Non-negative matrix factorization

Неотрицательные матричные разложения

Сергей Горбачев

12 октября 2018 г.

группа 162

Содержание

Постановка задачи

У нас есть неотрицательная прямоугольная матрица X размера $m \times n$
Мы хотим приблизить ее произведением двух прямоугольных неотрицательных матриц W и H

$$X \approx WH \equiv Q \qquad X, Q \in R_+^{m \times n} \tag{1}$$

$$W \in R_+^{m \times r},\ H \in R_+^{r \times n}$$

$$r < min(m, n)$$

Постановка задачи

Введя меру потерь (дивергенцию) $D(X, WH)$, получаем оптимизационную задачу:

$$(W^*, H^*) = \underset{W \geq 0,\ H \geq 0}{\arg\min}\ D(X, WH) \tag{2}$$

Примеры дивергенции

$$D(X, Q) = \sum_{i=1}^{m} \sum_{j=1}^{n} d(p_{ij}, q_{ij}) \tag{3}$$

$$d(p, q) > 0$$

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

Название	$d(p, q)$
норма l_1	$\lvert p - q \rvert$
норма Фробениуса	$(p - q)^2$
обобщённая дивергенция Кульбака-Лейблера	$p \ln \frac{p}{q} - p + q$
дивергенция Итакура-Саито	$\ln \frac{q}{p} + \frac{p}{q} - 1$
расстояние Хеллингера	$(\sqrt{p} - \sqrt{q})^2$
χ^2 Пирсона	$\frac{(\sqrt{p}-\sqrt{q})^2}{q}$
χ^2 Неймана	$\frac{(\sqrt{p}-\sqrt{q})^2}{p}$

Связь дивергенции с правдоподобием

Во многих случаях дивергенция соответсвует правдоподобию, то есть существует плотность $p(X|Q)$, такая, что при каких-то a и b

$$-\ln p(X|Q) = aD(X,Q) + b \tag{4}$$

Например норма Фробениуса соответсвует гауссовскому распределению

$$p(X|Q) = \prod_{ij} N(p_{ij}|\ q_{ij}, \sigma^2) = \prod_{ij} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(p_{ij}-q_{ij})^2}{2\sigma^2}} \tag{5}$$

Кульбака-Лейблера - пуассоновскому, Итакура-Саито - гамма распределению

Проблемы NMF

- NP-трудна
- Некорректно поставлена: если W_0, H_0 — решение задачи, то пара $W = W_0 D,\ H = D^{-1} H_0$ тоже может быть решением
- $D(X, WH)$ не выпукла по совокупности аргументов (W, H), поэтому обычно используют блочно-покоординатные методы минимизации, то есть поочереди обновляем W и H

$$H^t = f(X, W^{t-1}, H^{t-1}) \tag{6}$$

$$W^t = f(X^T, (H^t)^T, (W^{t-1})^T)^T$$

- Можем гарантировать только схождение к локальному минимуму

Стационарная точка

Локальный минимум по условиям Каруша-Куна-Таккера:

$$\begin{aligned} W^* \geq 0, \quad \nabla_W D(X, W^* H^*) &\geq 0, \quad W^* \otimes \nabla_W D(X, W^* H^*) = 0 \\ H^* \geq 0, \quad \nabla_H D(X, W^* H^*) &\geq 0, \quad H^* \otimes \nabla_H D(X, W^* H^*) = 0 \end{aligned} \tag{7}$$

Градиентный спуск

Поочередный градиентный спуск:

$$
\begin{aligned}
& h_{kj} \leftarrow h_{kj} - \nu_{kj} \frac{\partial D(X, WH)}{\partial h_{kj}} \\
& w_{ik} \leftarrow w_{ik} - \eta_{ik} \frac{\partial D(X, WH)}{\partial w_{ik}} \\
& k = 1, ..., r, i = 1, ..., m, j = 1, ..., n
\end{aligned}
\tag{8}
$$

В таком ввиде не подходит, так как не учитывает ограничение неотрицательности

Мультипликативные обновления

Выбираем шаг так, чтоб обновления были мультипликативными

$$
\begin{aligned}
\frac{\partial D(X, WH)}{\partial h_{kj}} &= [\nabla_H]_{kj} = [\nabla_H^+]_{kj} - [\nabla_H^-]_{kj} \\
\nu_{ki} &= \frac{h_{kj}}{[\nabla_H^+]_{kj}} \\
h_{kj} \leftarrow h_{kj} &- \frac{h_{kj}}{[\nabla_H^+]_{kj}}([\nabla_H^+]_{kj} - [\nabla_H^-]_{kj}) = h_{kj}\frac{[\nabla_H^-]_{kj}}{[\nabla_H^+]_{kj}}
\end{aligned}
\tag{9}
$$

В матричном ввиде:

$$
H \leftarrow H \otimes \nabla_H^- \oslash \nabla_H^+ \tag{10}
$$

$\otimes$ - поэлементное умножение матриц, $\oslash$ - поэлементное деление

Мультипликативные обновления

Плюсы

- просты в реализации
- маштабируются и приспосабливаются к работе с разряженными матрицами

Минусы

- Скорость сходимости невелика (можно увеличить обновляя одну матрицу несколько раз подряд)
- Если $h_{kj} = 0$, то он останется нулем, даже если $[\nabla_H]_{kj} < 0$, поэтому может не сойтись к локальному минимуму

NMF с нормой Фробениуса

Оптимизационная задача:

$$(W^*, H^*) = \underset{W \geq 0,\ H \geq 0}{\arg\min} ||X - WH||_F^2 \tag{11}$$

Без ограничения неотрицательности решение можно было бы получить с помощью SVD.

NMF с нормой Фробениуса

Для нормы Фробениуса мультипликативные обновления выглядят так:

$$\begin{aligned} \nabla_H &= W^T W H - W^T X \\ H &\leftarrow H \otimes (W^T X) \oslash (W^T W H) \end{aligned} \tag{12}$$

Известно, что при таком обновлении функция потерь монотонно невозрастает*

*Lee, Seung - Algorithms for Non-negative Matrix Factorization

Мультипликативные обновления с нормой Фробениуса

Модификации

- Отделять элементы W и H от нуля небольшой константой ε

$$H \leftarrow max(\varepsilon, H \otimes \nabla_H^- \oslash \nabla_H^+) \tag{13}$$

  Показанно, что этот алгоритм сходится к локальному решению такой задачи:

$$(W_\varepsilon^*, H_\varepsilon^*) = \underset{W \geq \varepsilon,\ H \geq \varepsilon}{\arg\min} ||X - WH||_F^2 \tag{14}$$

- После каждого шага реинициализировать ε только те элементы, для которых компонента градиента отрицательна

Метод попеременных наименьших квадратов

Alternating Least Squares (ALS): на каждом шаге находится решение задачи наименьших квадратов по одной из компонент, а затем берется проекция на неотрицательную область

$$\begin{aligned} H \leftarrow max(\arg\min_{Z \in R^{r \times n}} ||X - WZ||_F^2, 0) = \\ = max((W^T W)^{-1} W^T X, 0) \end{aligned} \tag{15}$$

Метод быстрый, но итерационный процесс не сходится, функция потерь не монотонна. Можно использовать для инициализации других методов.

Метод попеременных неотрицательных наименьших квадратов

Alternating Nonnegative Least Squares (ANLS): на каждом шаге находится покомпонентный минимум в неотрицательной области

$$H \leftarrow max(\arg\min_{H \geq 0} ||X - WH||_F^2, 0) \tag{16}$$

Есть разные методы нахождения покомпонентных минимумов, но все они очень долгие. Можно использовать для уточнения решения, найденного более простыми методами.

Метод иерархических попеременных наименьших квадратов

Hierarchical alternating least squares (HALS): на каждом шаге находится минимум в неотрицательной области по столбцу w_k или строке h_k

$$
\begin{aligned}
h_k \leftarrow & \operatorname*{arg\,min}_{h_k \geq 0} ||X - WH||_F^2 = \\
&= max(0, \frac{w_k^T X - \sum_{l \neq k} w_k^T w_l h_l}{w_k^T w_k})
\end{aligned}
\tag{17}
$$

Каждая итерация требует малых вычислений и сходится быстрее метода мультипликативных обновлений, однако чувствителен к начальным значениям

Метод иерархических попеременных наименьших квадратов

Hierarchical alternating least squares (HALS): на каждом шаге находится минимум в неотрицательной области по столбцу w_k или строке h_k

$$\begin{aligned} h_k \leftarrow \underset{h_k \geq 0}{\arg\min} \, ||X - WH||_F^2 = \\ = max(0, \frac{w_k^T X - \sum_{l \neq k} w_k^T w_l h_l}{w_k^T w_k}) \end{aligned} \tag{18}$$

Каждая итерация требует малых вычислений и сходится быстрее метода мультипликативных обновлений, однако чувствителен к начальным значениям

NMF с обобщённой дивергенцией Кульбака-Лейблера

Оптимизационная задача:

$$\begin{aligned}(W^*, H^*) &= \underset{W \geq 0,\ H \geq 0}{\arg\min}\ D_{KL}(X, WH) \\ &= \underset{W \geq 0,\ H \geq 0}{\arg\min} \sum_{i=1}^{m} \sum_{j=1}^{n} \left(p_{ij} \ln \frac{p_{ij}}{[WH]_{ij}} - p + [WH]_{ij} \right)\end{aligned} \tag{19}$$

NMF с обобщённой дивергенцией Кульбака-Лейблера

Метод мультипликативных обновлений

$$
\begin{aligned}
&[\nabla_H]_{k,j} = \sum_{i=1}^{m} w_{ik} - \sum_{i=1}^{m} w_{ik}\frac{x_{ij}}{[WH]_{ij}} \\
&h_{kj} \leftarrow h_{kj}\frac{\sum_{i=1}^{m} w_{ik}\frac{x_{ij}}{[WH]_{ij}}}{\sum_{i=1}^{m} w_{ik}} \\
&\nabla_H = W^T J_{m\times n} - W^T(X \oslash (WH)) \\
&H \leftarrow H \otimes (W^T(X \oslash (WH))) \oslash (W^T J_{m\times n})
\end{aligned} \tag{20}
$$

Функция потерь так же монотонно невозрастает

Инициализация алгоритмов

Методы сходятся локально, поэтому начальные значения играют большую роль

- Случайная
- Кластеризация столбцов X, центроиды инициализируют столбцы W, а H - матрица принадлежности к кластерам
- SVD для X находится r наибольших сингулярных троек (u_k, s_k, v_k)

$$W^0_{:k} = u_k^{+/-}\sqrt{s_k\frac{||v_k^{+/-}||}{||u_k^{+/-}||}}, \quad H^0_{k:} = v_k^{+/-}\sqrt{s_k\frac{||u_k^{+/-}||}{||v_k^{+/-}||}} \tag{21}$$

- ALS
- Мультистарт
  - генерируется с помощью ALS 10-20 пар
  - делается 10-20 итераций целевого метода на каждой паре
  - выбирается пара с наименьшим значением функционала

Применение к стохастическим матрицам

NMF может применяться к матрицам вероятностей, с единичной суммой столбцов или строк

$$\sum_{i=1}^{n} h_{ki} = 1 \quad \forall k$$

Показанно, что в локальных минимумах задачи минимизации с дивергенцией Кульбака-Лейблера, суммы по столбцам и строкам матриц X и Q^* совпадают, а значит матрица Q^* тоже будет стохастической, но нет алгоритмов гарантирующих сходимость к локальному минимуму

Применение к стохастическим матрицам

Так же мы можем потребовать, чтоб W и H тоже были стохастическими и учесть это при оптимизации

- Нормировка: после каждого шага нормируем матрицу

$$h_{kj} \leftarrow \frac{h_{kj}}{\sum_{l=1}^{r} h_{lj}} \tag{22}$$

  Может нарушаться монотонность убывания функции потерь.

- Репараметризация: перейдем от матрицы H к Z

$$h_{kj} \leftarrow \frac{z_{kj}}{\sum_{l=1}^{r} z_{lj}} \tag{23}$$

Содержание

References i

- Неотрицательные матричные разложения
- Lee, Seung - Algorithms for Non-negative Matrix Factorization