon^*) = \underset{A \geq \varepsilon, X \geq \varepsilon}{\operatorname{argmin}} D_{AB}^{(\alpha, \beta)}(P, AX). \tag{9}$$

ε -прореживание для АБ-дивергенции

Проредим решение задачи (9):

$$X = X_\varepsilon^* \otimes [X_\varepsilon^* > \varepsilon],$$

$$A = A_\varepsilon^* \otimes [A_\varepsilon^* > \varepsilon].$$

Теорема

Для матриц (A, X) , полученных из $(A_\varepsilon^*, X_\varepsilon^*)$ ε -прореживанием, верно следующее: $\forall i, k, j$

$$\left\{ \begin{array}{l} \left[\begin{array}{l} a_{ik} = 0, \quad [\nabla_A D_{AB}]_{ik} \geq -\mathcal{O}(\varepsilon), \\ a_{ik} > 0, \quad |[\nabla_A D_{AB}]_{ik}| \leq \mathcal{O}(\varepsilon), \end{array} \right. \\ \left[\begin{array}{l} x_{kj} = 0, \quad [\nabla_X D_{AB}]_{kj} \geq -\mathcal{O}(\varepsilon), \\ x_{kj} > 0, \quad |[\nabla_X D_{AB}]_{kj}| \leq \mathcal{O}(\varepsilon), \end{array} \right. \end{array} \right.$$

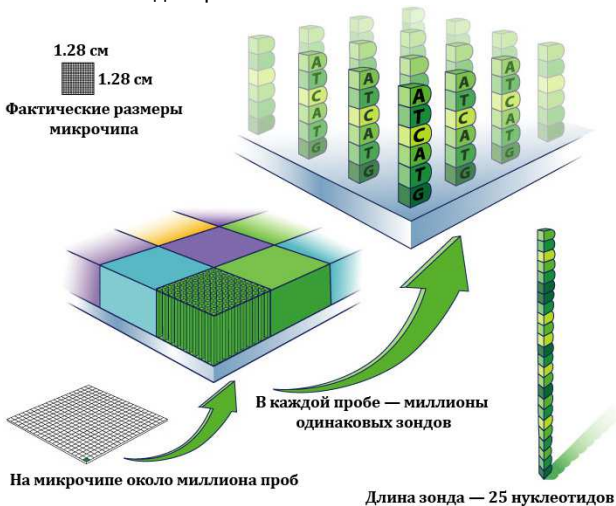
то есть, в точке (A, X) условия стационарности исходной задачи (7) выполняются с точностью до $\mathcal{O}(\varepsilon)$.

Результаты

- Для задачи неотрицательного матричного разложения предложен способ выбора оптимальной функции потерь в семействе АБ-дивергенций, основанный на методе согласования вклада.
- Предложен ε -модифицированный мультипликативный алгоритм неотрицательного матричного разложения с АБ-дивергенцией. Показано, что:
 - в ходе его применения функция потерь монотонно не возрастает;
 - алгоритм глобально сходится к стационарной точке оптимизационной задачи, отделённой от нуля;
 - модификация решения, полученная ε -прореживанием, даёт точку, условия стационарности исходной задачи в которой выполняются с точностью до $\mathcal{O}(\varepsilon)$.

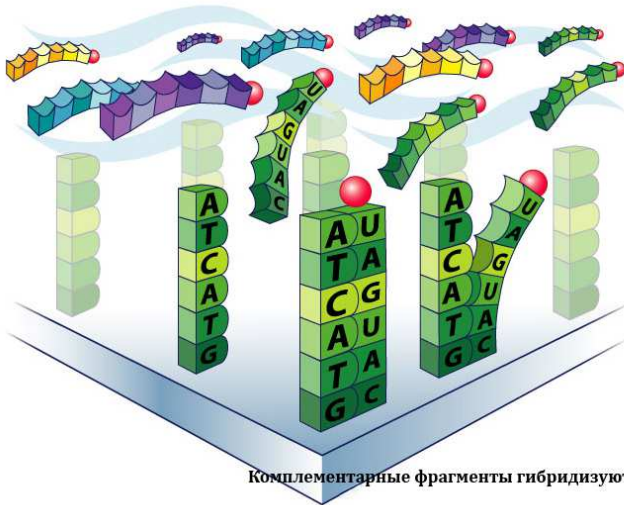
ДНК-микрочип

ДНК-микрочип — устройство, позволяющее оценивать экспрессию десятков тысяч генов одновременно.

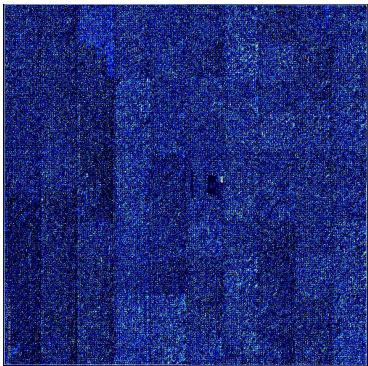


Гибридизация

Помеченные фрагменты одноцепочечной ДНК наносятся на чип



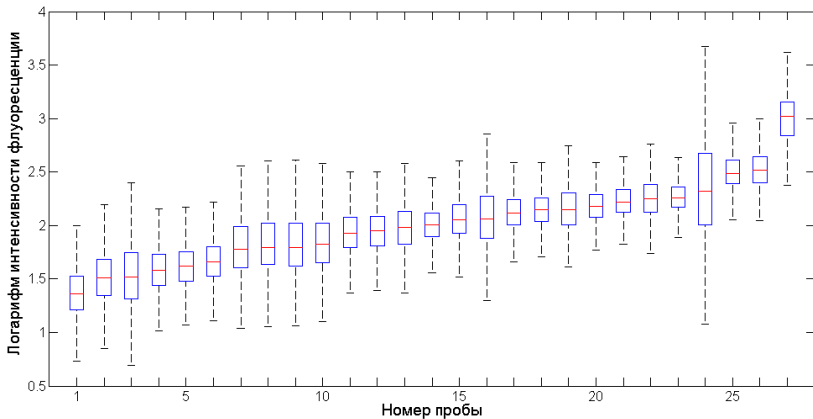
Получение оценок экспрессии



- 1 Изображение со сканера оцифровывается, получается вектор значений интенсивности флуоресценции проб.
- 2 Проводится предобработка интенсивностей:
 - фоновая поправка;
 - нормализация.
- 3 Значения предобработанных интенсивностей всех проб каждого гена **усредняются**, давая оценку экспрессии.

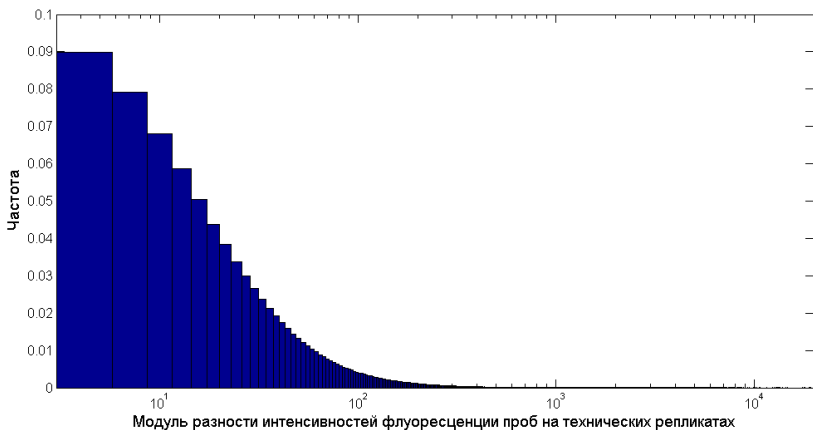
Распределения интенсивностей проб

Различия в физических свойствах проб вызывают систематические различия между интенсивностями флуоресценции:



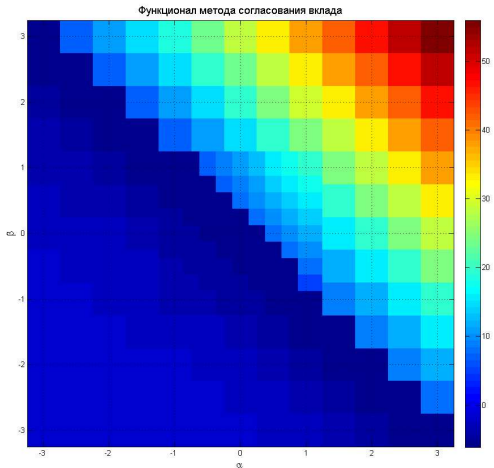
Шум

Технические репликаты — микрочипы, на которые был нанесён один и тот же образец:



Распределение шума имеет тяжёлые хвосты и не может быть оценено непосредственно — не существует экспериментов с известным сигналом.

Результаты оценки качества модели как функции от α и β



Функционал достигает минимума при $\alpha = -0.5$, $\beta = 0.75$, что достаточно близко, но всё же отличается от случая $\alpha = \beta = 0$, соответствующего традиционной логнормальной модели шума.

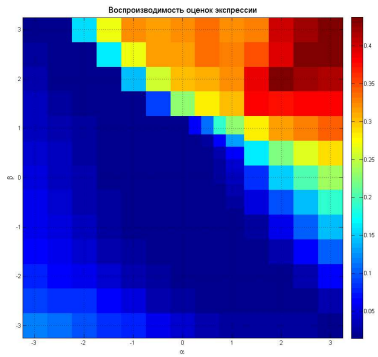
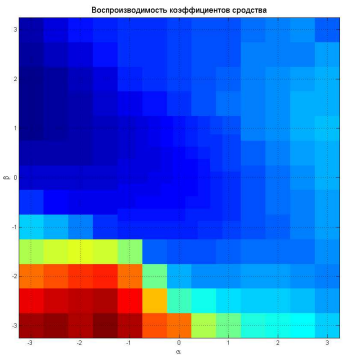
Результаты оценки качества модели как функции от α и β

Воспроизводимость коэффициентов сродства по двум подмножествам чипов:

$$rep_a = \frac{1}{G} \sum_{g=1}^G \frac{1}{P(g)} \sum_{p \in P(g)} \frac{|a_{pg}^1 - a_{pg}^2|}{a_{pg}^1 + a_{pg}^2}.$$

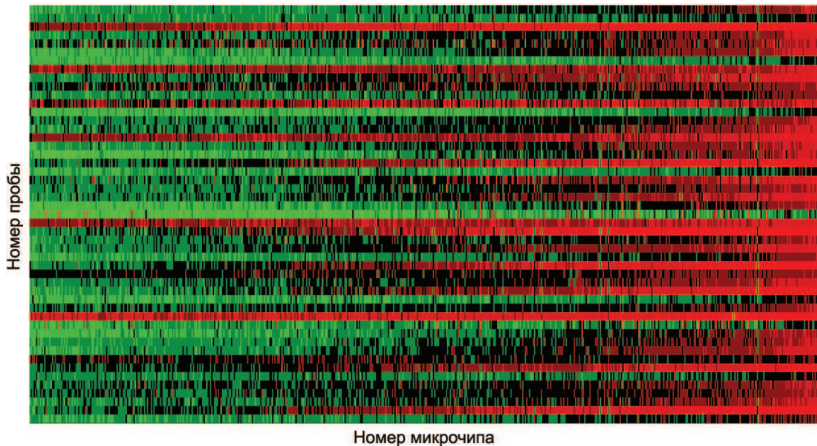
Воспроизводимость оценок экспрессии:

$$rep_c = \frac{1}{G} \sum_{g=1}^G \frac{1}{K} \sum_{k=1}^K \frac{|c_{gk}^1 - c_{gk}^2|}{c_{gk}^1 + c_{gk}^2}.$$



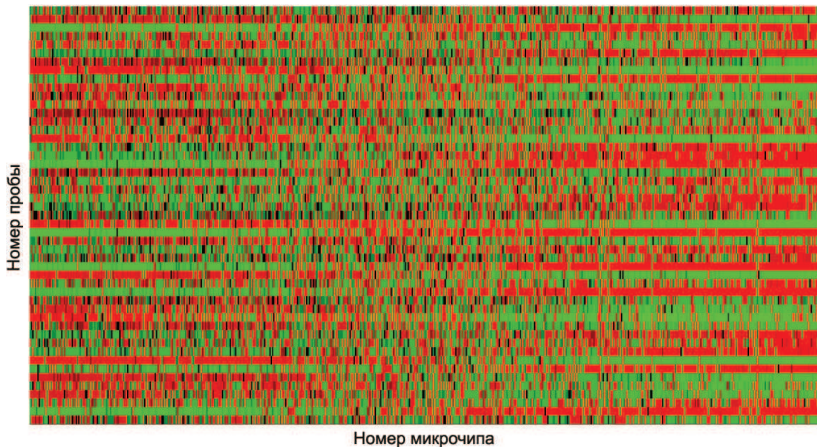
Альтернативный сплайсинг

Альтернативный сплайсинг — экспрессия **части** гена. Пробы к отсутствующим частям не оценивают экспрессию:



Эффект альтернативного сплайсинга в модели

На остатках настроенной модели виден эффект альтернативного сплайсинга:



Модель, учитывающая эффект кросс-гибридизации

Рассмотрим факторизованную модель ранга G :

$$I_{pk} \approx \hat{I}_{pk} = \sum_{g=1}^G a_{pg} c_{gk}.$$

Используем информацию о сходстве последовательностей проб и генов: положим $a_{pg} = 0$, если n_{pg} — число совпадающих нуклеотидов в пробе p и гене g — меньше 20.

Сформируем матрицу бинарных весов $W^A \in \{0, 1\}^{P \times G}$ с элементами $w_{pg}^A = [n_{pg} \geq 20]$.

Веса встраиваются в обновления матрицы A :

$$A \leftarrow W^A \otimes \max \left(\varepsilon, A \otimes \left(\left(\left(P^{[\alpha]} \otimes Q^{[\beta-1]} \right) X^T \right) \oslash \left(Q^{[\alpha+\beta-1]} X^T \right) \right)^{[\omega(\alpha, \beta)]} \right).$$

Неоднозначность разложения

NMF некорректно поставлена:

$$(A, X) \Leftrightarrow (AS, S^{-1}X),$$

где S — произвольная невырожденная матрица, сохраняющая неотрицательность компонент разложения.

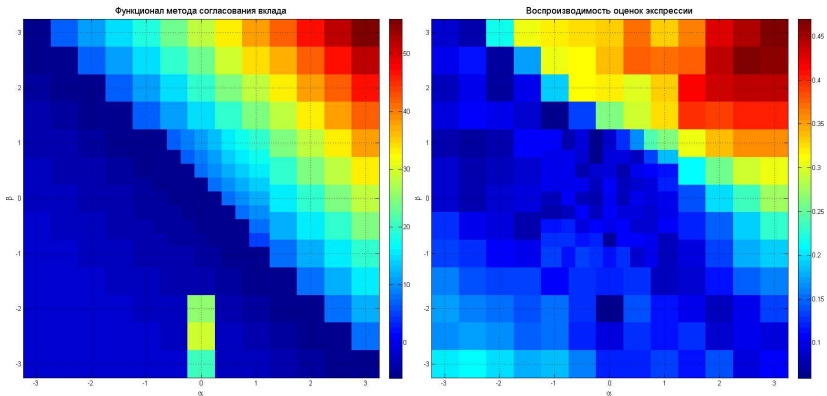
В частности, неотрицательность сохраняется, если $S \geq 0$ и $S^{-1} \geq 0$, т. е., S — обобщённая матрица перестановок \implies

Решение NMF находится с точностью до нормировки и перестановки столбцов A и строк X .

Достаточное условия единственности разложения с точностью до нормировки и перестановок: одна из двух матриц A и X содержит диагональную подматрицу размера r [Donoho, Stodden, 2004, Laurberg et al., 2008].

В рассматриваемой модели выполняется.

Результаты оценки качества модели как функции от α и β



Функционал метода согласования вклада достигает минимума при $\alpha = -0.5$, $\beta = 0.75$.

Публикации

- Рябенко, Е. А. (2014). Мультипликативный метод неотрицательного матричного разложения с АБ-дивергенцией и его сходимости. *Машинное обучение и анализ данных*, 1(7), 800–816.
- Крайнова, Н. А., Хаустова, Н. А., Макеева, Д. С., Федотов, Н. Н., Гудим, Е. А., Рябенко, Е. А., Шкурников, М. Ю., Галатенко, В. В., Сахаров, Д. А., Мальцева, Д. В. (2013). Оценка потенциальных референсных генов для нормализации данных ПЦР-РВ в экспериментах с клетками линии HeLa. *Биотехнология*, 1, 42–50.
- Рябенко, Е. А. (2012). Настройка нелинейной модели данных экспериментов с экспрессионными ДНК-микрочипами. *Математическая биология и биоинформатика*, 7(2), 554–566.
- Sakharov, D. A., Maltseva, D. V., Riabenko, E. A., Shkurnikov, M. U., Northoff, H., Tonevitsky, A. G., Grigoriev, A. I. (2012). Passing the anaerobic threshold is associated with substantial changes in the gene expression profile in white blood cells. *European journal of applied physiology*, 112(3), 963–972.
- Riabenko, E., Kogadeeva, M., Gavrilyuk, K., Sokolov, E., Shanin, I., Tonevitsky, A. G. (2012). Comparing Affymetrix Human Gene 1.0 ST preprocessing methods on tissue mixture data. 6th International Conference on Bioinformatics and Biomedical Engineering (iCBBE) (pp. 631–634). Shanghai, China.
- Мальцева, Д. В., Рябенко, Е. А., Сизова, С. В., Яшин, Д. В., Хаустова, С. А., Шкурников, М. Ю. (2012). Влияние физической нагрузки на экспрессию генов HSPBP1, PGLYRP1 и HSPA1A в лейкоцитах человека. *Бюллетень экспериментальной биологии и медицины*, 153(6), 846–850.
- Riabenko, E. A., Tonevitsky, E. A., Tonevitsky, A. G., Grigoriev, A. I. (2011). Structural Peculiarities of Human Genes Which Expression Increases in Response to Stress. *American Journal of Biomedical Sciences*, 3(2), 90–94.
- Рябенко, Е. А., Когадеева, М. С. (2011). Нижняя граница числа комплементарных нуклеотидов при моделировании кросс-гибридизации. ММРО-15, г. Петрозаводск. (с. 540–542). МАКС Пресс.
- Когадеева, М. С., Рябенко, Е. А. (2011). Математическая модель данных микрочипов ДНК, учитывающая эффекты кросс-гибридизации и насыщения. ММРО-15, г. Петрозаводск. (с. 536–539). МАКС Пресс.

Библиография I

- Affymetrix. (2005). Guide to Probe Logarithmic Intensity Error (PLIER) Estimation. Technical note. URL: http://media.affymetrix.com/support/technical/technotes/plier_technote.pdf
- Affymetrix. (2007). Sample Data, Gene 1.0 ST Data Set. URL: http://www.affymetrix.com/support/technical/sample_data/gene_1_0_array_data.affx (доступен после регистрации).
- Amari, S. (1985). *Differential-Geometrical Methods in Statistics*. Springer Verlag.
- Basu, A., Harris, I.R., Hjort, N., Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85, 549–559.
- Cichocki, A., Cruces, S., Amari, S. (2011). Generalized Alpha-Beta Divergences and Their Application to Robust Nonnegative Matrix Factorization. *Entropy*, 13(1), 134–170.
- Donoho, D., Stodden, V. (2004). When does non-negative matrix factorization give a correct decomposition into parts? In S. Thrun, L. K. Saul, B. Schölkopf (Eds.), *Advances in Neural Information Processing Systems* (pp. 1141–1148).
- Edgar, R., Domrachev, M., Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1), 207–210.
- Gillis, N., Glineur, F. (2012). Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization. *Neural Computation*, 24(4), 1085–105.
- Hibi, R., Takahashi, N. (2011). A modified multiplicative update algorithm for euclidean distance-based nonnegative matrix factorization and its global convergence. *Neural Information Processing*, 7063, 655–662.
- Hyvärinen, A. (2006). Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6, 695–708.
- Hyvärinen, A. (2007). Some extensions of score matching. *Computational Statistics & Data Analysis*, 51(5), 2499–2512.
- Hyvärinen, A. (2008). Estimation theory and information geometry based on denoising. *Proc. Workshop on Information Theory in Science and Engineering*. Tampere, Finland.

Библиография II

- Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2), 249.
- Jones, M., Hjort, N., Harris, I.R., Basu, A. (1998). A comparison of related density-based minimum divergence estimators. *Biometrika*, 85, 865–873.
- Minami, M., Eguchi, S. (2002). Robust blind source separation by beta divergence. *Neural Computation*, 14(8), 1859–86.
- Laurberg, H., Christensen, M. G., Plumbley, M. D., Hansen, L. K., Jensen, S. H. (2008). Theorems on positive data: on the uniqueness of NMF. *Computational Intelligence and Neuroscience*, (2), 764206.
- Lee, D. D., Seung, S. H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788–791.
- Lee, D. D., Seung, S. H. (2001). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems* (pp. 556–562).
- Li, C., Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America*, 98(1), 31–36.
- Minka, T. (2005). Divergence measures and message passing. Microsoft Research Cambridge, UK. <http://research.microsoft.com/en-us/um/people/minka/papers/message-passing/minka-divergence.pdf>
- Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F., Spencer, F. (2004). A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association*, 99(468), 909–917.