

Рекомендательные системы

По материалам Евгения Соколова

Занятие 1

Опрос

- Какие рекомендательные системы вы знаете?

Опрос

- Какие рекомендательные системы вы знаете?
- Рекомендации чего вы хотели бы получать?

Рекомендательные системы

- Фильмы, видео
- Музыка
- Книги
- Приложения
- Товары
- Посты в социальных сетях
- Баннерные системы
- Люди (социальные сети, сервисы знакомств)
- Услуги (рестораны, отели, ...)
- Научные публикации



Рекомендательные системы

- Рекомендательные системы сокращают объём информации, необходимый для принятия решения
- Не нужно читать отзывы на 1000 фильмов — модель сама выберет лучший
- Netflix: 2/3 просмотренных фильмов найдены через рекомендательную систему
- Amazon: 35% продаж через полки рекомендаций
- Youtube: 60% просмотров благодаря рекомендациям

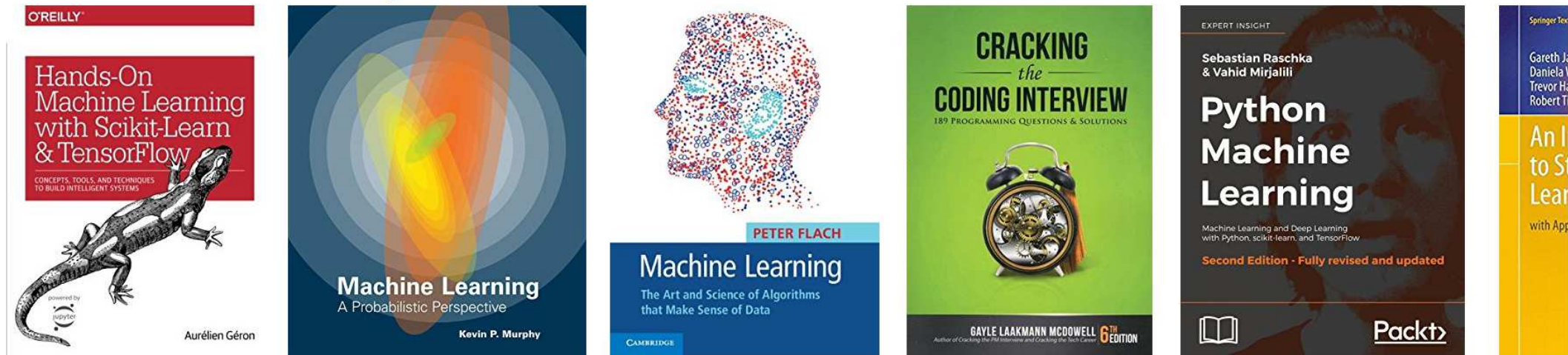
Amazon

Try Amazon Prime today and get unlimited fast, FREE shipping [See more](#)



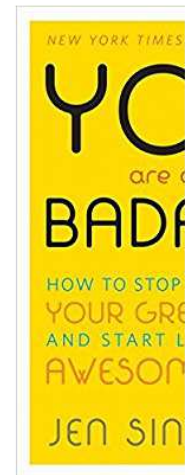
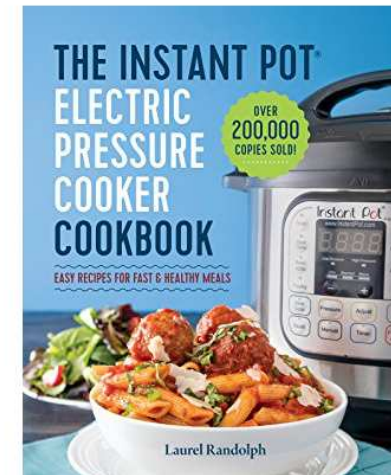
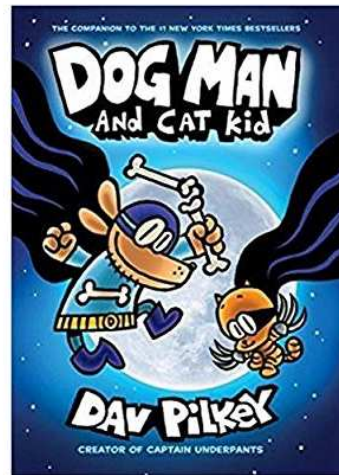
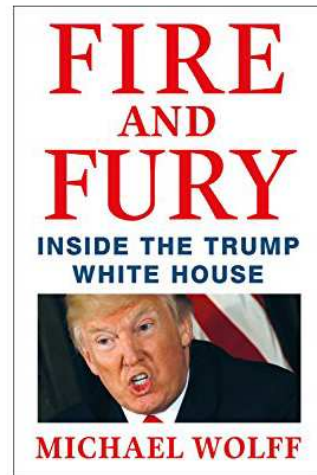
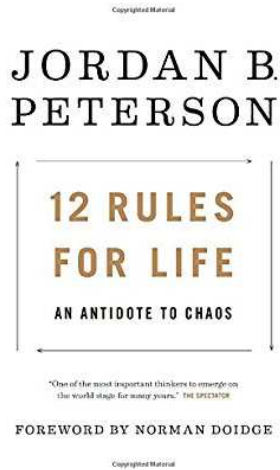
Amazon

Recommendations for you in Books

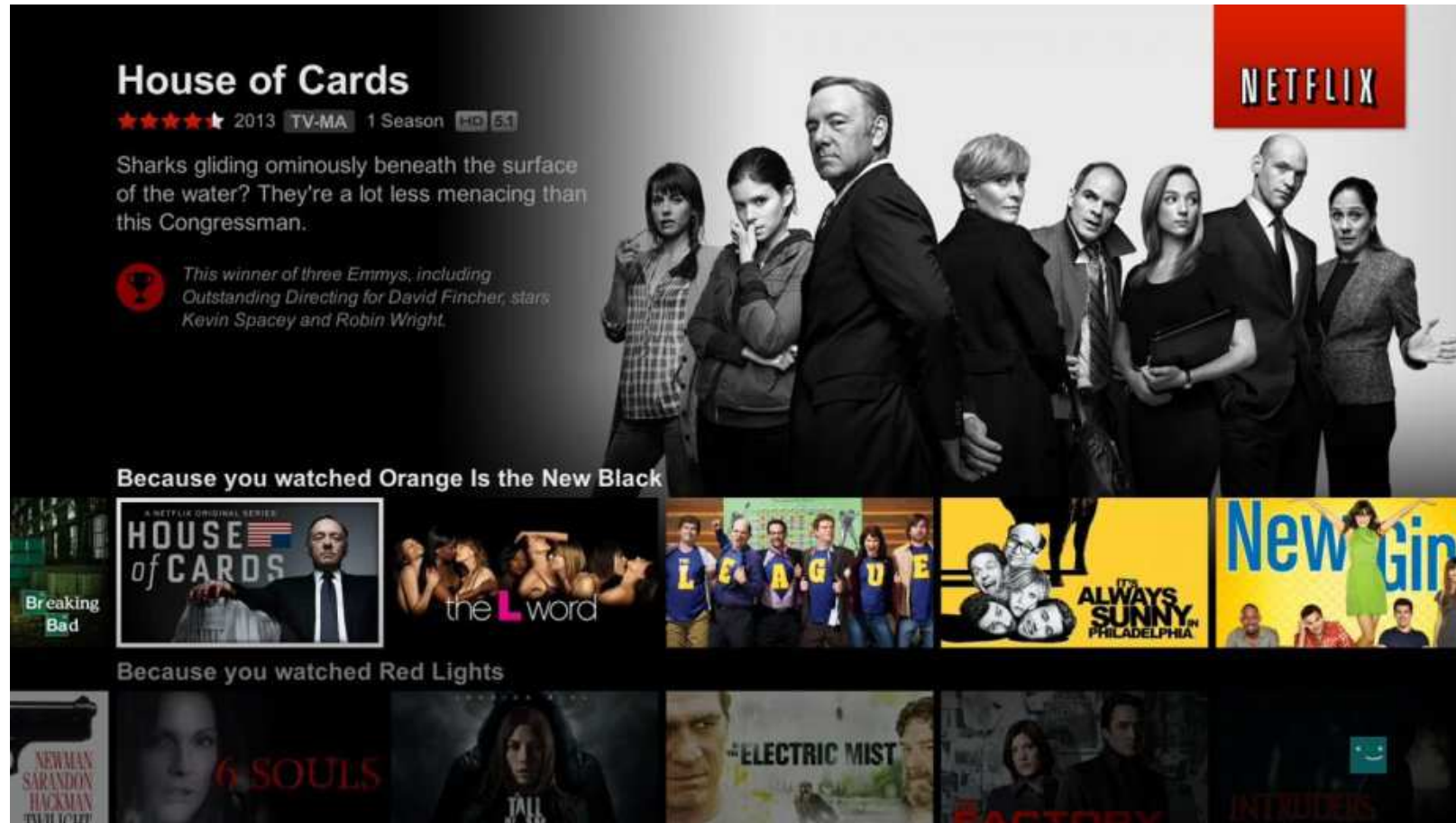


Amazon

Books best sellers [See more](#)




Netflix



House of Cards

★★★★★ 2013 TV-MA 1 Season HD 5.1

Sharks gliding ominously beneath the surface of the water? They're a lot less menacing than this Congressman.

 This winner of three Emmys, including Outstanding Directing for David Fincher, stars Kevin Spacey and Robin Wright.

Because you watched Orange Is the New Black

Breaking Bad | **HOUSE of CARDS** | the L word | LAGUE | ALWAYS SUNNY IN PHILADELPHIA | New Girl

Because you watched Red Lights

NEWMAN SARANDON HACKMAN TWILIGHT | 6 SOULS | TALL MEADOWS | THE ELECTRIC MIST | FACTORY | INTRODUCERS

Netflix

Profile Type	Score Image A	Score Image B
Comedy	5.7	6.3
Romance	7.2	6.5



Image A



Image B

Рекомендации контента

- Медийный бум приводит к взрывному росту объёмов информации в сети
- Рекомендательные системы помогают ориентироваться
- Для авторов — поиск целевой аудитории
- Пионеры в Китае — Toutiao (более 100 миллионов активных пользователей) и другие платформы

Цели с точки зрения продавца

- ?

Цели с точки зрения продавца

- Продать больше товаров
- Продать больше редких товаров
- Повысить лояльность пользователя
- Лучше понять покупателей

Цели с точки зрения покупателя

Цели с точки зрения покупателя

- Купить то, что нужно
- Понять, что покупать вместе с данным товаром
- Понять, что интересно (если нет задачи купить что-то конкретное)

Краткая история

- Начало 90-х: одна из первых рекомендательных систем (GroupLens, рекомендации записей в Usenet)
- Начало 2000-х: активные исследования, коммерциализация
- 2006: Netflix Prize
- 2007: первая конференция RecSys

Netflix Prize

- Предсказываем, какую оценку пользователь поставит фильму
- Метрика: RMSE
- Задача: улучшить на 10% качество предсказания
- Конкурс шёл с 02.10.2006 по 21.09.2009
- Главный приз: \$1,000,000
- Размеры:
 - 500 тысяч пользователей
 - 17 тысяч фильмов
 - 10^8 рейтингов

Netflix Prize

- Одно из первых крупных соревнований по анализу данных (предшественник kaggle и т.д.)
- Первый большой открытый набор данных для тестирования алгоритмов рекомендаций
- Алгоритмы, разработанные участниками конкурса, до сих пор популярны в индустрии
- Netflix Prize привёл к большой популярности RMSE как метрики качества рекомендаций (не самый лучший результат)

Netflix Prize



На основе чего можно строить
рекомендации?

На основе чего можно строить рекомендации?

- Данные по другим пользователям — «что смотрят люди с похожими на мои интересами?»
- Данные по объектам (фильмам) — «какие фильмы похожи на те, которые мне понравились?»

Типичная рекомендательная система

- Объект: пара «user-item»
- Целевая переменная: клики, длинные клики, досмотры, покупки, дослушивания, лайки и т.д.
- Решаем задачу классификации/регрессии/ранжирования

Типичная рекомендательная система

- Объект: пара «user-item»
- Целевая переменная: клики, длинные клики, досмотры, покупки, дослушивания, лайки и т.д.
- Решаем задачу классификации/регрессии/ранжирования

Особенности:

- Выбор целевой переменной
- Выбор метрики качества
- Факторы для модели
- Слишком много товаров/видео/песен/...

Отбор кандидатов

- Простая и быстрая модель, которая отбирает тысячи товаров для данного пользователя
- Сложная модель применяется только к отобранным кандидатам

Основные подходы

- Есть методы, разработанные напрямую для рекомендаций
- Коллаборативная фильтрация
 - Рекомендации на основе сходства действий пользователей
- Контентные рекомендации

Memory-based models

Обозначения

- Множество товаров: I
- Множество пользователей: U
- Множество пар «пользователь-товар», для которых известны оценки: R
- Если для пары (u, i) известен рейтинг, то будем писать $\exists r_{ui}$
- Оценки — рейтинги фильмов, индикаторы покупки товара и т.д.

Оценки

- Оценки (или фидбэк) бывают явные и неявные
- Явные оценки
 - Пользователь поставил оценку фильму/товару
 - Пользователь написал отзыв
 - Пользователь поставил лайк
- Неявные оценки
 - Пользователь посмотрел фильм
 - Пользователь добавил товар в корзину
 - Пользователь долго смотрел на запись в социальной сети
- Неявные оценки более шумные, но их больше

Сходство пользователей

- $I_{uv} = \{i \in I \mid \exists r_{ui} \text{ и } \exists r_{vi}\}$ — множество товаров, которые оценили и пользователь u , и пользователь v
- Сходство пользователей:

$$w_{uv} = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - \bar{r}_v)^2}},$$

где \bar{r}_u и \bar{r}_v — средние рейтинги пользователей

User-based collaborative filtering

- Дан пользователь u_0
- Найдём пользователей, которые похожи на него:

$$U(u_0) = \{v \in U \mid w_{u_0 v} > \alpha\}$$

- Порекомендуем те товары, которые часто покупались пользователями из $U(u_0)$

User-based collaborative filtering

Пользователи		Товары					
		1	1	0		1	
		0	1	1			1
					1	1	0
			1	1		0	
			1				1

User-based collaborative filtering

		Товары					
Пользователи	1	1	1	0		1	
	0	0	1	1			1
					1	1	0
			1	1		0	
		1	1				1

User-based collaborative filtering

		Товары					
Пользователи	1	1	1	0		1	
	2	0	1	1			1
	3				1	1	0
	4		1	1		0	
	5		1				1

Похожие пользователи

User-based collaborative filtering

		Товары					
Пользователи	1	1	0		1		
	0	1	1			1	
				1	1	0	
		1	1		0		
		1				1	

Похожие пользователи

User-based collaborative filtering

Недостатки:

- Много параметров, которые сложно выбирать
 - Какой порог сходства для пользователей?
 - Сколько похожих пользователей должны были купить товар, чтобы мы его порекомендовали?
- Требуется хранить всю матрицу оценок

Есть и другие методы, основанные на сходствах, но все обладают теми же недостатками.