**Гибридная рекомендательная система на основе графа:**

Пусть – набор пользователей,

– набор фильмов,

– матрица размером , где каждая ячейка содержит оценку пользователяна соответствующий фильм (если оценка отсутствует, значение устанавливается равным нулю).

– матрица размером , содержащая дополнительную информацию о пользователях (пол, страна проживания, возраст и т.д.),

– матрица, содержащая дополнительную информацию об объектах (жанр, год выхода и т.д.).

**Задача:** предсказать рейтинги фильмов для каждого пользователя.

Алгоритм:

**Вход:**  ,

**Выход:** матрица размером из наиболее вероятных оценок пользователей на фильмы

Пусть – граф сходства пользователей (Similarity Graph, пользователи - узлы)

Тогда:

Для каждого пользователя в :  
 Для каждого пользователя != в :

– процент объектов со схожими рейтингами между двумя пользователями ()

– агрегированное сходство пользователей и , высчитывающееся на основе

=

Пусть – граф признаков пользователей, который мы строим на основе графа схожести пользователей (. Данный граф представляет собой матрицу размером , где – количество пользователей и – количество признаков, и содержит для каждого пользователя такие признаки как Degree Centrality (количество связей, имеющихся у узла), Closeness Centrality (мера близости узла ко всем остальным узлам в сети), Average Neighbor Degree (средняя степень соседа) и т.д.

–матрица категоризированных характеристик пользователей размером , получающаяся в результате обработки дополнительной информации о пользователях ()

Тогда – матрица размером , получающаяся в результате объединения матриц и , и содержащая для каждого пользователя как его признаки, так и его характеристики

Пропустим эту матрицу через автоэнкодер и получим матрицу из векторов-признаков. Количество измерений при этом уменьшится до

Пусть – оптимальное количество кластеров для

Тогда набор кластеров, который мы получаем в результате разделения векторов из на кластеров методом k-средних

Сгенерируем матрицу принадлежностей пользователей кластерам

Пусть матрица приближённых рейтингов фильмов в кластерах

Тогда:

Для каждого кластера из :

Для каждого фильма из :

Если есть пользователи, которые оценили фильм в :

[] = среднее от оценок этих пользователей на фильм в матрице

Либо если есть схожие фильмы фильму , которые имеют оценку в кластере :

[] = среднее от оценок этих пользователей на похожие фильмы на фильм в

Либо:

[] = среднее от всех оценок пользователей из кластера в

Приближаем рейтинги пользователей путём перемножения матриц и и деления каждого элемента на общее количество фильмов

 