

Рекомендательные модели на основе матричной факторизации

Игнатов Дмитрий Игоревич ♦

- ♦ Национальный исследовательский университет Высшая школа экономики
Факультет компьютерных наук
Департамент анализа данных и искусственного интеллекта

2017



План занятия

- 1 Расширение модели SVD, SVD++ и time-SVD. Градиентный спуск.
- 2 Факторизационные машины (Factorization Machines).

Оглавление

1 Модель на основе идеи SVD. Градиентный спуск.

2 Факторизационные машины

Модель на основе идеи SVD. Градиентный спуск.

- Yehuda Koren, Robert M. Bell: Advances in Collaborative Filtering. Recommender Systems Handbook 2011: 145-186 (2015 edition)
- Yehuda Koren, Robert M. Bell, Chris Volinsky: Matrix Factorization Techniques for Recommender Systems. IEEE Computer 42(8): 30-37 (2009)

- Базовая модель:

$$r_{ui} = \mu + b_i + b_u + \mathbf{q}_i^T \mathbf{p}_u + e_{ui}$$

- Целевая функция:

$$\min_{b, \mathbf{q}, \mathbf{p}} = \sum_{(u,i) \in R} (r_{ui} - \mu - b_i - b_u - \mathbf{q}_i^T \mathbf{p}_u)^2 + \lambda(b_i^2 + b_u^2 + \|\mathbf{q}_i\|^2 + \|\mathbf{p}_u\|^2), \text{ где}$$

$$R = \{(u, i) \mid \text{оценка } r_{ui} \text{ определена}\}$$

- Метод градиентного спуска:

$$e_{ui} = r_{ui} - \hat{r}_{ui}$$

$$b_u \leftarrow b_u + \gamma \cdot (e_{ui} - \lambda \cdot b_u)$$

$$b_i \leftarrow b_i + \gamma \cdot (e_{ui} - \lambda \cdot b_i)$$

$$\mathbf{q}_i \leftarrow \mathbf{q}_i + \gamma \cdot (e_{ui} \cdot \mathbf{p}_u - \lambda \cdot \mathbf{q}_i)$$

$$\mathbf{p}_u \leftarrow \mathbf{p}_u + \gamma \cdot (e_{ui} \cdot \mathbf{q}_i - \lambda \cdot \mathbf{p}_u)$$

Модель SVD++.

Yehuda Koren: Factor in the neighbors: Scalable and accurate collaborative filtering. TKDD 4(1): 1:1-1:24 (2010)

- Можно добавлять дополнительную информацию в модель. Например, введём дополнительный набор факторов для продуктов y_i , которые будут характеризовать пользователя на основе того, что он просматривал, но не оценивал.
- Модель принимает вид

$$\hat{r}_{u,i} = \mu + b_u + b_i + \mathbf{q}_i^\top \left(\mathbf{p}_u + \frac{1}{\sqrt{|N(u)|}} \sum_{a \in N(u)} \mathbf{y}_a \right),$$

где $N(u)$ – множество продуктов, которые просматривал этот пользователь ($\frac{1}{\sqrt{|N(u)|}}$ т.н. нормализация).

Факторизация для случая неявного отклика.

Yifan Hu, Yehuda Koren, Chris Volinsky: Collaborative Filtering for Implicit Feedback Datasets. ICDM 2008: 263-272

- $p_{ui} = [r_{ui} > 0]^1$ – бинаризованные предпочтения.
- Целевая функция:
$$\min_{\mathbf{x}^*, \mathbf{y}^*} = \sum_{(u,i)} c_{ui} (p_{ui} - \mathbf{x}_u^T \mathbf{y}_i)^2 + \lambda (\|\mathbf{x}_u\|^2 + \|\mathbf{y}_i\|^2), \text{ где}$$

суммирование производится по всем парам (u, i) .
- $c_{ui} = 1 + \alpha r_{ui}$ – возможный выбор степени достоверности, где r_{ui} , например, частота просмотров телесериала, а $\alpha = 40$.
- SGD решение неэффективно, применяется схема ALS.

¹Нотация Айверсона: $[z] = 1$, если z истинно, иначе $[z] = 0$

Время в коллаборативной фильтрации: модель timeSVD++.

Yehuda Koren: Collaborative filtering with temporal dynamics. Commun. ACM 53(4): 89-97 (2010)

- Пример: давайте добавим время, т.е. будем рассматривать базовые предикторы и характеристики пользователя как функции от времени:

$$\hat{r}_{u,i} = \mu + b_u(t) + b_i(t) + \mathbf{q}_i^\top \mathbf{p}_u(t),$$

где

$$b_i(t) = b_i + b_{i, \text{Bin}(t)},$$

$$b_u(t) = b_u + \alpha_u \text{dev}_u(t) + b_{u,t},$$

$$p_{u,f}(t) = p_{u,f} + \alpha_{u,f} \text{dev}_u(t) + p_{u,f,t} + \frac{1}{\sqrt{|N(u)|}} \sum_{b \in N(u)} y_{bf},$$

$$\text{dev}_u(t) = \text{sign}(t - t_u) |t - t_u|^\beta.$$

- Один из основных компонентов модели, взявшей Netflix Prize.

Реализации

- [Software at RecSys Wiki](#)
- [Surprise](#) (новинка)

Оглавление

1 Модель на основе идеи SVD. Градиентный спуск.

2 Факторизационные машины

Factorization Machines: базовые статьи

- Steffen Rendle: Factorization Machines with libFM. ACM TIST 3(3): 57 (2012) [PDF](#)
- Steffen Rendle: Factorization Machines. ICDM 2010: 995-1000 [PDF](#)

Factorization Machines: пример

- $U = \{\text{Alice (A), Bob (B), Charlie (C), } \dots \}$
- $I = \{\text{Titanic (TI), Notting Hill (NH), Star Wars (SW), Star Trek (ST), } \dots \}$
- $S = \{(A, \text{TI}, 2010-1, 5), (A, \text{NH}, 2010-2, 3), (A, \text{SW}, 2010-4, 1), \dots$
 $(B, \text{SW}, 2009-5, 4), (B, \text{ST}, 2009-8, 5), (C, \text{TI}, 2009-9, 1), (C, \text{SW}, 2009-12, 5)\}$

Кодирование данных

Feature vector \mathbf{x}															Target y							
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	0	...	0.3	0.3	0.3	0	...	13	0	0	0	0	...	5	$y^{(1)}$
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	0	...	0.3	0.3	0.3	0	...	14	1	0	0	0	...	3	$y^{(2)}$
$\mathbf{x}^{(3)}$	1	0	0	...	0	0	1	0	...	0.3	0.3	0.3	0	...	16	0	1	0	0	...	1	$y^{(2)}$
$\mathbf{x}^{(4)}$	0	1	0	...	0	0	1	0	...	0	0	0.5	0.5	...	5	0	0	0	0	...	4	$y^{(3)}$
$\mathbf{x}^{(5)}$	0	1	0	...	0	0	0	1	...	0	0	0.5	0.5	...	8	0	0	1	0	...	5	$y^{(4)}$
$\mathbf{x}^{(6)}$	0	0	1	...	1	0	0	0	...	0.5	0	0.5	0	...	9	0	0	0	0	...	1	$y^{(5)}$
$\mathbf{x}^{(7)}$	0	0	1	...	0	0	1	0	...	0.5	0	0.5	0	...	12	1	0	0	0	...	5	$y^{(6)}$
	A	B	C	...	TI	NH	SW	ST	...	TI	NH	SW	ST	...	Time	TI	NH	SW	ST	...		
	User				Movie					Other Movies rated						Last Movie rated						

Двумерная модель

A. Factorization Machine Model

1) *Model Equation*: The model equation for a factorization machine of degree $d = 2$ is defined as:

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (1)$$

where the model parameters that have to be estimated are:

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times k} \quad (2)$$

And $\langle \cdot, \cdot \rangle$ is the dot product of two vectors of size k :

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle := \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (3)$$

A row \mathbf{v}_i within \mathbf{V} describes the i -th variable with k factors. $k \in \mathbb{N}_0^+$ is a hyperparameter that defines the dimensionality of the factorization.

Двумерная модель

A 2-way FM (degree $d = 2$) captures all single and pairwise interactions between variables:

- w_0 is the global bias.
- w_i models the strength of the i -th variable.
- $\hat{w}_{i,j} := \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ models the interaction between the i -th and j -th variable. Instead of using an own model parameter $w_{i,j} \in \mathbb{R}$ for each interaction, the FM models the interaction by factorizing it. We will see later on, that this is the key point which allows high quality parameter estimates of higher-order interactions ($d \geq 2$) under sparsity.

Связь с моделью коллаборативной фильтрации

$$n := |U \cup I|, x_j := \delta(j = i \vee j = u)$$

$$\hat{y}(x) = w_0 + w_u + w_i + \langle \mathbf{v}_u, \mathbf{v}_i \rangle$$

Ссылки на реализации FM и модификации

- LibFM <http://www.libfm.org>
- <https://github.com/ibayer/fastFM> [статья]
- <https://github.com/coreylynch/pyFM>
- <https://github.com/scikit-learn-contrib/polylearn> [статья]

Вопросы и контакты

www.hse.ru/staff/dima

Спасибо!

dmitrii.ignatov[at]gmail.com

dignatov[at]hse.ru