

Repeated learning in recommender systems

This paper addresses the issue of evaluating the quality of recommender systems in the long term, taking into account the evolution of consumers and product assortments. We consider the dynamical system of changes in consumers and products over time. The main purpose of the study is to identify the conditions under which degeneracies in audience, assortment, or transaction distribution occur in a given repeated machine learning system, and how such phenomena depend on the learning algorithms and recommendation models. Using the obtained results, we presented a model that is able to increase the metrics in the recommendation systems without degenerating the distributions on products and customers. We conduct a series of computational experiments on the synthetic datasets, the results of the experiments correspond to the theoretical predictions derived from the dynamical model.

1 Introduction

Recommender systems play a crucial role in various domains, including e-commerce, entertainment, and information retrieval, by assisting users in discovering relevant items or content. Evaluating the quality and performance of recommender systems over the long term is essential to ensure their effectiveness as consumer preferences and product assortments evolve. In this context, understanding the dynamics of changes in consumer behavior and product availability is paramount for maintaining the system's relevance and utility.

Research on Echo Chambers in E-commerce recommender systems and hidden feedback loops in predictive models has shed light on the complexities of maintaining system efficiency amidst evolving dynamics. These studies have emphasized the importance of addressing issues such as algorithmic biases, filter bubbles, and feedback loops that can impact the performance and reliability of recommendation models.

Existing literature has explored various methods to tackle similar challenges, highlighting the complexity of optimizing recommendation metrics while ensuring the stability of distribution patterns over time. Building upon this foundation, our study introduces a novel approach that focuses on preventing degeneracies in distribution while enhancing recommendation system metrics. By presenting a theoretical framework that outlines the conditions under which degeneracies occur and conducting computational experiments on synthetic datasets, we aim to provide valuable insights into the interplay between learning algorithms, recommendation models, and system degeneration.

The main goal of our research is to determine the sufficient conditions under which degeneracies in audience or product distribution arise due to the evolution operator. We aim to establish theoretical theorems regarding the conditions for degeneracy in a general setting and conduct experiments to identify algorithms that trigger degeneration in dynamic systems. By combining theoretical analysis with empirical results, we seek to advance the understanding of how learning algorithms and recommendation models influence the long-term performance and stability of recommender systems.

2 Problem statement

Рассмотрим модель динамической рекомендательной системы. Пусть C - множество покупателей, W - множество товаров, $|C|, |W| < \infty$. Покупатели и товары описываются конечным числом признаков, то есть $C \subset \mathbb{R}^d, W \subset \mathbb{R}^l$. На каждом шаге t имеется распределения с плотностями f_c^t и f_w^t на множествах C и W соответственно, при $t = 0$

оба распределения известны. Данные функции отражают распределения признаков среди покупателей и товаров в системе. В последствии эти распределения будут меняться в процессе работы системы. Эти изменения отражает оператор эволюции $D : R_d \times R_l \rightarrow R_d \times R_l$, где R_k - множество всех функций плотности на \mathbb{R}^k . Таким образом, $(f_c^{t+1}, f_w^{t+1}) = D(f_c^t, f_w^t)$. Цель работы - определить, при каких достаточных условиях на оператор эволюции D в пределе возникает вырождение распределения покупателей или распределения товаров. Под вырождением распределения покупателей имеется в виду, что $f_c^t \rightarrow f_c$, где носитель f_c - подпространство \mathbb{R}^d меньшей размерности. Аналогично определяем вырождение для распределения товаров. Теперь рассмотрим конкретный пример динамической системы. Пусть имеется функция $\mu_t : C \times W \rightarrow [0, 1]$ - агрегированный фидбек⁶ для каждой пары из товара и пользователя за время работы системы, $\mu_0(c, w) = \delta > 0 \forall c, w$. Теперь опишем одну итерацию в модели работы динамической системы:

1. Происходит сэмплирование выборок произвольного размера покупателей и товаров из распределений f_c^t и f_w^t : $c_1^t, \dots, c_n^t \sim f_c^t$, $w_1^t, \dots, w_k^t \sim f_w^t$.
2. Рекомендательная система подбирает для каждого покупателя из выборки подмножество товаров из w_1^t, \dots, w_k^t .
3. Покупатели совершают покупку с вероятностью $u(c, w)$, где $u : C \times W \rightarrow [0, 1]$ - фиксированная функция полезности товара для данного пользователя. Далее рекомендательная система дообучается на полученном фидбэке.
4. Шаги 2-3 повторяются произвольное количество раз.
5. Обозначим за M_{t+1} - набор фидбэков, полученных за все повторения шага 3 на данной итерации, то есть мультимножество из троек вида (c, w, m) , где $c \in c_1^t, \dots, c_n^t$, $w \in w_1^t, \dots, w_k^t$, $m \in \{0, 1\}$. 1 обозначает совершение покупки, 0 - её отсутствие. Далее обновляется функция μ_t . Пусть сначала $\mu_{t+1} = \mu_t$. Далее для каждой тройки (c, w, m) из M_{t+1} функцию μ_{t+1} изменяется так: $\mu_{t+1}(c, w) = \lambda m + (1 - \lambda)\mu_{t+1}(c, w)$, где $\lambda \in [0, 1]$ - некоторая константа, отражающее скорость забывания старого фидбэка.
6. Далее происходит обновление распределений товаров и пользователей в соответствии с новой функцией μ_{t+1} : $f_w^{t+1}(w_i) = \frac{\sum_{c \in C} \mu_{t+1}(c, w_i)}{\sum_{w \in W, c \in C} \mu_{t+1}(c, w)}$. Аналогично обновляется и f_c^t . В данной работе мы рассматриваем как общую постановку проблемы вырождения оператора эволюции, так и описанный выше конкретный пример динамической системы.
- ⁷ Цель - доказать теоремы о достаточных условиях вырождения для общей постановки проблемы, а также провести эксперименты для описанного примера системы и выявить алгоритмы, которые вызывают вырождение динамической системы.

0. Лучше все писать в настоящем времени

1. Здесь нужна какая-то большая (или несколько больших) работа, подтверждающая утверждения о важной роли рекомендательных систем и т.п.

2. В этом параграфе есть несколько утверждений, каждое должно быть подкреплено ссылкой на литературу

3. Очень коротко и по верхам. Нужно добавить ссылки на литературу и сказать пару слов о каждой работе. Если вы сможете привязать работу к вашей теме - будет еще лучше

4. Здесь и дальше по всему тексту - вместо короткого дефиса нужно длинное тире. В латехе оно ставится как три дефиса "---"

5. Обсудите с Антоном, у него могут быть другие предпочтения по нотации/оформлению формул, но мы на курсе рекомендуем правила

- * вектор-функции и векторы пишутся прописными жирными буквами

- * матрицы пишется заглавными жирными буквами

(см презентацию с прошлого занятия)

Нужно пройти по всему тексту

6. обратная связь. Я так понял, текст планируется потом перевести на английский. Но пока вы пишете на русском, лучше использовать термины, устоявшиеся в русскоязычной литературе

7. Цель описана несколько топорно. А если не удастся доказать теорему? Может быть там теоремы нет, но есть теоретический анализ. Нужно развить и описать более аккуратно.