

Рекомендательные системы

Практикум на ЭВМ, весна 2018

Попов Артём Сергеевич

МГУ имени М. В. Ломоносова, факультет ВМК, кафедра ММП

29 марта 2018 г.

Введение

Корреляционные методы

Латентные модели

Classification-based

Практические аспекты

Рекомендации фильмов на сайте Кинопоиск



...его собирает свои истории его режиссер Мартин МакДона.

21



Интервью

Сэм Рокуэлл: «Может, я стану следующей Суперженщиной?»

Звезда «Трех билбордов» обсудила с КиноПоиском пользу Эдипова комплекса и алкоголя и рассказала, как была снята сцена с выбрасыванием человека из окна.

61



Если вам понравился этот фильм, не пропустите

☰ скрыть оцененные фильмы



[Фарго](#)
Fargo



[Таинственная река](#)
Mystic River



[Гран Торино](#)
Gran Torino



[Законопослушный... гражданин](#)
Law Abiding Citizen



[Залечь на дно в Брюгге](#)
In Bruges

● Знаете похожие фильмы? [Порекомендуйте их...](#)

☞ все рекомендации к фильму (7)

4. [Излом времени](#) 50 337 624
A Wrinkle In Time

5. [Чёрная Пантера](#) 39 080 670
Black Panther

16.03 — 18.03

[подробнее »](#)



Результаты уик-энда

Зрители

2 452 899

— 2 395 440

▼ 49%

Деньги

659 586 622 руб.

— 636 541 912

▼ 49%

Цена билета

268,90 руб.

▲ 1,57

▲ 1%

16.03 — 18.03

[подробнее »](#)



Лучшие фильмы — Top 250

43. [Пробуждение](#)
Awakenings

8.446

Рекомендации фильмов на сайте Кинопоиск



...его собирает свои истории его режиссер Мартин МакДона.

21



Интервью

Сэм Рокуэлл: «Может, я стану следующей Суперженщиной?»

Звезда «Трех билбордов» обсудила с КиноПоиском пользу Эдипова комплекса и алкоголя и рассказала, как была снята сцена с выбрасыванием человека из окна.

61

Если вам понравился этот фильм, не пропустите

скрыть оцененные фильмы



[Фарго](#)
Fargo



[Таинственная река](#)
Mystic River



[Гран Торино](#)
Gran Torino



[Законопослушный гражданин](#)
Law Abiding Citizen



[Залечь на дно в Брюгге](#)
In Bruges

Знаете похожие фильмы? Рекомендуем их...

все рекомендации к фильму (7)

4. [Излом времени](#) 50 337 624
A Wrinkle In Time

5. [Чёрная Пантера](#) 39 080 670
Black Panther

16.03 — 18.03

[подробнее »](#)



Результаты уик-энда

Зрители

2 452 899

2 395 440

▼ 49%

Деньги

659 586 622 руб.

636 541 912

▼ 49%

Цена билета

268,90 руб.

1,57

▲ 1%

16.03 — 18.03

[подробнее »](#)



Лучшие фильмы — Top 250

43. [Пробуждение](#)
Awakenings

8.446

Рекомендации от сервиса Яндекс.Маркет





Яндекс.Маркет

Электроника

Мобильные телефоны

Смартфоны

Описание





ЕЩЕ 8

✓ Товар добавлен в корзину

Samsung Galaxy S9 64Gb SM-G960FD Ультрафиолет...
Цвет товара: ультрафиолет
HEY APPLE

- 1 +

48 700 P


В корзине 1 товар на сумму 48 700 P


Продолжить покупки


Оформить заказ

В магазине «HEY APPLE» часто покупают


Дополните ваш заказ и сэкономьте на доставке


Apple Magic Mouse 2 White Bluetooth
Мыши
4 400 P
В корзину


Наушники Marshall Major II Bluetooth
Наушники и Bluetooth-гарнитуры
7 990 P
В корзину


Беспроводная акустика JBL Charge 3 Black (JBLCHARGE3BLKEU)
Портативная акустика
7 990 P
Цвет: ●
В корзину

Рекомендации от сервиса Яндекс.Маркет

✓ Товар добавлен в  корзину

Samsung Galaxy S9 64Gb SM-G960FD Ультрафи...

Цвет товара: ультрафиолет

HEY APPLE

- 1 +

48 700 ₽


В корзине 1 товар на сумму 48 700 ₽

Продолжить покупки

Оформить заказ

В магазине «HEY APPLE» часто покупают

Дополните ваш заказ и сэкономьте на доставке




Apple Magic Mouse 2 White Bluetooth

Мыши

4 400 ₽

В корзину




Наушники Marshall Major II Bluetooth

Наушники и Bluetooth-гарнитуры

7 990 ₽

В корзину



Беспроводная акустика JBL Charge 3 Black (JBLCHARGE3BLKEU)

Портативная акустика









7 990 ₽

Цвет: ●


В корзину

Рекомендации песен от сервиса Яндекс.Музыка

Нравится [альтернатива](#)? А это слышали?

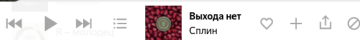
	All My Life Foo Fighters	4:22
	One Step Closer Linkin Park	2:37
	The Only Exception Paramore	4:27
	Polly Nirvana	2:53
	Unintended Muse	3:57
	Cut The Cord Shinedown, Brent Smith	3:44
	Miss Nothing The Pretty Reckless	3:13
	Love Hurts Incubus	3:57

[Рок](#): лучшее за неделю



Кракатук

4:38



Слушайте музыку в высоком качестве и без рекламы

Оформить подписку

Формальная постановка задачи

Дано:

- ▶ U — множество субъектов (users/пользователи)
- ▶ I — множество объектов (items/товары/ресурсы)
- ▶ Y — множество возможных действий
- ▶ T — множество транзакций

$$T = \{(u_j, i_j, y_j) \mid u \in U, i \in I, y \in Y\}_{j=1}^N$$

Пример: сайт с музыкой

- ▶ U — пользователи сайта
- ▶ I — песни на сайте
- ▶ Y_1 — прослушать песню на 70%, $Y_1 = \{e\}$
 Y_2 — поставить оценку песне, $Y_2 = \{0, 1, 2, 3, 4, 5\}$
 $Y = Y_1 \cup Y_2$

Формальная постановка задачи

- ▶ Пусть $Y = \{\mathbb{Z}_+ \cup \{0\}\}$ (для простоты изложения)
- ▶ Работать со списком транзакций неудобно, заведём матрицу пользователи-айтемы (bag of items)
 $X \in \mathbb{R}^{|U| \times |I|}$:

$$X_{ui} = \sum_{j=1}^N [u_j = u][i_j = i] y_j, \text{ если } (u, i) \text{ встречалась в } X$$

Не все ячейки X заполнены, но не значит, что они нулевые!

Основные задачи:

- ▶ Предсказать незаполненные ячейки X
- ▶ Посчитать близости $\rho(u, u')$, $\rho(i, i')$, $\rho(u, i)$
- ▶ Сформировать рекомендации для всех u (по всем i)

Пример. Покупки в интернет-магазине

- ▶ U — множество интернет-пользователей
- ▶ I — множество товаров в магазине
- ▶ Y — пользователь купил товар
- ▶ $X_{ui} = \mathbb{I}[u \text{ купил товар } i]$

Задачи, которые можно решать:

- ▶ рекомендовать клиенту другие товары
- ▶ рекомендовать клиенту товары во время его следующей покупки
- ▶ информировать клиента о наличии товара

Пример. Конкурс Netflixprize

Конкурс (ссылка) проходил в с 2006 по 2009 год

Призовой фонд: 1 000 000 долларов

- ▶ U — множество пользователей сервиса
- ▶ I — множество фильмов
- ▶ Y — оценка фильма
- ▶ $X_{ui} = \mathbb{I}[u \text{ рейтинг, выставленный } u \text{ для } i]$

Метрика качества: MSE

Что необычного:

- ▶ Один из первых конкурсов с большим призовым фондом
- ▶ Один из первых больших датасетов для рекомендаций
- ▶ Многие методы появились во время решения конкурса
- ▶ Много методов рекомендаций для оптимизации MSE

Тривиальные рекомендации

Пусть $Y = \{1\}$ (1, если купил)

Идея: клиенты, купившие i_0 , также купят $I(i_0)$

1. Пусть пользователь u_0 купил товар i_0
2. Множество пользователей, покупавших товар i_0

$$U(i_0) = \{u \in U | x_{ui_0} \neq \emptyset, u \neq u_0\}$$

3. Множество товаров, близких данному товару

$$I(i_0) = \{i \in I | \text{sim}(i, i_0) > \delta\}$$

$$\text{sim}(i, i_0) = \frac{|U(i_0) \cap U(i)|}{|U(i_0) \cup U(i)|}$$

4. Взять наибольшие по $\text{sim}(i, i_0)$ элементы из $I(i_0)$

Пример данных

Таблица: Матрица X — покупки пользователей

	телефон	наушники	power-bank	sd-карта	тостер	блендер
Вова	+					+
Дима	+	+	+	+		
Женя	+	+			+	+
Юра	+	+				
Маша		+	+		+	+
Рома		+				
Лёша		+			+	

Что порекомендуется Вове во время покупки телефона?

Пример данных

Таблица: Матрица X — покупки пользователей

	телефон	наушники	power-bank	sd-карта	тостер	блендер
Вова	+					+
Дима	+	+	+	+		
Женя	+	+			+	+
Юра	+	+				
Маша		+	+		+	+
Рома		+				
Лёша		+			+	

Что порекомендуется Вове во время покупки телефона?
Наушники.

Пример данных

Таблица: Матрица X — покупки пользователей

	телефон	наушники	power-bank	sd-карта	тостер	блендер
Вова	+					+
Дима	+	+	+	+		
Женя	+	+			+	+
Юра	+	+				
Маша		+	+		+	+
Рома		+			+	
Лёша		+			+	

Что порекомендуется Роме?

Пример данных

Таблица: Матрица X — покупки пользователей

	телефон	наушники	power-bank	sd-карта	тостер	блендер
Вова	+					+
Дима	+	+	+	+		
Женя	+	+			+	+
Юра	+	+				
Маша		+	+		+	+
Рома		+			+	
Лёша		+			+	

Что порекомендуется Роме?
Блендер.

Проблемы подхода

- ▶ Рекомендации тривиальные (всё самое популярное)
- ▶ Не учитываются интересы пользователя u_0
- ▶ Проблема холодного старта (нечего рекомендовать новым пользователям)
- ▶ Хранение матрицы X

User-based рекомендации

Идея: клиенты, похожие на u_0 , также купили купят $I(u_0)$

1. Множество пользователей, похожих на u_0

$$U(i_0) = \{u \in U \mid \underset{user}{\text{sim}}(u, u_0) > \delta_1, u \neq u_0\}$$

2. Множество пользователей, купивших товар i

$$V(i) = \{u \in U \mid x_{ui} \neq \emptyset\}$$

3. Множество товаров, близких данному пользователю

$$I(u_0) = \{i \in I \mid \text{sim}(u_0, i) > \delta_2\}$$

$$\text{sim}(u_0, i) = \frac{|U(u_0) \cap V(i)|}{|U(u_0) \cup V(i)|}$$

4. Взять наибольшие по $\text{sim}(u_0, i)$ элементы из $I(u_0)$

Проблемы подхода

- ▶ Нет рекомендаций для нетипичных пользователей
- ▶ Проблема холодного старта
- ▶ Хранение матрицы X

Item-based подход

Идея: с товарами, купленными u_0 , часто покупают $I(u_0)$

1. Множество товаров, близких хоть какому-то из товаров u_0

$$I(u_0) = \{i \in I \mid \exists i_0 : x_{u_0 i_0} \neq \emptyset, \underset{item}{\text{sim}}(i, i_0) > \delta\}$$

2. Взять наибольшие по $\text{sim}(i, i_0)$ элементы из $I(u_0)$

Недостатки:

- ▶ Снова тривиальность
- ▶ Проблема холодного старта
- ▶ Хранение матрицы X

User-based KNN

Пусть $Y = \{1, 2, 3, \dots, K\}$ (рейтинги)

$$\hat{x}_{ui} = \bar{x}_u + \frac{\sum_{u' \in U_\alpha} \text{sim}(u, u')(x_{u'i} - \bar{x}_{u'})}{\sum_{u' \in U_\alpha} \text{sim}(u, u')}$$

\hat{x}_{ui} — предсказания рейтинга

$\bar{x}_u = \frac{1}{|I(u)|} \sum_{i \in I(u)} x_{ui}$ — средние рейтинги пользователя

$U_\alpha(u) = \{u' | \text{sim}(u, u') > \alpha\}$ — близкие пользователи

$I(u)$ — множество оценённых товаров

Item-based KNN

$$\hat{x}_{ui} = \bar{x}_i + \frac{\sum_{i' \in I_\alpha} \text{sim}(i, i')(x_{ui'} - \bar{x}_{i'})}{\sum_{i' \in I_\alpha} \text{sim}(i, i')}$$

\hat{x}_{ui} — предсказания рейтинга

$\bar{x}_i = \frac{1}{|U(i)|} \sum_{u \in U(i)} x_{ui}$ — средние рейтинги товара

$I_\alpha = \{i' | \text{sim}(i, i') > \alpha\}$ — близкие товары

$U(i)$ — множество пользователей, оценивших товар

Параметры метода

Функции близости:

- ▶ Корреляция Пирсона
- ▶ Косинусная мера близости
- ▶ Мера Жаккарда

Почему KNN?

Параметры метода

Функции близости:

- ▶ Корреляция Пирсона
- ▶ Косинусная мера близости
- ▶ Мера Жаккарда

Почему KNN?

$$\sum_{i=1}^N w_i(x)(\alpha - y_i)^2 \rightarrow \min_{\alpha}$$

$$a(x) = \frac{\sum_{i=1}^N w_i(x)y_i}{\sum_{i=1}^N w_i(x)}$$

Итоги

Корреляционные методы:

- ▶ Интуитивные и понятные
 - ▶ Легко реализовать для небольших множеств U и I
 - ▶ Нет никаких теоретических обоснований
 - ▶ Не ставится никакой задачи оптимизации, работа метода зависит только от понимания задачи
 - ▶ Проблема холодного старта
 - ▶ Проблема работы с большой матрицей
- Необходимы специальные модели для работы с матрицей, например, map-reduce**

Латентные модели

Идея: для $u \in U$ построить вектор $p_u \in \mathbb{R}^g$, $g \ll |U|$
для $i \in I$ построить вектор $q_i \in \mathbb{R}^h$, $h \ll |I|$
 $\hat{X}_{ui} = F(p_u, q_i)$

Способы построения моделей:

- ▶ Жёсткая кластеризация ($p_{uc} = \mathbb{I}[u \text{ в кластере } c]$)
- ▶ Мягкая кластеризации (тематические модели)
(p_{uc} — оценка принадлежности u кластеру c)
- ▶ Матричные разложения
(p_u — столбцы/строки каких-то матриц после разложения)
- ▶ Специальные методы обучения представлений
(p_u — вектор из модели skip-gram)
- ▶ End-to-end построение представлений

Матричные разложения: SVD

Хотим найти разложение матрицы X :

$$X = PQ^T \quad \text{или} \quad X = P\Sigma Q^T$$

p_u — строки матрицы P , q_i — строки матрицы Q

Использование сингулярного разложения (SVD):

$$\|X - P\Sigma Q^T\|^2 \rightarrow \min_{P, Q, \Sigma}$$

$$PP^T = I \quad QQ^T = I$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_d), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_d \geq 0$$

Можно записать так:

$$\sum_{u \in U} \sum_{i \in I} (x_{ui} - p_u^T \Sigma q_i)^2 \rightarrow \min_{P, Q, \Sigma}$$

SVD: недостатки и преимущества

- Если X_{ui} неизвестно, мы будем считать его нулём
- Все вектора одной сущности ортогональны между собой, сложно искать похожие
- Неинтерпретируемые
- + Можно использовать Truncated SVD для уменьшения размерности
- + Много готовых реализаций

В качестве представлений можно также использовать:

$$p_u = P_u \Sigma \quad q_i = Q_i$$

$$p_u = P_u \sqrt{\Sigma} \quad q_i = \sqrt{\Sigma} Q_i$$

LFM, Latent Fator Model

Не будем учитывать неизвестные элементы как нулевые¹:

$$\sum_{(u,i) \in T} (x_{ui} - p_u^T q_i)^2 \rightarrow \min_{P,Q}$$

Оптимизация модели с помощью метода SGD

Можно учитывать регуляризацию:

$$\sum_{(u,i) \in T} \left((x_{ui} - p_u^T q_i)^2 + \lambda \|p_u\|^2 + \mu \|q_i\|^2 \right) \rightarrow \min_{P,Q}$$

Можно учитывать средний вклад пользователя и товара:

$$\sum_{(u,i) \in T} (x_{ui} - \hat{x}_u - \hat{x}_i - p_u^T q_i)^2 \rightarrow \min_{P,Q}$$

¹Tacaks G., Pilaszy I., Nemeth B., Tikk D. Scalable collaborative filtering approaches for large recommendation systems

LFM, Latent Fator Model

Можно делать неотрицательные компоненты:

$$\sum_{(u,i) \in T} (x_{ui} - p_u^T q_i)^2 \rightarrow \min_{p \geq 0, q \geq 0}$$

Обучение с помощью метода проекции градиента

Можно использовать функцию β (пример: σ , если $x_{ui} \in [0, 1]$):

$$\sum_{(u,i) \in T} (x_{ui} - \beta(p_u^T q_i))^2 \rightarrow \min_{P, Q}$$

Можно использовать вместо квадратичной ошибки любую другую, например hinge loss

ALS для LFM's

Можно использовать метод ALS для обучения

Идея: в точке оптимума L должно выполняться:

$$\frac{\partial L}{\partial p_u} = 0, \quad \frac{\partial L}{\partial q_i} = 0$$

Зафиксируем переменные Q :

$$\sum_{u \in U} \left(\|X_u - Qp_u\|^2 + \frac{\lambda}{2} \|p_u\|^2 \right) \rightarrow \min_P$$

ALS для LFMс

Можно использовать метод ALS для обучения

Идея: в точке оптимума L должно выполняться:

$$\frac{\partial L}{\partial p_u} = 0, \quad \frac{\partial L}{\partial q_i} = 0$$

Зафиксируем переменные Q :

$$\sum_{u \in U} \left(\|X_u - Qp_u\|^2 + \frac{\lambda}{2} \|p_u\|^2 \right) \rightarrow \min_P$$

Задача минимизации решается аналитически:

$$p_u = (Q^T Q + \lambda I)^{-1} Q^T X_u$$

Аналогично, можно решить задачу, зафиксировав P :

$$q_u = (P^T P + \mu I)^{-1} P^T X_i$$

ALS для LFM's

Будем решать итерационно:

- ▶ Зафиксировав Q , пересчитываем P
- ▶ Зафиксировав P , пересчитываем Q

Используем разложение Холецкого вместо обращения

- ▶ Хорошо и быстро работает
- ▶ Можно обобщить на случай неотрицательных разложения
(положительная срезка $x \rightarrow \max(x, 0)$)
- ▶ Легко обновлять профили пользователей после прихода новых оценок

Интерпретация ALS

Распишем формулу принятия решения:

$$\hat{X}_{ui} = q_i^T p_u = q_i^T (Q^T Q + \lambda I)^{-1} Q^T X_u = \sum_{j \in I} q_i^T W q_j X_{uj}$$

$$W = (Q^T Q + \lambda I)^{-1} = L L^T$$

$$\hat{X}_{ui} = \sum_{j \in I} q_i^T L^T L q_j X_{uj} = \sum_{j \in I} (L q_i)^T L q_j X_{uj}$$

На что похоже?

Интерпретация ALS

Распишем формулу принятия решения:

$$\hat{X}_{ui} = q_i^T p_u = q_i^T (Q^T Q + \lambda I)^{-1} Q^T X_u = \sum_{j \in I} q_i^T W q_j X_{uj}$$

$$W = (Q^T Q + \lambda I)^{-1} = L L^T$$

$$\hat{X}_{ui} = \sum_{j \in I} q_i^T L^T L q_j X_{uj} = \sum_{j \in I} (L q_i)^T L q_j X_{uj}$$

На что похоже?

Корреляционные методы:

$$\hat{x}_{ui} = \bar{x}_i + \frac{\sum_{i' \in I_\alpha} \text{sim}(i, i') (x_{ui'} - \bar{x}_{i'})}{\sum_{i' \in I_\alpha} \text{sim}(i, i')}$$

Неявные и явные предпочтения

Явные (explicit):

- ▶ Проставил рейтинг фильму
- ▶ Лайкнул запись
- ▶ Написал рецензию на товар

Пользователь явно сообщает своё отношение к объекту

Неявные (implicit):

- ▶ Просмотрел страницу фильма
- ▶ Посетил страницу пользователя
- ▶ Купил товар в интернет-магазине

Если есть доступ к неявным предпочтениям, как их учитывать?

Можно ли строить латентные модели по бинарным данным?

Implicit ALS

Пусть x_{ui} — неявный фидбек

Пусть s_{ui} — показатель неявного интереса

$$s_{ui} = \begin{cases} 1, & x_{ui} \geq 0 \\ 0, & x_{ui} = 0 \end{cases}$$

Пусть c_{ui} — уровень доверия показателю s_{ui}

$$c_{ui} = 1 + \alpha x_{ui}$$

Модель Implicit ALS (оптимизация с помощью ALS):

$$\sum_{(u,i) \in T} c_{ui} (s_{ui} - p_u^T q_i)^2 \rightarrow \min_{P, Q}$$

Модели cbow и skip-gram

В модель cbow по словам контекста предсказывается слово:

$$\mathcal{L}(U, V) = \sum_{i=1}^N \log p(w_i | w_{i-k}^{i+k}) \rightarrow \max_{U, V}$$

$$p(w_i | w_{i-k}^{i+k}) = \operatorname{softmax}_{w_i \in W} \left\langle v_{w_i}, \sum_{j=-k, j \neq 0}^k u_{w_{i+j}} \right\rangle$$

В модели skip-gram по слову предсказывается его контекст:

$$\mathcal{L}(U, V) = \sum_{i=1}^N \sum_{j=-k, j \neq 0}^k \log p(w_{i+j} | w_i) \rightarrow \max_{V, U}$$

$$p(c | w) = \operatorname{softmax}_{c \in W} \langle v_c, u_w \rangle = \frac{\exp(\langle v_c, u_w \rangle)}{\sum_{c'} \exp(\langle v_{c'}, u_w \rangle)}$$

Модель paragraph2vec

paragraph2vec (PV-DBOW) — расширение моделей word2vec на представления документов

По словам из контекста и текущему документу предсказываем слово:

$$\mathcal{L}(U, V) = \sum_{d \in D} \sum_{i=1}^{N_d} \log p(w_i | d, w_{i-k}^{i+k}) \rightarrow \max_{U, V}$$

$$p(w_i | d, w_{i-k}^{i+k}) = \operatorname{softmax}_{w_i \in W} \left\langle v_{w_i}, \sum_{\substack{j=-k \\ j \neq 0}}^k u_{w_{i+j}} + u_d \right\rangle$$

Адаптация модели под рекомендации

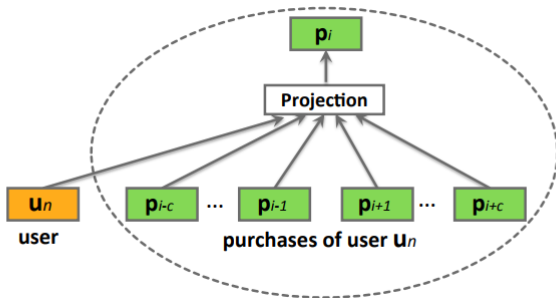
По пользователю предсказываем товары:

$$\mathcal{L}(U, V) = \sum_{(u,i) \in T} \log p(i|u) \rightarrow \max_{V, U}$$

Может хорошо работать в задачах, где у пользователя есть константные предпочтения:

- ▶ музыка
- ▶ фильмы

user2vec



По товарам пользователя предсказываем другие его товары (user2vec):

$I(u)$ — товары пользователя

$$\mathcal{L}(U, V) = \sum_{u \in U} \sum_{i=1}^{I(u)} \log p(i|u, \text{sample} \sim I(u) \setminus i) \rightarrow \max_{U, V}$$

Адаптация модели под рекомендации

Товары, которые покупаются одновременно, похожи (product2vec):

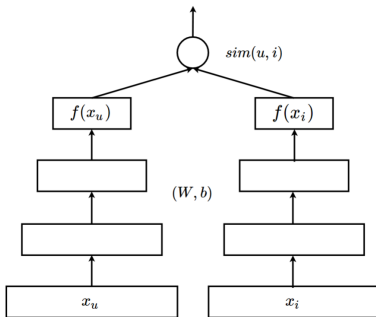
$I(i)$ — товары, которые покупались в связке

$$\mathcal{L}(U, V) = \sum_{i \in I} \sum_{j \in I(i)} \log p(j|i) \rightarrow \max_{V, U}$$

Может хорошо работать в задачах, где нет константных предпочтений:

- покупки в интернет-магазине

Deep semantic similarity based personalized recommendation

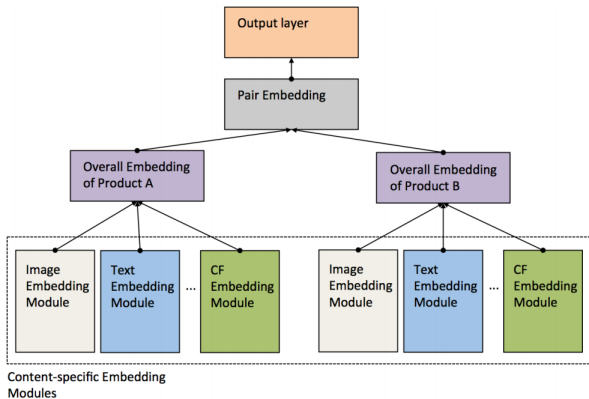


Учимся предсказывать близость пользователя и товара²

Обучаем с negative sampling

²Deep Learning based Recommender System: A Survey and New Perspectives

Комбинация разных факторов: content2vec



Обучаем с negative sampling

Задача рекомендаций, как задача классификации/регрессии

Пусть $y \in \{1, 2, \dots, K\}$ (рейтинги для фильмов)

Признаковые описания:

- ▶ user: пол, возраст, интересы, one-hot вектор user
- ▶ item: жанр фильма, описание, one-hot вектор item

Обучающая выборка: все пары (u, i) , для которых известен y

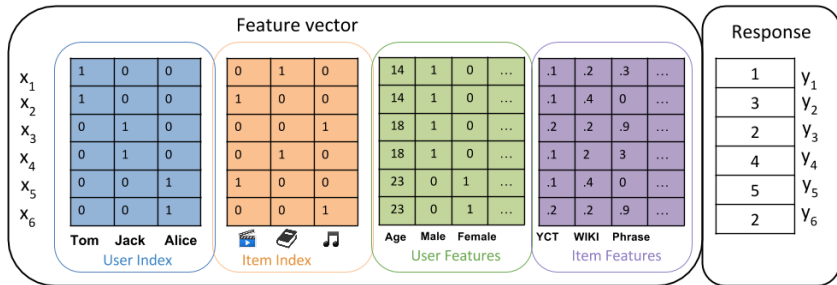
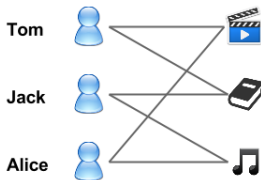
Обучение: обучаем любой алгоритм классификации/регрессии

Выдача рекомендаций: для каждого user выдаём items с наибольшим предсказанным y

Сложности подхода

- ▶ Как учитывать взаимодействие пользователей и товаров?
- ▶ Как учитывать негативные примеры? (пользователь не покупает товар)
- ▶ Как отбирать кандидатов для вычисления y ?

Множества признаков



Множества признаков

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					Time	Last Movie rated						

В качестве признаков добавлена история пользователя

Проблема отсутствия взаимодействия

- ▶ Не учитываются взаимодействия пользователя и товара

Проблема отсутствия взаимодействия

- ▶ Не учитываются взаимодействия пользователя и товара

У нас есть информация о взаимодействиях x_{ui} !

Количество таких признаков $|U| + |I|$ для пары (u, i)

Проблема отсутствия взаимодействия

- ▶ Не учитываются взаимодействия пользователя и товара

У нас есть информация о взаимодействиях x_{ui} !

Количество таких признаков $|U| + |I|$ для пары (u, i)

Пусть наша модель изначально была линейной:

$$\hat{x}_{ui} = \sum_f w_f x_f + \sum_{u' \in U} w_{u'} [u = u'] + \sum_{i' \in I} w_{i'} [i = i']$$

Добавим в качестве признака историю пользователей:

$$\begin{aligned} \hat{x}_{ui} = \sum_f w_f x_f + \sum_{u' \in U} w_{u'} [u = u'] + \sum_{i' \in I} w_{i'} [i = i'] + \\ + \sum_{i'} w_{ui'} x_{ui'} + \sum_{u'} w_{u'i} x_{u'i} \end{aligned}$$

Квадратичная модель

Пойдём дальше: хотим добавить признак индикатор пары
 $[user = u, item = i] = [user = u][item = i] = x_u x_i$

Таких признаков $|U| \times |I|$ (больше чем объектов!) — легко переобучиться

По сути, теперь модель не линейная, а квадратичная:

$$\hat{x}_{user,item} = w_0 + \sum_{u' \in U} w_{u'} x_{u'} + \sum_{i' \in I} w_{i'} x_{i'} + \sum_{u' \in U} \sum_{i' \in I} w_{u'i'} x_{u'} x_{i'}$$

Факторизационные машины

Пусть $w_{ui} = \langle v_u, v_i \rangle$, где $v_u, v_i \in \mathbb{R}^m$

Модель «Factorization machine»^{3 4} (FM):

$$\hat{x}_{user,item} = w_0 + \sum_{u' \in U} w_{u'} x_{u'} + \sum_{i' \in I} w_{i'} x_{i'} + \sum_{u' \in U} \sum_{i' \in I} w_{u'i'} x_{u'} x_{i'}$$

Обучение модели с помощью SGD (или ALS или MCMC)

³<https://www.csie.ntu.edu.tw/~b97053/paper/Rendle2010FM.pdf>

⁴https://mk-minchul.github.io/Factorization_Machine/

Проблема отсутствия отрицательных примеров

Пусть $y \in \{1\}$ (покупка товара)

Предложенный метод в лоб не работает, так как нет объектов отрицательного класса

Что делать?

Проблема отсутствия отрицательных примеров

Пусть $y \in \{1\}$ (покупка товара)

Предложенный метод в лоб не работает, так как нет объектов отрицательного класса

Что делать?

Сэмплировать негативные примеры

Генерация негативных примеров

- ▶ Все, которых нет в выборке (невозможно)
- ▶ Случайные из равномерного распределения
- ▶ Случайные с вероятностями, пропорциональными популярности объекта
- ▶ Объекты, которые рекомендует какой-то алгоритм, но они не были куплены
- ▶ Комбинация стратегий

Проблема выбора списка рекомендаций

Невозможно получить оценки сразу для всех товаров

Давайте проведём отбор кандидатов:

- ▶ Только популярные
- ▶ Только находящиеся в той же категории, что и текущий
- ▶ Только те, которые уже покупал пользователь
- ▶ Которые близки (sim) к текущему
- ▶ Заранее подготовленные списки
- ▶ Которые считаются вероятными у других подходов к рекомендациям

О подходе

- ▶ Очень хорошее качество
- ▶ Не так часто упоминается в статьях...
- ▶ ... но именно так часто делают на практике
- ▶ Легко ансамблировать разные другие алгоритмы рекомендаций
- ▶ Легко учитывать контент — текст, картинки

Ещё раз о метриках. . .

Метрика RMSE — не всегда хороша, т.к. задача точно предсказать оценку обычно не стоит

- ▶ подобрать рекомендации для пользователя
- ▶ отранжировать их по релевантности
- ▶ точные оценки не важны, важен порядок

Какие метрики лучше?

L_u — истинные предпочтения u

$R_u(k)$ — лучшие k рекомендаций

$$\text{precision@k} = \frac{|L_u \cap R_u(k)|}{|R_u(k)|} \quad \text{recall@k} = \frac{|L_u \cap R_u(k)|}{|L_u|}$$

$$\text{hitrate@k} = [L_u \cap R_u(k) \neq \emptyset]$$

Другие метрики

- ▶ Разнообразие (diversity): например, число рекомендаций из разных категорий, или степень различия рекомендаций между сессиями пользователей
- ▶ Новизна (novelty): сколько среди рекомендаций объектов, новых для пользователей?
- ▶ Покрытие (coverage): доля объектов, которые хоть раз побывали в числе рекомендованных
- ▶ Прозорливость (serendipity): способность угадывать непопулярные предпочтения

Можно оптимизировать сумму функционалов

Ещё несколько фактов

- ▶ История действий пользователя построена с учётом существующих методов рекомендации
- ▶ Можно смотреть на результаты онлайн-метрик (полученная прибыль, полученное количество кликов)
- ▶ Хотелось бы, чтобы пользователю рекомендовалось то, что он не купил бы без рекомендаций
- ▶ А/Б тестирование для тестирования качества рекомендаций на практике

To be continued...

- ▶ Методы с латентными переменными
- ▶ Матричные разложения
- ▶ Оценивание качества рекомендаций
- ▶ Продвинутое реализации корреляционных методов