

Адаптивный нелинейный метод восстановления матрицы по частичным наблюдениям

О. Городницкий, М. Трофимов

Московский Физико-Технический Институт

Курс: Машинное Обучение и Анализ Данных
Группа 374, 2016 Весна

Цель исследования и задача

Цель исследования

Проверить гипотезу о параметризуемости отображений, получить использующий её метод.

Задача

Заполнить пропуски в матрице (тензоре).

Предложение

Использовать настраиваемую функцию представлений вместо фиксированной. Согласно гипотезе это позволит адаптировать функцию для конкретной задачи.

- 1 Generalized Low Rank Models, Madeleine Udell, Corinne Horn, Reza Zadeh, and Stephen Boyd May 5, 2015.
- 2 Matrix factorization techniques for recommender systems, Koren, R. Bell, and C. Volinsky, 2009
- 3 Parallel Matrix Factorization for Recommender System, Hsiang-Fu Yu, Cho-Jui Hsieh, Inderjit S. Dhillon, 2013
- 4 ADADELTA: An Adaptive Learning Rate Method, Matthew D. Zeiler, 2012

Постановка задачи

Даны:

- 1 Целевая матрица $M \in \mathbb{R}^{m \times n}$
- 2 множество индексов её известных элементов (наблюдений): $\mathcal{O} = \{(i_k, j_k)\}_{k=1}^l$, где $i_k \in [1, \dots, m], j_k \in [1, \dots, n]$.

Найти:

- 1 Аппроксимирующую матрицу A , минимизирующую MSE на множестве $\mathcal{D}_{test} = \{(i_k, j_k)\}_{k=1}^h$

Оптимизационная задача

$$\mathbf{A}^* = \operatorname{argmin}_{\mathbf{A} \in \mathbb{R}^{m \times n}} \frac{1}{|\mathcal{D}_{test}|} \sum_{(i_k, j_k) \in \mathcal{D}_{test}} (A[i_k, j_k] - M[i_k, j_k])^2$$

Матричная факторизация

- 1 Фиксируем ранг k целевой матрицы M
- 2 $M = UV^T$, $U \in R^{m \times k}$, $V \in R^{n \times k}$
- 3 Каждый элемент результирующей матрицы вычисляется, как скалярное произведение некоторых векторов (представлений), относящихся к определенным строке и столбцу

Итерационное сингулярное разложение

- 1 Применяем усеченное SVD к матрице, полученной на предыдущем шаге.
- 2 Заменяем потерянные элементы M величинами, полученными из усеченного SVD на предыдущем шаге.
- 3 Повторяем процесс
- 4 Так как SVD может быть применено только к полным матрицам, на первом шаге все потерянные элементы M заменяются средними по строкам.

Нейросетевая факторизация

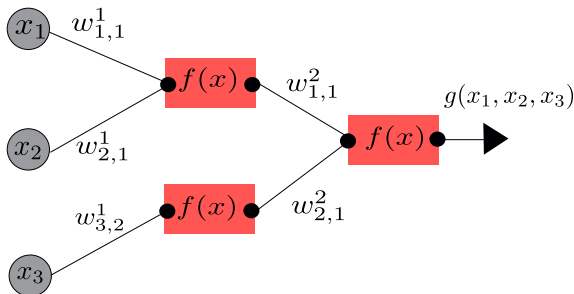
Требуется найти параметризацию

$\mathcal{F}(i, j, W) : d_1 \times d_2 \times R^m \rightarrow R$ и параметры $W \in R^m$ такие, что на некотором $\mathcal{D}_{test} = \{(i_k, j_k)\}_{k=1}^h$ достигается минимум MSE:

$$\mathcal{F}^*, W^* = \operatorname{argmin}_{\mathcal{F}, W} \frac{1}{|\mathcal{D}_{test}|} \sum_{(i_k, j_k) \in \mathcal{D}_{test}} (M[i_k, j_k] - \mathcal{F}(i_k, j_k, W))^2$$

Параметризация и параметры

$$g(x_1, x_2, x_3) = f(w_{1,1}^2 \times f(w_{1,1}^1 x_1 + w_{2,1}^1 x_2) + w_{2,1}^2 \times f(w_{3,2}^1 x_3))$$



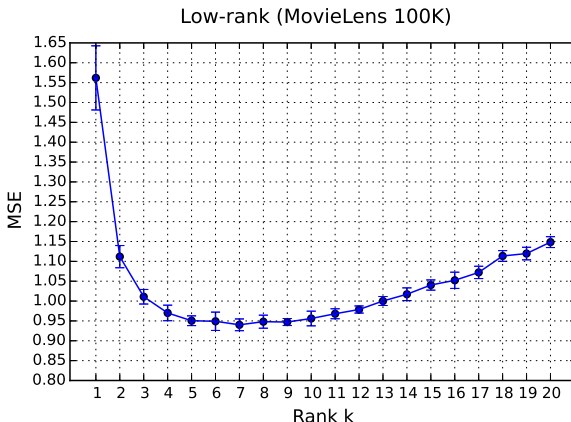
Параметризация - структура, параметры - веса

Теорема (Колмогоров) - любая непрерывная функция может быть приближена с произвольной точностью с помощью суперпозиций операции сложения и нелинейной функции одного аргумента.

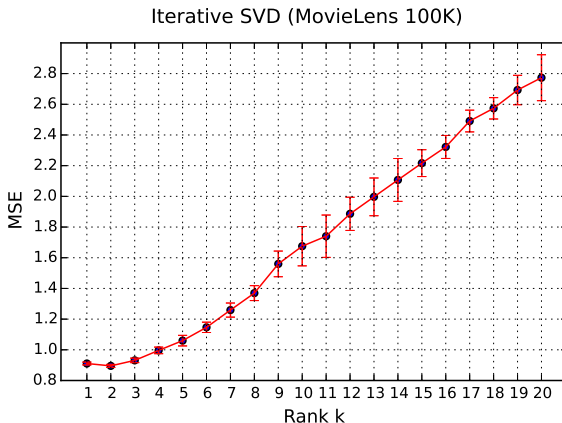
- 1 Исследовать задачу для случая, где \mathcal{F} определяется обучающим алгоритмом
- 2 Найти параметризацию, являющуюся оптимальной для данного набора данных
- 3 Получить качество рассматриваемых решений
- 4 Сравнить качество методов Матричной факторизации и Итерационного сингулярного разложения с точностью метода Нейросетевой факторизации.

Набор данных MovieLens 100K:

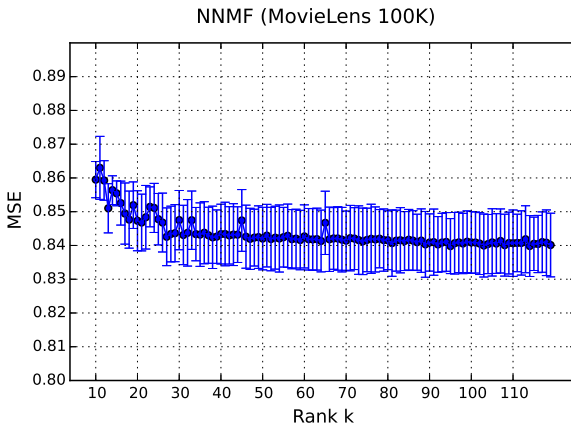
- 1 1000 пользователей
- 2 1700 фильмов
- 3 80000 оценок (ненулевых элементов)
- 4 каждый пользователь оценил не менее 20 фильмов
- 5 оценки выставлялись по шкале [1,2,3,4,5]



Зависимость ошибки от ранга k для матричной факторизации
Лучший результат: $\text{MSE} = 0.94$ при ранге $k = 7$



Зависимость ошибки от ранга k для Iterative SVD Лучший результат: $MSE = 0.895$, при ранге $k = 2$



Зависимость ошибки от длины представлений k для NNMF
Лучший результат: $\text{MSE} = 0.841$, при ранге $k = 95$

Лучшие MSE

	Low Rank	Iterative SVD	NNMF
MSE	0.94	0.895	0.841
Rank	7	2	95
SD	0.02	0.01	0.01

Метод извлекает пользу из больших значений ранга и устойчив при его значениях, превосходящих оптимум.

Метод показал себя успешно в рамках данной метрики. Возможно дальнейшее исследование работы метода с применением дополнительных техник (композиций, расширения множества скрытых переменных, ...) и метрик качества (nDCG, MAP, ...).