

**Д. А. Хвесюк**

(ГГУ имени Ф. Скорины, Гомель)

Науч. рук. **Н. А. Аксёнова**, ст. преподаватель

### **РЕШЕНИЕ ПРОБЛЕМЫ ХОЛОДНОГО СТАРТА В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ**

В современном мире большинство компаний внедряют в свои продукты различные модели машинного обучения. Существует огромное множество задач, которые решают данные модели. Одной из таких задач является построение рекомендательной системы. Такие системы разрабатываются, чтобы предлагать пользователям именно товар (контент), который они с большей вероятностью купят (посмотрят). Актуальным подходом в решении данной задачи является коллаборативная фильтрация (collaborative filtering). Данный метод основывается на предположении, что у определенного товара может существовать целевая аудитория, которая интересуется и другими товарами. Тогда если пользователь купит один товар из этого множества, то хорошим решением будет рекомендацию данному пользователю и других товаров, которые выбираются похожими пользователями. Для построения такой модели собирается прореженная матрица взаимодействий пользователей и товаров. Далее методом матричных разложений мы можем получить такие матрицы, которые будут содержать некоторое векторное представление для товаров и пользователей. Данные представления можно использовать для восстановления пропущенных значений в матрице, либо поиска похожих объектов путей расстояния между объектами. Хотя данный подход и хорошо себя показывает в некоторых приложениях, у него имеется один существенный недостаток – проблема холодного старта. Из-за того, что модель строится из взаимодействий пользователь-товар, мы не можем строить модель на старте разработки, когда нашим приложением еще никто не пользуется. Также происходит и с новыми товарами, пока новым товаром никто не пользовался, мы не можем подобрать для него правильное векторное представление. Холодный старт со стороны товара может решить другими подходами получения векторных представлений. Например, мы можем взять за координаты признаковое описание наших товаров, если таковое имеется. Такой подход называется content-based. Для работы с

497

контентом можно воспользоваться методами NLP, например модификацией модели word2vec – doc2vec. Холодный старт со стороны пользователя уже от части может решиться построением представлений для товаров. Однако также можно построить и модель поиска похожих пользователей на основе опроса предпочтений (часто используется при первом подключении к стриминговому сервисам). Надо помнить, что данные методы дают рекомендации хуже самой коллаборативной фильтрации, поэтому были придуманы гибридные системы, совмещающие оба подхода.