Рекомендательный сервис на основе Графов знаний

Кинопоиск 1.3.3.7

https://github.com/GrandArchTemplar/kp

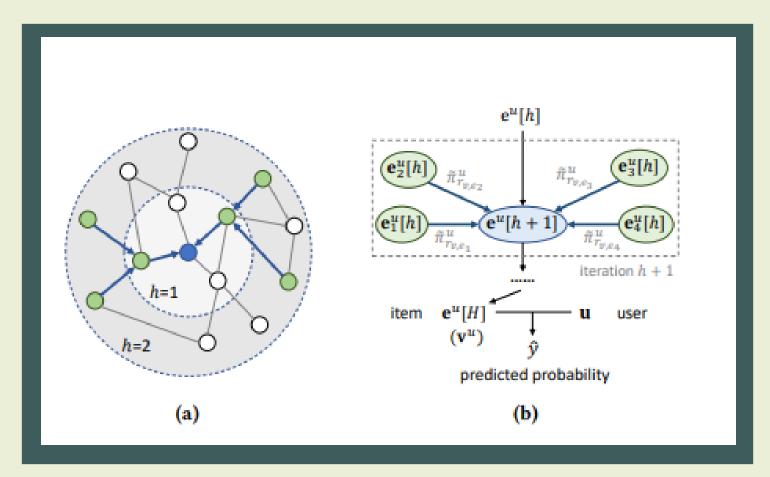
Выполнили студенты группы R4295½:

Павел Гилев Тушавин Глеб Шипицын Иван Андреева Полина

Алгоритм KGCN

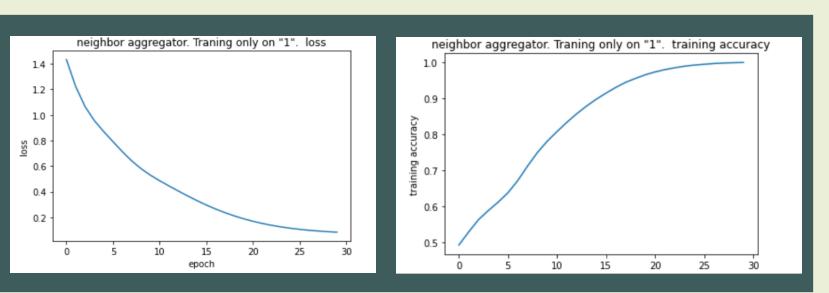
Идеи:

- эмбеддинги фильмов на основе соседей
- влияние отношений на принятие решения
- трансдуктивный метод



Еще немного про алгоритм

- Попытки улучшить.
- Ограничения
- Итоговое решение: схожесть вкусов максимальное совпадение понравившихся фильмов
- Результаты:



Точность на валидационном сете для известных юзеров (усредненная по всем юзерам):

F1 = 0.656

Точность на валидационном сете для новых юзеров при поиске самых похожих:

F1 = 0.395

Данные

- Рейтинги с movielens(20m)
- Граф знаний на основе imdb-movie датасета
- Некоторые данные пришлось урезать из-за нехватки вычислительных мощностей

Neo4ј в теории

Плюсики

- Графовые бд для графа знаний
- Есть приятная библиотека на јаva для работы с ней
- Поддерживает из коробки не только http, но и bolt
- Система транзакций

Минусики

- Дефолтный проект с сайта не всегда делает то, что задумано
- Есть драйвера под руби, перл и пхп, но нет под раст :с

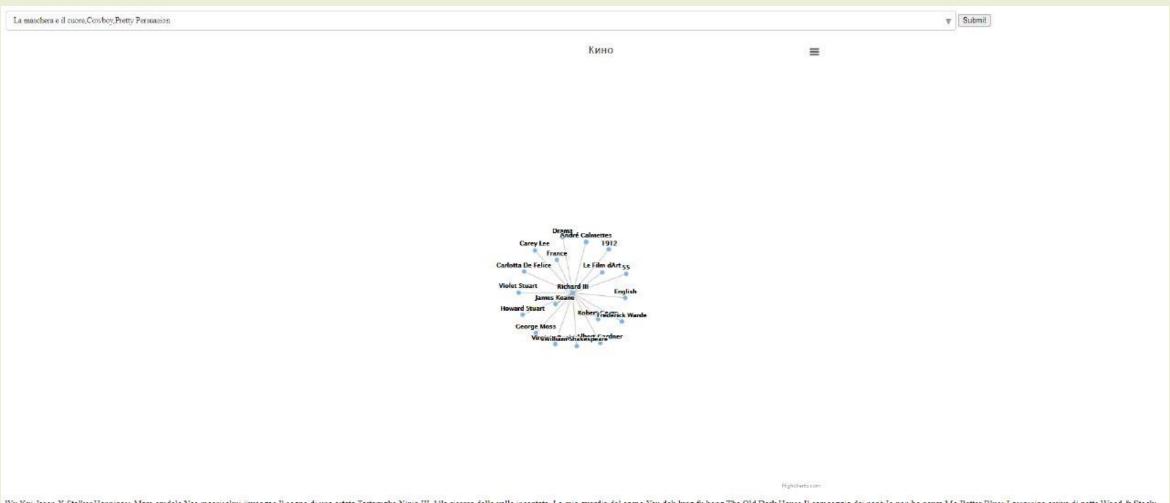
Neo4і в реальности

- Делать запрос всего графа знаний очень больно
- Кажется нет встроенной ленивости
- Не очень удобно не файлом подгружать и выгружать много данных

Визуализация

- Визуализировать перемножение векторов непоказательно
- Визуализировать пути -- не соответствует алгоритму
- Визуализировать эмбеддинги в каком-либо цветом спектре и их преобразования в ходе тренировки/поиска -- тоже не очень показательно
- Решили сделать второй вариант, но проблемы с бд :с

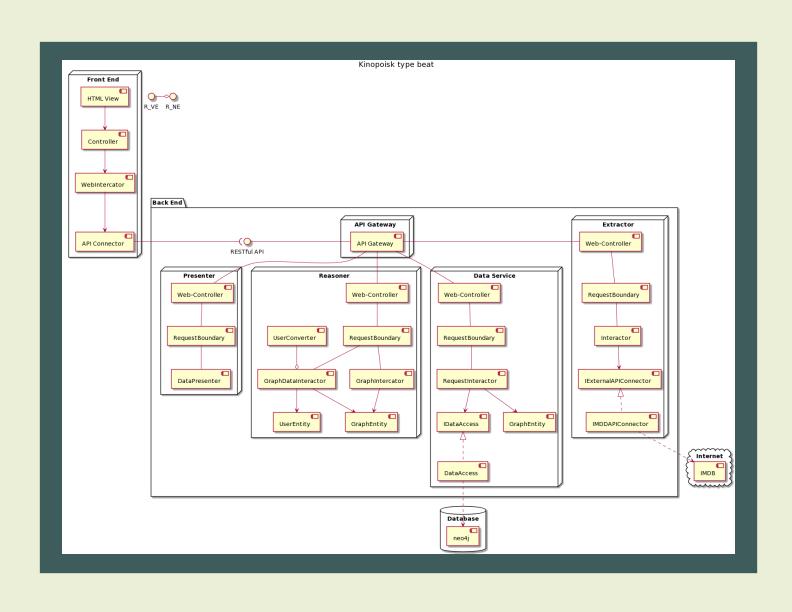
Визуализация как примерно она должна быть



Wu Kui, Jason X, Stalker, Happiness, Mare crudele, Nae meorisokui jiwoogae, Il sogno di una estate, Tartarughe Ninja III, Alla ricerca della valle incantata, La mia guardia del corpo, Yau doh lung fu bong, The Old Dark House, Il campeggio dei papă, Io non ho paura, Mo Better Blues, Lassassino arriva di notte, Wood & Stock:

Sexo, Corto circuito 2, Pazzo di te, Heavens Fall, Lamica, Lo squartatore di New York, Fred Claus - Un fratello sotto lalbero, Kenny, Went to Coney Island on a Mission from God... Be Back by Five, Koda, Contenders serie 7, Kangwon-do ui him, Anche gli angeli mangiano fagioli, V2 - jäätynyt enkeli, Il rompiscatole, Milioni in pericolo Pioggia di soldi, Il colpo di grazia, Long quan, Vibrator, Matrix Reloaded La schivata, Big Daddy - Un papà speciale, Alan, Lodio, Un secret, Il terrore viene dalla pioggia, Scary Movie 2, La sposa siriana Linfiltrato, Derecho de familia, Lucia y el sexo, De unge àt: Erik Nietzsche sagaen del 1

Диаграмма компонентов



Спасибо за внимание! Вопросы???