НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Неотрицательные матричные разложения

Научно-исследовательский семинар
Факультет компьютерных наук

Выполнил студент группы БПМИ182
Антон Медведев

2020

- Данные часто неотрицательны по своей природе
  - компьютерное зрение
  - рекомендательные системы
  - обработка аудиосигнала
  - ...

- Данные часто неотрицательны по своей природе
  - компьютерное зрение
  - рекомендательные системы
  - обработка аудиосигнала
  - ...
- Кластеризация

- Данные часто неотрицательны по своей природе
  - компьютерное зрение
  - рекомендательные системы
  - обработка аудиосигнала
  - ...
- Кластеризация
- Понижение размерности

Постановка задачи

$$P \approx AX \equiv Q$$

Постановка задачи

$P \approx AX \equiv Q$

$P, Q \in \mathbb{R}_+^{m \times n} \quad A \in \mathbb{R}_+^{m \times r} \quad X \in \mathbb{R}_+^{r \times n} \quad r < \min(m, n)$

Постановка задачи

$$P \approx AX \equiv Q$$
$$P, Q \in \mathbb{R}_+^{m\times n} \quad A \in \mathbb{R}_+^{m\times r} \quad X \in \mathbb{R}_+^{r\times n} \quad r < \min(m, n)$$
$$(A^*, X^*) = \arg\min_{A,X\geqslant 0} D(P, AX)$$

Постановка задачи

$P \approx AX \equiv Q$
$P, Q \in \mathbb{R}_+^{m \times n} \quad A \in \mathbb{R}_+^{m \times r} \quad X \in \mathbb{R}_+^{r \times n} \quad r < \min(m, n)$
$(A^*, X^*) = \arg\min\limits_{A,X \geqslant 0} D(P, AX)$

Определение

Функция $D(P, Q)$ называется дивергенцией, если

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

Постановка задачи

$P \approx AX \equiv Q$
$P, Q \in \mathbb{R}_+^{m\times n} \quad A \in \mathbb{R}_+^{m\times r} \quad X \in \mathbb{R}_+^{r\times n} \quad r < \min(m, n)$
$(A^*, X^*) = \operatorname*{arg\,min}_{A,X\geqslant 0} D(P, AX)$

Определение

Функция $D(P, Q)$ называется дивергенцией, если

- $D(P, Q) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_{ij}, q_{ij})$
- $d(p, q) \geqslant 0$

Постановка задачи

$$P \approx AX \equiv Q$$
$$P, Q \in \mathbb{R}_+^{m\times n} \quad A \in \mathbb{R}_+^{m\times r} \quad X \in \mathbb{R}_+^{r\times n} \quad r < \min(m, n)$$
$$(A^*, X^*) = \operatorname*{arg\,min}_{A,X\geqslant 0} D(P, AX)$$

Определение

Функция $D(P, Q)$ называется дивергенцией, если

- $D(P, Q) = \sum_{i=1}^{m}\sum_{j=1}^{n} d(p_{ij}, q_{ij})$
- $d(p, q) \geqslant 0$
- $d(p, q) = 0 \iff p = q$

- $d_F(p, q) = (p - q)^2$

- $d_F(p, q) = (p - q)^2$
- $d_{KL}(p, q) = p \ln \frac{p}{q} - p + q$

Постановка задачи
Дивергенции

- $d_F(p, q) = (p - q)^2$
- $d_{KL}(p, q) = p \ln \frac{p}{q} - p + q$
- $d_{IS}(p, q) = \ln \frac{q}{p} + \frac{p}{q} - 1$

Постановка задачи

Дивергенции

- $d_F(p, q) = (p - q)^2$
- $d_{KL}(p, q) = p \ln \frac{p}{q} - p + q$
- $d_{IS}(p, q) = \ln \frac{q}{p} + \frac{p}{q} - 1$
- $\begin{cases} \frac{1}{\alpha(\alpha-1)}(p^\alpha q^{1-\alpha} - \alpha p + (\alpha - 1)q), \alpha \notin \{0, 1\} \\ p \ln \frac{p}{q} - p + q, \quad \alpha = 1 \\ q \ln \frac{q}{p} - q + p, \quad \alpha = 0 \end{cases}$

- $d_F(p,q) = (p-q)^2$
- $d_{KL}(p,q) = p\ln\frac{p}{q} - p + q$
- $d_{IS}(p,q) = \ln\frac{q}{p} + \frac{p}{q} - 1$
- $\begin{cases} \frac{1}{\alpha(\alpha-1)}(p^\alpha q^{1-\alpha} - \alpha p + (\alpha-1)q), \alpha \notin \{0,1\} \\ p\ln\frac{p}{q} - p + q, \quad \alpha = 1 \\ q\ln\frac{q}{p} - q + p, \quad \alpha = 0 \end{cases}$
- $\begin{cases} \frac{1}{\beta(\beta+1)}(p^{\beta+1} - q^{\beta+1} - (\beta+1)q^\beta(p-q)), \beta \notin \{0,-1\} \\ p\ln\frac{p}{q} - p + q, \quad \beta = 0 \\ \ln\frac{q}{p} + \frac{p}{q} - 1, \quad \beta = -1 \end{cases}$

- NMF NP-трудна

- NMF NP-трудна
- решение задачи не является единственным

- NMF NP-трудна
- решение задачи не является единственным
- $D(P, AX)$ не выпукла по совокупности аргуметов (A, X)

- NMF NP-трудна
- решение задачи не является единственным
- $D(P, AX)$ не выпукла по совокупности аргуметов (A, X)
  - $X^t = f(P, A^{t-1}, X^{t-1})$
  - $(A^t)^T = f(P^T, (X^t)^T, (A^{t-1})^T)$

- NMF NP-трудна
- решение задачи не является единственным
- $D(P, AX)$ не выпукла по совокупности аргуметов (A, X)
  - $X^t = f(P, A^{t-1}, X^{t-1})$
  - $(A^t)^T = f(P^T, (X^t)^T, (A^{t-1})^T)$
- лучшее, что можно гарантировать — сходимость к стационарной точке

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$(A^*, X^*) = \underset{A,X \geqslant 0}{\arg\min} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

NMF с нормой Фробениуса

Мультипликативные обновления

$$(A^*, X^*) = \underset{A, X \geqslant 0}{\arg\min} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-)$$

NMF с нормой Фробениуса

Мультипликативные обновления

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$(A^*, X^*) = \arg\min_{A, X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$X \leftarrow X \otimes (A^T P) \oslash (A^T AX)$$

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$X \leftarrow X \otimes (A^T P) \oslash (A^T AX)$$

Оптимизация базового алгоритма:

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$X \leftarrow X \otimes (A^T P) \oslash (A^T AX)$$

Оптимизация базового алгоритма:

$$X \leftarrow \max\left(\varepsilon, X \otimes (A^T P) \oslash (A^T AX)\right)$$

NMF с нормой Фробениуса

Мультипликативные обновления

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$X \leftarrow X \otimes (A^T P) \oslash (A^T AX)$$

Оптимизация базового алгоритма:

$$X \leftarrow \max\left(\varepsilon, X \otimes (A^T P) \oslash (A^T AX)\right)$$

$$(A_\varepsilon^*, X_\varepsilon^*) = \arg\min_{A,X \geqslant \varepsilon} \|P - AX\|_F^2$$

$$(A^*, X^*) = \arg\min_{A,X \geqslant 0} \|P - AX\|_F^2$$

Базовый метод решения — поочерёдный градиентный спуск

$$\nabla_X = \nabla_X^+ - \nabla_X^-$$

$$X \leftarrow X - (X \oslash \nabla_X^+) \otimes (\nabla_X^+ - \nabla_X^-) = X \otimes \nabla_X^- \oslash \nabla_X^+$$

$$\nabla_X = 2A^T AX - 2A^T P$$

$$X \leftarrow X \otimes (A^T P) \oslash (A^T AX)$$

Оптимизация базового алгоритма:

$$X \leftarrow \max\left(\varepsilon, X \otimes (A^T P) \oslash (A^T AX)\right)$$

$$(A_\varepsilon^*, X_\varepsilon^*) = \arg\min_{A,X \geqslant \varepsilon} \|P - AX\|_F^2$$

$$A^* = A_\varepsilon^* \otimes [A_\varepsilon^* > \varepsilon] \quad X^* = X_\varepsilon^* \otimes [X_\varepsilon^* > \varepsilon]$$

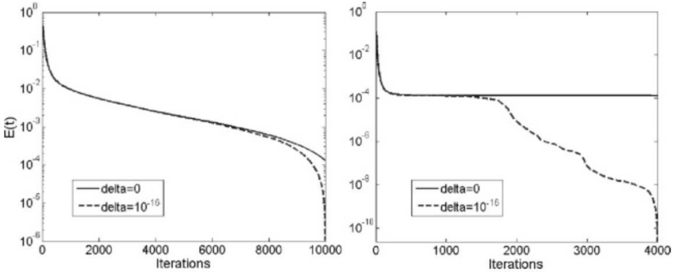
NMF с нормой Фробениуса

Сравнение MU-алгоритмов

$e^t = \|P - A^t X^t\|_F$

$E^t = \frac{e^t - e_{min}}{e^0 - e_{min}}$

e_{min} — наименьшая ошибка



[Gillis, Glineur, 2012]: E^t для обычного и модифицированного MU-алгоритмов на плотных (слева) и разреженных (справа) данных.

NMF с нормой Фробениуса

Сложность MU-алгоритма

Пусть Z — число ненулевых компонент в P, тогда на обновление X требуется:

шаг	число операций
$\mathrm{M}_1 = A^T P$	2Zr
$\mathrm{M}_2 = A^T A$	$2mr^2$
$\mathrm{M}_3 = M_2 X$	$2nr^2$
$\mathrm{X} \leftarrow X \otimes M_1 \oslash M_3$	2nr

Можно вычислить M_1 и M_2, а потом сделать несколько итераций по X!

MU-алгоритмы популярны, потому что

- просты в реализации

MU-алгоритмы популярны, потому что

- просты в реализации
- хорошо масштабируются и легко приспосабливаются к работе с разреженными данными

MU-алгоритмы популярны, потому что

- просты в реализации
- хорошо масштабируются и легко приспосабливаются к работе с разреженными данными
- были предложены в самой первой работе по NMF

MU-алгоритмы популярны, потому что

- просты в реализации
- хорошо масштабируются и легко приспосабливаются к работе с разреженными данными
- были предложены в самой первой работе по NMF

Главный минус — низкая скорость сходимости!

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2)$$

NMF с нормой Фробениуса

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2) = \max(0, (A^T A)^{-1} A^T P)$$

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2) = \max(0, (A^T A)^{-1} A^T P)$$

- Проблема: проецирование портит решение

NMF с нормой Фробениуса
ALS-алгоритм

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2) = \max(0, (A^T A)^{-1} A^T P)$$

- Проблема: проецирование портит решение
- Решение: на каждом шаге умножать обновляемую компоненту на $\alpha^* = \arg\min_{\alpha \geqslant 0} \|P - \alpha AX\|_F^2 = \frac{(PX^T, A)}{(A^T A, XX^T)}$

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2) = \max(0, (A^T A)^{-1} A^T P)$$

- Проблема: проецирование портит решение
- Решение: на каждом шаге умножать обновляемую компоненту на $\alpha^* = \arg\min_{\alpha \geqslant 0} \|P - \alpha AX\|_F^2 = \frac{(PX^T, A)}{(A^T A, XX^T)}$

Свойства алгоритма:

- метод очень грубый

NMF с нормой Фробениуса
ALS-алгоритм

$$X \leftarrow \max(0, \arg\min_X \|P - AX\|_F^2) = \max(0, (A^T A)^{-1} A^T P)$$

- Проблема: проецирование портит решение
- Решение: на каждом шаге умножать обновляемую компоненту на $\alpha^* = \arg\min_{\alpha \geqslant 0} \|P - \alpha AX\|_F^2 = \frac{(PX^T, A)}{(A^T A, XX^T)}$

Свойства алгоритма:

- метод очень грубый
- годится для инициализации других алгоритмов

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2$$

NMF с нормой Фробениуса
HALS-алгоритм

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

NMF с нормой Фробениуса

HALS-алгоритм

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

$$x_{kj} \leftarrow \underset{x_{kj} \geqslant 0}{\arg\min} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \underset{X_{k:} \geqslant 0}{\arg\min} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \underset{X_{k:} \geqslant \varepsilon}{\arg\min} \|P - AX\|_F^2 = \max(\varepsilon, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

NMF с нормой Фробениуса

HALS-алгоритм

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:k}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:k}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant \varepsilon} \|P - AX\|_F^2 = \max(\varepsilon, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:k}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

Свойства алгоритма:

- чувствителен к начальному приближению

NMF с нормой Фробениуса
HALS-алгоритм

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell:}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant \varepsilon} \|P - AX\|_F^2 = \max(\varepsilon, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell:}}{A_{:k}^T A_{:k}})$$

Свойства алгоритма:

- чувствителен к начальному приближению
- сходится быстрее, по сравнению с MU

NMF с нормой Фробениуса
HALS-алгоритм

$$x_{kj} \leftarrow \arg\min_{x_{kj} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P_{:j} - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} x_{\ell j}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant 0} \|P - AX\|_F^2 = \max(0, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

$$X_{k:} \leftarrow \arg\min_{X_{k:} \geqslant \varepsilon} \|P - AX\|_F^2 = \max(\varepsilon, \frac{A_{:k}^T P - \sum_{\ell \neq k} A_{:\ell}^T A_{:\ell} X_{\ell :}}{A_{:k}^T A_{:k}})$$

Свойства алгоритма:

- чувствителен к начальному приближению
- сходится быстрее, по сравнению с MU
- можно ускорить, используя внутренние итерации

- случайная инициализация

NMF с нормой Фробениуса

Инициализация

- случайная инициализация
- кластеризация

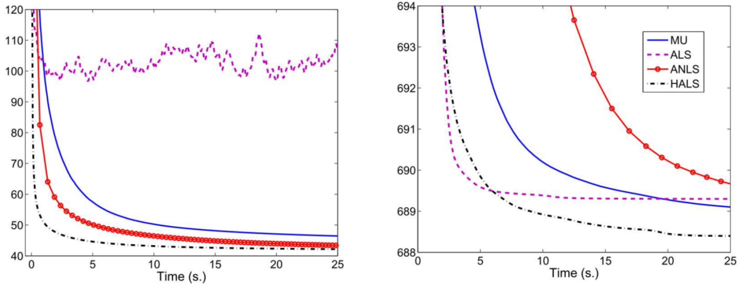
NMF с нормой Фробениуса
Инициализация

- случайная инициализация
- кластеризация
- ALS-алгоритм

NMF с нормой Фробениуса

Инициализация

- случайная инициализация
- кластеризация
- ALS-алгоритм
- мультистарт
  - с помощью ALS генерируем 10-20 пар матриц
  - делаем 10-20 итераций целевого метода на каждой паре
  - выбираем пару матриц с наименьшим значением функционала



[Gillis, 2014]: зависимость относительной точности приближения на плотных (слева) и разреженных (справа) данных.

Основные источники:

- Доклад Евгения Рябенко
- Запись выступления Евгения Рябенко
- A tutorial on NMF

Дополнительные источники:

- Диссертация Евгения Рябенко