

Гибридные рекомендательные системы



Юлия Пономарева

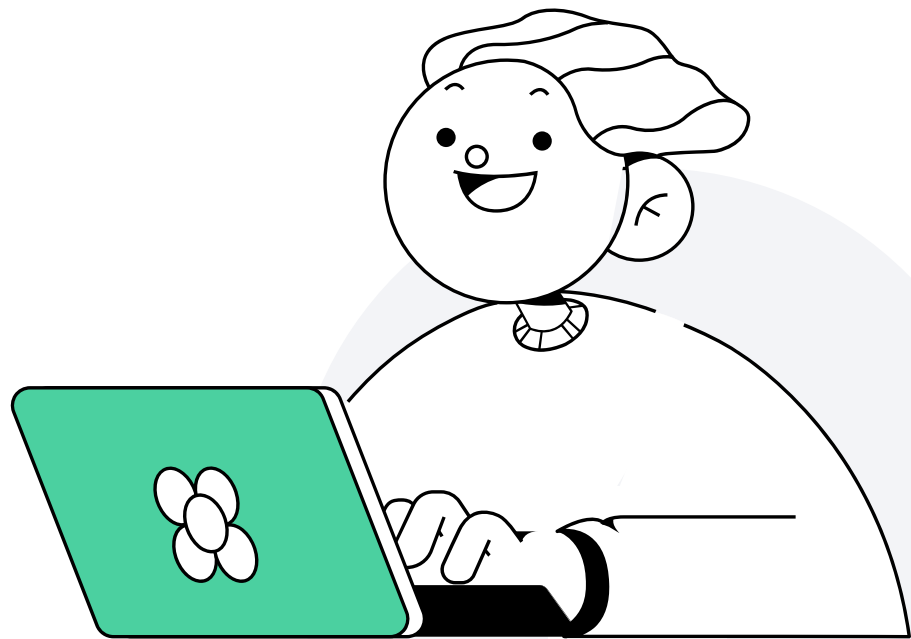
О спикере:

- Data Scientist
- автор YouTube-канала [machine learrning](#)
- работала в ITMO и Napoleon IT



План занятия

- ① Типология гибридных систем
- ② LightFM



Типология гибридных систем



1

Стекинг (обогащение фич)

Фичи по содержанию:

- Фильм про зомби
- Снят в 90-е

Предсказания других алгоритмов:

- Вероятность понравиться по ALS — 0.8
- Предсказанная оценка по kNN — 4.7

Блендинг (взвешивание)

$$0,3 * \begin{array}{|c|c|} \hline \text{алгоритм A} \\ \hline \text{item1} & 5.0 \\ \hline \text{item2} & 0.0 \\ \hline \text{item3} & 3.0 \\ \hline \end{array} + 0,7 * \begin{array}{|c|c|} \hline \text{алгоритм B} \\ \hline \text{item1} & 3.0 \\ \hline \text{item2} & 5.0 \\ \hline \text{item3} & 4.0 \\ \hline \end{array} = \begin{array}{|c|c|} \hline \text{гибрид} \\ \hline \text{item1} & 3.6 \\ \hline \text{item2} & 3.5 \\ \hline \text{item3} & 3.7 \\ \hline \end{array}$$

Комбинирование фич

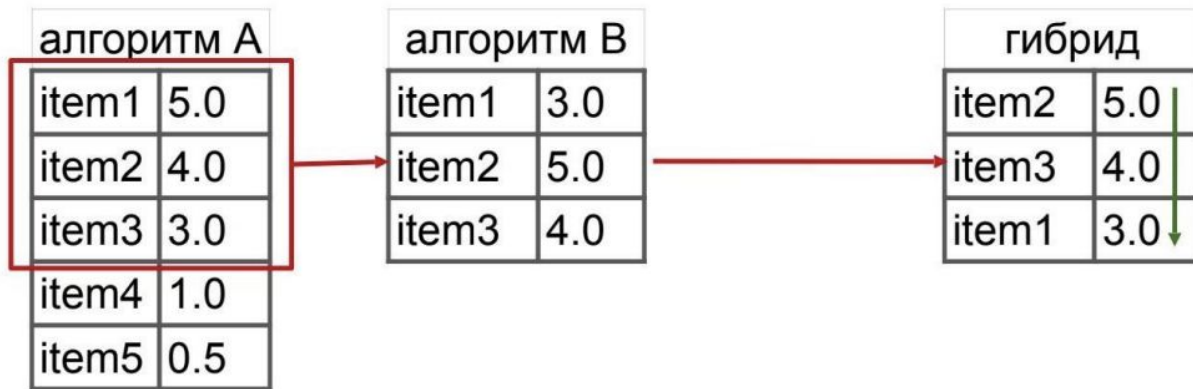
Фичи по содержанию:

- Фильм про зомби
- Снят в 90-е

Коллаборативные фичи:

- А ещё нравится пользователю А
- Но не нравится пользователю Б

Каскадные рекомендации



Резюме

- Гибридизация часто улучшает качество рекомендаций
- Иногда положительно сказывается на разнообразии
- Не гарантирует решение всех проблем, связанных с тем или иным подходом



Практика



Гибридная рекомендательная система

Задача: рекомендации на главной странице сервиса в разделе «Персональная подборка»

Что делать

1. Датасет ml-latest
2. Вспомнить подходы, которые мы разобрали
3. Выбрать понравившийся подход к гибридным системам

LightFM



2

Действующие лица

$u \in U$ — пользователи

$i \in I$ — объекты

$(u, i) \in U \times I = S_+ \sqcup S_-$ — положительные и отрицательные взаимодействия

Действующие лица

- Обучаются векторы для каждой фичи
- Пользователь - сумма векторов его фичей

F^U – всевозможные свойства пользователей

$f_u \subset F^U$ – свойства пользователя u , в том числе id

e_f^U – вектор скрытых факторов фичи f

Действующие лица

- Обучаются векторы для каждой фичи
- Объект - сумма векторов его фичей

F^I — всевозможные свойства объектов

$f_i \subset F^I$ — свойства объекта i , в том числе id

e_f^I — вектор скрытых факторов фичи f

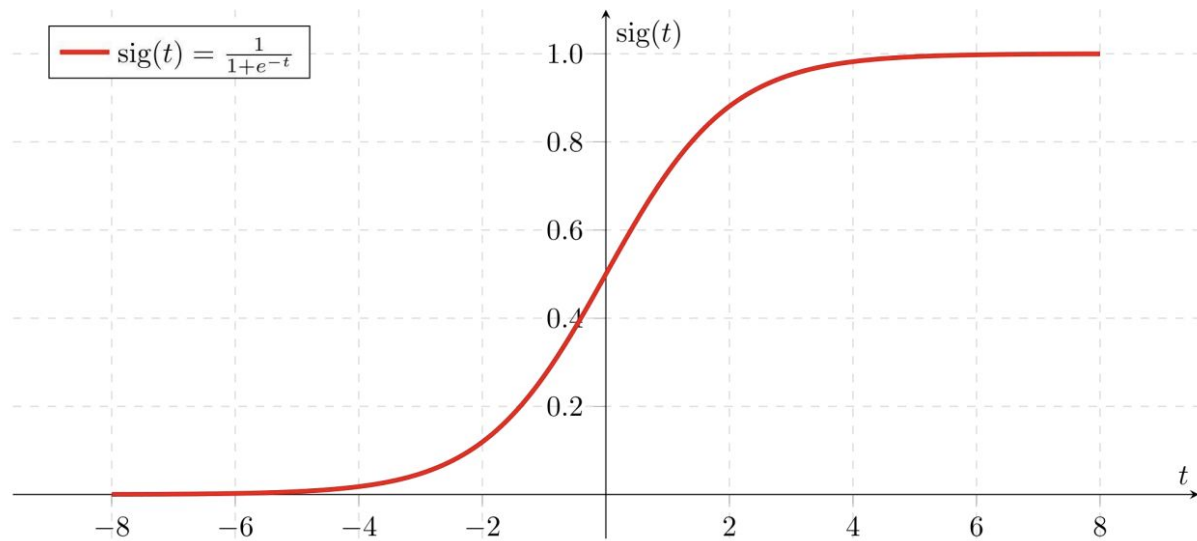
Предсказанная релевантность

$$\hat{r}(u, i) = \sigma \left(\left(\sum_{f \in f_u} \mathbf{e}_f^U \right) \cdot \left(\sum_{f \in f_i} \mathbf{e}_f^I \right) + \sum_{f \in f_i} b_f^U + \sum_{f \in f_i} b_f^I \right)$$



скалярное произведение

Сигмоида



Функция правдоподобия

$$L(e^U, e^I, b^U, b^I) = \prod_{(u,i) \in S^+} \hat{r}_{ui} \times \prod_{(u,i) \in S^-} (1 - \hat{r}_{ui})$$

максимум находится с помощью SGD

Логарифм правдоподобия

$$\begin{aligned}\ln L(e^U, e^I, b^U, b^I) &= \ln \prod_{(u,i) \in S^+} \hat{r}_{ui} \prod_{(u,i) \in S^-} (1 - \hat{r}_{ui}) \\ &= \sum_{(u,i) \in S^+} \ln \hat{r}_{ui} + \sum_{(u,i) \in S^-} \ln (1 - \hat{r}_{ui}) \\ &= \sum_{(u,i) \in S} p_{ui} \ln \hat{r}_{ui} + (1 - p_{ui}) \ln (1 - \hat{r}_{ui})\end{aligned}$$

Негативный семплинг

- Все наблюдаемые взаимодействия — положительные
- Часть наблюдаемых можно считать отрицательными
- Отбирать, какие именно, можно случайно

Практика



LightFM

Задача: рекомендации на главной странице сервиса в разделе «Персональная подборка»

Что делать

1. Датасет тот же ml-latest
2. Использовать LightFM
3. Взять любого пользователя, посмотреть на результаты предсказаний

Сколько есть времени

20 минут

Гибридные рекомендательные системы

