

## РАЗРАБОТКА АЛГОРИТМА РЕКОМЕНДАЦИИ ВЫБОРА ПРЕДМЕТА ДЛЯ ОБУЧЕНИЯ

Нехай И.А.

Белорусский государственный университет информатики и радиоэлектроники, г. Минск, Беларусь,  
igorn1995@yandex.by

Abstract. Presented algorithm for university or master course subject recommendation based on collaborative filtering approach.

Разрабатываемое программное средство создаётся для решения проблемы выбора предметов во время обучения на первой и второй ступенях высшего образования, а также для анализирования популярности тех или иных направлений обучения среди студентов и магистрантов. Данное программное средство может успешно использоваться для рекомендаций по выбору курсов среди студентов и магистрантов, а также для оценки качества материалов, преподаваемых в рамках конкретного курса. Также данный подход помогает определиться с выбором конкретного направления для продолжения образования на второй ступени обучения выпускникам.

Для реализации данного проекта необходим сбор начальной базы оценок для различных предметов. Для этого требуется произвести сбор статистики среди студентов и магистрантов, которые должны оценить заданные курсы по десятибалльной шкале в зависимости от личных предпочтений. Эти данные заносятся в базу для дальнейшей обработки, при этом от каждого студента не требуется оценивать все возможные курсы и предметы, достаточно лишь произвести оценку заданного подмножества всех предметов. Важно отметить, что данные после ввода в базу обезличиваются. Для работы алгоритма рекомендации все данные загружаются в матрицу  $Y$ , где каждый столбец соответствует опрошенному студенту, а каждая строка – оценённому предмету. В ячейках хранятся оценки предметов или нули, если предмет не был оценён. На базе данной матрицы также создаётся дополнительная матрица  $R$ , аналогичная предыдущей, которая указывает, был ли оценён тот или иной предмет студентом. Для вычисления предсказаний используется алгоритм коллаборативной фильтрации. Коллаборативная фильтрация вырабатывает рекомендации, основанные на модели предшествующего поведения пользователя. Эта модель может быть построена исключительно на основе поведения данного пользователя или – что более эффективно – с учетом поведения других пользователей со сходными характеристиками [1]. Для оптимальности рекомендаций наиболее предпочтительным является второй вариант, однако он требует больше трудоёмкости при сборе данных, необходимых для формирования статистической базы. Для запуска процесса обучения необходимо реализовать функцию вычисления стоимости с поддержкой L2-регуляризации во избежание переобучения модели, что приводит к плохим результатам при предсказании рекомендаций для новых пользователей программного средства [2]. Функция

стоимости при реализации алгоритма коллаборативной фильтрации имеет следующий вид:

$$J(x, \theta) = 0.5 \sum_{(i,j):r(i,j)=1} ((\theta^{(j)})^T x^{(i)} - y^{(i,j)})^2 + 0.5 \lambda \left( \sum_{i=1}^n \sum_{k=1}^n (x_k^{(i)})^2 + \sum_{j=1}^n \sum_{k=1}^n (\theta_k^{(j)})^2 \right); \quad (1)$$

где  $x$  – матрица признаков.

Следует отметить, что при увеличении размера матрицы признаков увеличивается и точность рекомендаций, однако это также негативно влияет на производительность алгоритма.

Функция (1) минимизируется при помощи градиентного спуска или любого другого алгоритма оптимизации, после чего становится возможным выполнить предсказание для произвольного студента по формуле:

$$P = \theta^T x; \quad (2)$$

где  $x$  – вектор оценок данного студента.

Для реализации оптимизации функции (1) в данном программном средстве используется метод сопряжённых градиентов, который показывает хорошие результаты на практике. В целях производительности, для увеличения скорости сходимости данного алгоритма используется метод Полака-Райбера.

Кроме непосредственно реализации алгоритма коллаборативной фильтрации, в программе также поддерживается возможность непосредственно совершения предсказания для нового пользователя на основе выставленных им оценок для нескольких предметов и на основе оценок студентов с аналогичными предпочтениями по формуле (2).

Данное программное средство также может использоваться для вывода статистических графиков с целью визуализации корреляции между различными предметами, получения статистики по числу обучавшихся различным предметам, для вывода графиков изменения популярности тех или иных предметов с течением времени.

### Литература

1. М. Джонс – Рекомендательные системы. [Электронный ресурс]. – Режим доступа: <https://www.ibm.com/developerworks/ru/library/os-recommender1/index.html>.
2. Э. Ён – Машинное обучение. [Электронный ресурс]. – Режим доступа: <https://www.coursera.org/learn/machine-learning>.