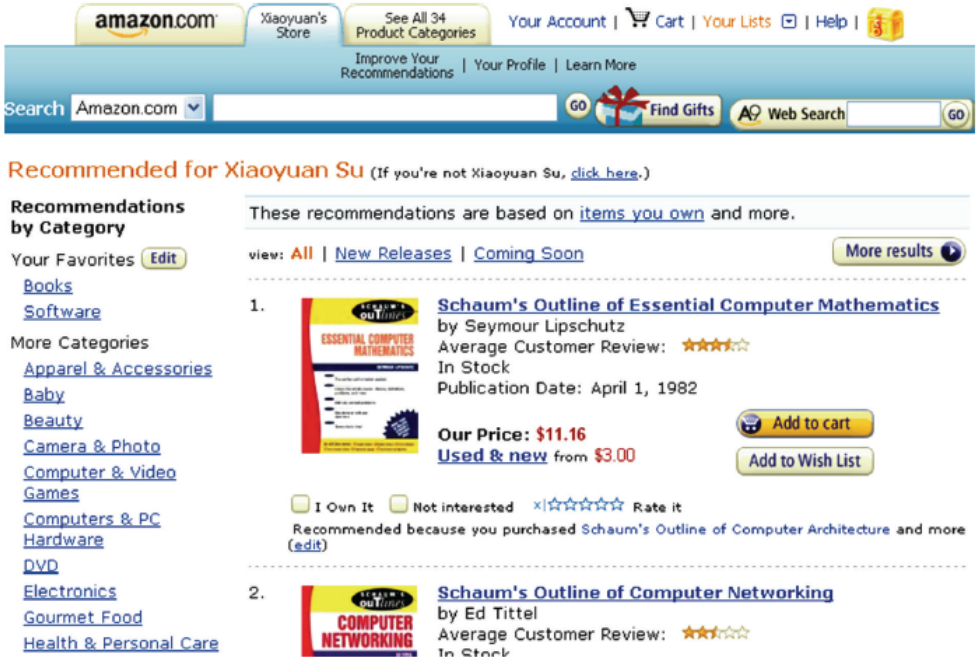
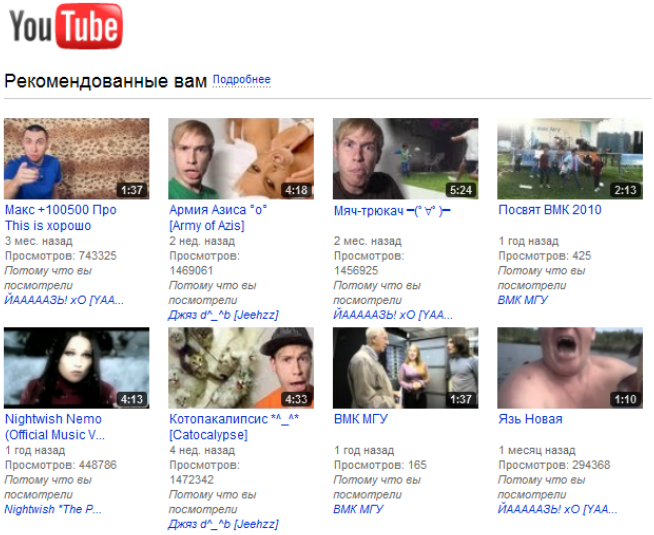
# Коллаборативная фильтрация с учетом временного фактора

**Выполнил:**
студент 428 группы
*Гомзин А.Г.* (428 гр.)

Руководитель: *Коршунов А.В.*

# Введение. Системы рекомендаций



# Постановка задачи

- Исследовать основные алгоритмы коллаборативной фильтрации
- Разработать систему для автоматического построения рекомендаций пользователей и объектов на основе коллаборативной фильтрации, учитывающую изменение предпочтений пользователей с течением времени.
- Исследовать влияние временного фактора на результаты коллаборативной фильтрации

# Коллаборативная фильтрация

| Пользователи \ Объекты | 1 | 2 | ... | $i$ | ... | $m$ |
|---|---|---|---|---|---|---|
| 1 | 5 | 3 | | 1 | 2 | |
| 2 | | 2 | | | | 4 |
| : | | | 5 | | | |
| $u$ | 3 | 4 | | 2 | 1 | |
| : | | | | | 4 | |
| $n$ | | | 3 | 2 | | |
| $a$ | 3 | 5 | | ? | 1 | |

**Алгоритм:**

1. Для каждого пользователя $u$ вычислим, насколько его интересы совпадают с интересами пользователя $a$
2. После этого выберем множество пользователей, наиболее близких к $a$
3. Предскажем оценку на основе оценок объекта $i$ "соседями"из предыдущего шага

# Коллаборативная фильтрация с учетом временного фактора

**Постановка задачи**

- *Объекты* – музыкальные композиции и исполнители
- *Пользователь:* набор прослушанных треков (включая время прослушивания)
- *Задача* – найти алгоритм рекомендации, учитывающий изменчивость вкусов пользователей

**Исходные данные**

Социальная сеть *last.fm*.

$\approx$ 6000 пользователей с историями прослушивания: 15 случайных пользователей, их друзья и друзья друзей.

# Коллаборативная фильтрация с учетом временного фактора

**Вектор интересов по исполнителям**

$$V_{Artist}(u, a)_\alpha = \sum_{\substack{i \in I \\ artist(i)=a}} e^{-\alpha \times time(i)}$$

**Вектор интересов по композициям**

$$V_{Track}(u, s)_\alpha = \sum_{\substack{i \in I \\ track(i)=s}} e^{-\alpha \times time(i)}$$

$u$ – текущий пользователь, $I$ – множество прослушиваний, $\alpha$ –параметр.
*Прослушивание* $i$: $artist(i)$ – исполнитель, $track(i)$ – название композиции, $time(i)$ – время прослушивания.

# Коллаборативная фильтрация с учетом временного фактора

**Мера близости интересов интересов по исполнителям**

$$sim_{Artist}(u, v) = \frac{V_{Artist}(u)_\alpha \cdot V_{Artist}(v)_0}{\mid V_{Artist}(u)_\alpha \mid \cdot \mid V_{Artist}(v)_0 \mid}$$

**Мера близости интересов по композициям**

$$sim_{Track}(u, v) = \frac{V_{Track}(u)_\alpha \cdot V_{Track}(v)_0}{\mid V_{Track}(u)_\alpha \mid \cdot \mid V_{Track}(v)_0 \mid}$$

**Общая мера близости интересов**

$$sim(u, v) = W_{Artist} sim_{Artist}(u, v) + W_{Track} sim_{Track}(u, v)$$

$u$ – текущий пользователь, $\alpha$ – параметр, $W_{Artist}$, $W_{Track}$ – весовые коэффициенты.

# Коллаборативная фильтрация с учетом временного фактора

**Предсказание неявной оценки**

$$p_{u,a} = \frac{\sum_{v \in K} V_{Artist}(v, a)_0 \times sim(u, v)}{\sum_{v \in K} sim(u, v)}$$

$u$ – текущий пользователь. $K$ – множество $k$ наиболее близких по интересам к $u$ пользователей.

# Оценка качества

**Разработаны два метода оценки качества**

- Качество предсказания
- Качество выбора кандидатов

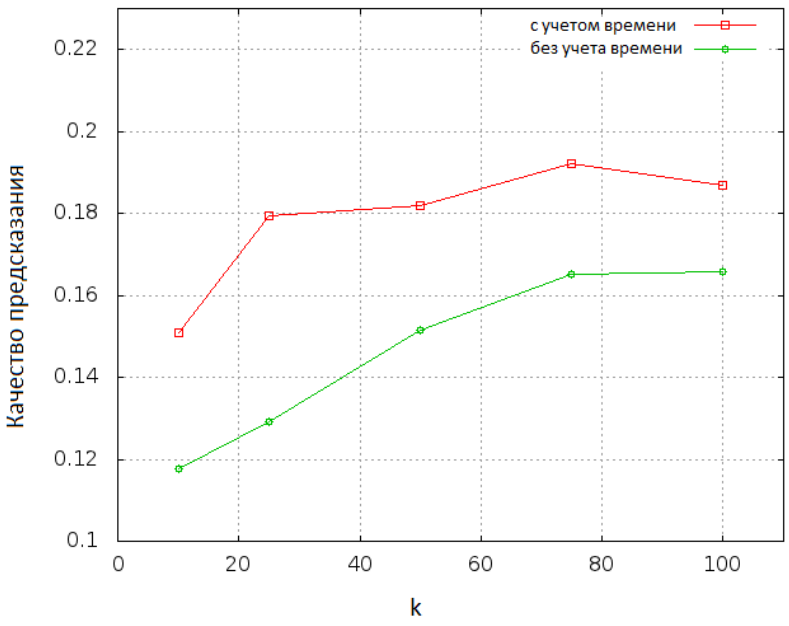
# Результаты

| | Качество предсказания | | Качество выбора кандидатов | |
|---|---|---|---|---|
| k | + время | базовый | + время | базовый |
| 10 | 0.15073 | 0.11764 | 0.01481 | 0.01297 |
| 25 | 0.17941 | 0.12911 | 0.02026 | 0.01705 |
| 50 | 0.18191 | 0.15147 | 0.02509 | 0.02291 |
| 75 | 0.19215 | 0.16509 | 0.02867 | 0.02492 |
| 100 | 0.18676 | 0.16580 | 0.03143 | 0.02763 |

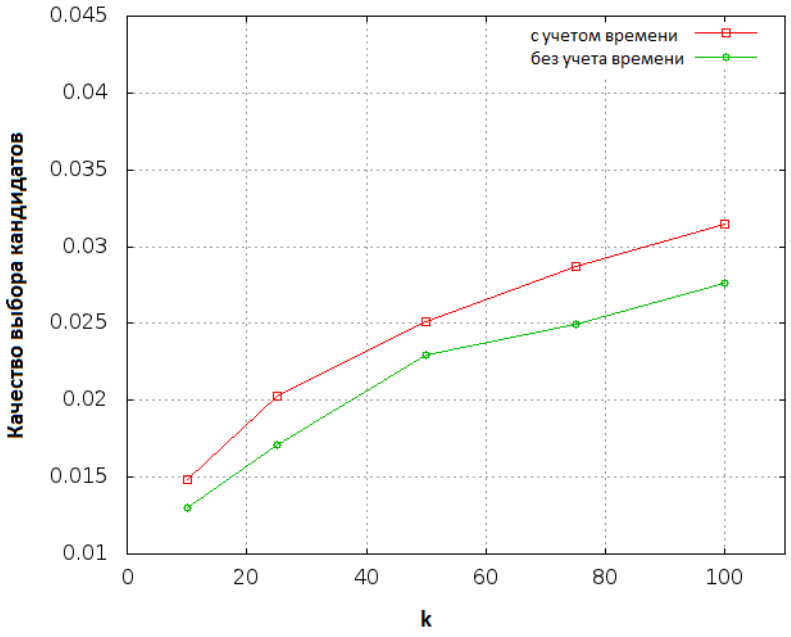
Таблица: *Качество предсказания* и *качество выбора кандидатов* при $\alpha = 0.01$

$$|\ K\ | = 20,\ \mathrm{T} \approx 19 \text{ ноября } 2011 \text{ года.}$$
$$W_{Artist} = 1,\ W_{Track} = 2$$

# Результаты




Качество предсказания

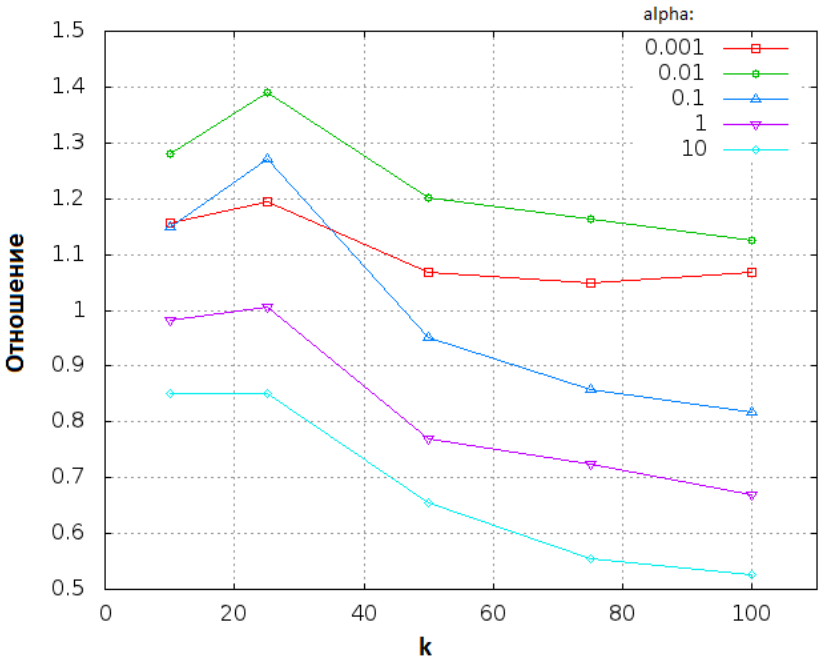



Качество выбора кандидатов

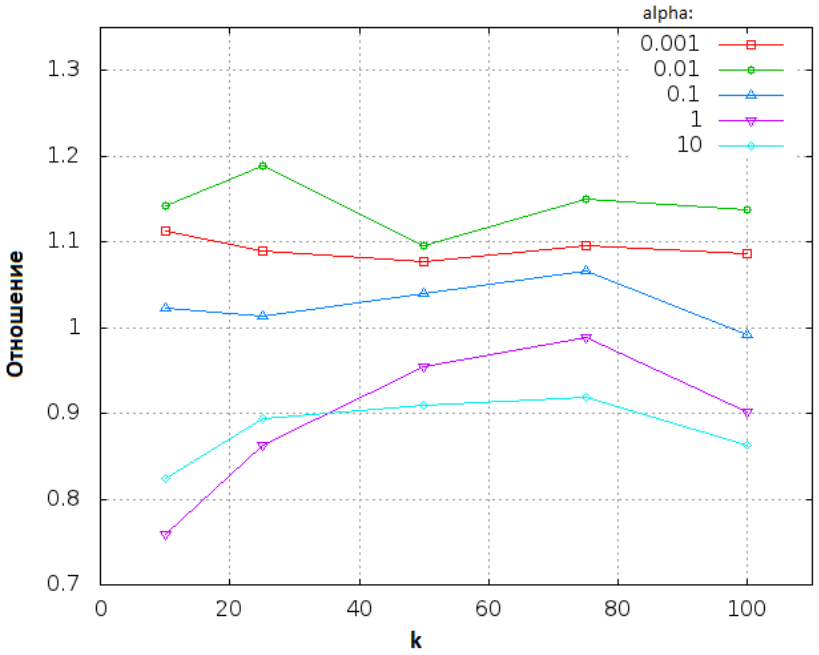
$$|\ K\ | = 20,\ \mathrm{T} \approx 19 \text{ ноября } 2011 \text{ года}$$
$$W_{Artist} = 1,\ W_{Track} = 2,\ \alpha = 0.01$$

# Результаты

*Отношение значения метрики для алгоритма с учетом времени к значению метрики для алгоритма без учета времени для различных $\alpha$*




Качество предсказания




Качество выбора кандидатов

$$|\ K\ | = 20,\ \mathrm{T} \approx 19 \text{ ноября } 2011 \text{ года}$$

$$W_{Artist} = 1,\ W_{Track} = 2$$

# Выводы

- Исследован алгоритм коллаборативной фильтрации
- Разработан алгоритм, рекомендующий исполнителей музыкальных композиций с учетом меняющихся со временем музыкальных интересов пользователей
- Разработаны способы оценки влияния временного фактора на результаты рекомендаций
- Экспериментально доказано, что учет временного фактора улучшает результаты систем рекомендаций

# Оценка качества. Качество предсказания

# Оценка качества. Качество выбора кандидатов