

# ГИБРИДНАЯ МОДЕЛЬ СТОХАСТИЧЕСКОГО ГРАДИЕНТНОГО СПУСКА И ВЫБОРОК ГИББСА В РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМАХ

Петик М.Г., Канищев И.С.

Вятский государственный университет,  
Институт компьютерных и физико-математических наук,  
кафедра математического моделирования.  
Россия, 610000, г.Киров, ул.Московская, д.36,  
E-mail: [marypetik@gmail.com](mailto:marypetik@gmail.com), [kanishchev.ilya@gmail.com](mailto:kanishchev.ilya@gmail.com)

Рекомендательные системы являются одним из важнейших научных направлений в области анализа данных. Такие системы имеют очень большое прикладное значение применительно к спорту, образованию и маркетингу. В последнем случае мы заинтересованы в построении прогноза того как клиент оценит качество товара.

Типичной для рекомендательных систем является ситуация, когда количество клиентов исчисляется сотнями тысяч, а количество товаров десятками тысяч, при этом анализируемые данные представлены в форме списка. В нашей работе предложено использование гибридного алгоритма на основе факторных разложений с использованием выборки Гиббса. Эффективность метода доказана в наших экспериментах на платформе *Kaggle inClass*, где использовались классические данные MovieLens.

Формат данных Netflix, которые использовались в ходе наших экспериментов, исключительно прост: 1) индекс клиента, 2) индекс видеофильма, 3) рейтинг по пятибалльной шкале как клиент оценивает видеофильм. В настоящее время данные Netflix рассматриваются как классические в области рекомендательных систем.

В использованных данных числа клиентов и видеофильмов исчисляются сотнями и десятками тысяч, соответственно. Известные рейтинги составляют не более одного процента от теоретически полного объема информации. По этой причине данные даны в форме списка, но не в традиционной форме матрицы. Отметим, что методы стохастического градиентного спуска (СГС) проявили себя очень эффективными при работе со списками данных, когда подавляющее большинство отношений отсутствует и может быть востребовано в качестве прогноза.

Выборки Гиббса замечательны тем, что для них не требуются явно выраженные совместные распределения, а нужны лишь условные вероятности для каждой переменной, входящей в распределение. Алгоритм на каждом шаге берет одну случайную величину и выбирает ее значение при условии фиксированных остальных. Можно показать, что последовательность получаемых значений образуют возвратную цепь Маркова, устойчивое распределение которой является как раз искомым совместным распределением.