

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

УДК 519.2

Отчет об исследовательском проекте на тему:
Тензорные бандиты и их приложения
(промежуточный, этап 1)

Выполнил студент:

группы БПМИ212, 3 курса

Горбач Марина Павловна

Принял руководитель проекта:

Самсонов Сергей Владимирович

Научный сотрудник, аспирант

Международная лаборатория стохастических алгоритмов и
анализа многомерных данных, ФКН НИУ ВШЭ

Москва 2024

Содержание

Аннотация	3
1 Введение	4
2 Обзор литературы	5
3 План дальнейшей работы	7
Список литературы	8

Аннотация

Алгоритмы многоруких бандитов являются мощным теоретическим аппаратом, который позволяет эффективно решать множество практических задач, такие как задачи рекомендательных систем и динамического ценообразования. При этом множество существующих алгоритмов страдает от так называемого “проклятья размерности” - оценки сложности алгоритма в худшем случае зависят линейно от количества действий, которых может быть очень много. Для того, чтобы преодолеть эту теоретическую проблему, недавно была предложена парадигма низкоранговых тензорных бандитов, основанная на предположении низкого ранга у тензора ожидаемых наград. В рамках проекта ожидается обобщить эту парадигму на новые виды тензорных разложений и проверить работоспособность подхода на ряде практических задач.

Ключевые слова

Алгоритм многоруких бандитов, низкоранговые приближения, разложение Таккера, тензорные вычисления

1 Введение

В современном мире принятие решений в условиях неопределенности становится всё более важной и трудоёмкой задачей, так как пространство выбора, как и объёмы данных, постоянно растёт. Например, как описано в [5], у сервисов, предоставляющих персонализированный контент, постоянно возникает необходимость подобрать рекламу, новые предложения и т. д. на основании данных контента и самого пользователя. Подобные задачи осложняются тем, что постоянно появляется новый контент, новые пользователи, а также данные непрерывно меняются. Соответственно, большинство ранее распространённых способов, таких как коллаборативная фильтрация, стали непригодными к использованию. Чтобы удовлетворять современным требованиям, алгоритм должен быстро обучаться на новых данных и учитывать обновления, а также быстро принимать решения.

Одним из способов решения таких задач являются алгоритмы многоруких бандитов, так как они позволяют быстро реагировать на изменение среды, а также учитывать контекст. Алгоритмы многоруких бандитов – алгоритмы, в которых агенту необходимо сделать выбор, где каждый вариант представлен ручкой. При выборе каждая из ручек возвращает некоторую награду, максимизировать которую – цель агента. В данной работе речь пойдёт о тензорном представлении алгоритмов такого типа.

В контексте задачи о рекомендательной системе [15] это будет устроено следующим образом. Например, пользователю нужно порекомендовать музыку, тогда каждый выбор в алгоритме будет характеризоваться набором из жанра, исполнителя, названия произведения и других характеристик. Такой набор далее будет представлен тензором, что поможет в последствии оценить ожидаемую награду и определить оптимальный вариант. После его использования будет получен ответ от среды: пользователь прослушает трек, добавит его в плейлист или же пропустит, таким образом алгоритм будет общаться со средой и, следовательно, обучаться. Работа с тензорами в данном случае помогает эффективно моделировать контекст, который обладает множеством возможных значений.

На данный момент существует множество алгоритмов для решения задачи многорукого бандита. Один из наиболее популярных – алгоритм верхних доверительных границ, так как он показывает хорошие результаты на широком спектре задач [2]. Однако, применять его для тензорного варианта задачи затруднительно, так как операция разложения тензора не является выпуклой [15].

Для решения этой проблемы была создана серия алгоритмов, использующих низкоранговые приближения, так как это существенно снижает сложность алгоритма. Данная

работа подразумевает исследование существующих методов, их дальнейшее их развитие для новых типов тензорных разложений, а также применение результатов к практической задаче.

Исследование было начато с реализации алгоритмов, использующих разложение Таккера, которые были описаны в [15]. В качестве базовой модели был реализован известный алгоритм Tensor-Elimination, идея которого в том, чтобы постепенно уменьшать вероятность выбора не оптимальных ручек на основании данных предыдущих ходов. Были проведены эксперименты на искусственно сгенерированных данных. Следующим шагом будет реализация других уже известных алгоритмов, основанных на разложении Таккера. Дальнейшее развитие проекта подразумевает, предполагает создание алгоритмов, использующих другие тензорные разложения, в том числе тензорный поезд. Итогом работы должно стать применение разработанных алгоритмов в решении практической задачи.

2 Обзор литературы

Тензорные разложения В качестве самого первого источника для знакомства с особенностями тензорных разложений использовались материалы курса «Тензорные вычисления» [16]. В них содержалась основная информация об операциях с тензорами и низкоранговых разложениях, в том числе о разложении Таккера [12].

Помимо алгоритмов, основанных на разложении Таккера, планируется исследование алгоритмов с другими тензорными разложениями, в том числе с широко распространённым сейчас тензорным поездом [6].

Многие алгоритмы, рассматриваемые в данной работе использует операцию дополнения тензоров (tensor completion). Для знакомства сданной операцией использовалось [3]. Она подразумевает восстановление неизвестных элементов тензора на основании тех, что даны. Обычно это требуется для построения оценки тензора, моделирующего распределение награды по ручкам, построенной на основе данных, полученных во время уже сделанных шагов. Один из вариантов реализации операции дополнения также описан в [7].

Многорукие бандиты В последнее время алгоритмы многоруких бандитов развивались достаточно активно. Для знакомства с существующими алгоритмами и изучения основных идей использовались [4], [10], где описаны различные алгоритмы многоруких бандитов, в том числе стохастические и контекстные. Один из наиболее популярных на данный момент алгоритмов в силу хороших результатов – алгоритм верхних доверительных границ [1]. Его идея заключается в том, что для каждой ручки строится доверительный интервал, в котором

ождается её награда. Это делается на основании предыдущих шагов. В этом алгоритме решения принимаются исходя из принципа Оптимизма перед лицом неопределенности [4]. Он утверждает, что выбор должен делаться исходя из предположения, что среда будет настолько выгодной агенту, насколько это возможно. Поэтому данный алгоритм сравнивает ручки по верхним границам доверительных интервалов. Стоит также заметить, что по мере обучения алгоритма, он получает всё больше данных, что позволяет сужать доверительный интервал и, следовательно, делать оценки более точными.

Так как зачастую считать вышеописанный доверительный интервал достаточно проблематично, ещё один подход набирает популярность – алгоритм Томпсоновского сэмплирование [11], [8]. Идея данного алгоритма строится на предположении, что награда каждой ручки имеет некоторое распределение, неизвестное агенту, но которое не меняется со временем. Действуя в таком предположении, агент инициализирует для каждой ручки некоторое априорное распределение, которое обычно выбирается из общей информации о задаче. Далее после каждого шага, параметры распределения обновляются. Утверждается, что после некоторого числа итераций будет получено достаточно точное приближение реального распределения, на основании которого уже можно будет сделать выбор оптимальной ручки.

Обратим внимание также на метод, который является неотъемлемой частью большинства алгоритмов многоруких бандитов – метод рандомизации. Так как агенту никогда не известно распределение награды для каждой ручки, то нельзя исключать возможность того, что данные, которые были по ней получены содержат сильное смещение. Поэтому, с некоторой вероятностью, даже зная оптимальную ручку, алгоритм будет выбирать произвольную. Это уменьшит вероятность того, что наилучший выбор был отсеян в начале из-за недостаточного количества данных.

Контекстные многорукие бандиты Алгоритмы многоруких бандитов, которые описаны выше, обучаются исключительно на данных получаемых из общения со средой, которые представлены величинами наград. Однако, в реальности обычно существует какая-то информация о среде, которую агент не может учесть таким образом, однако, она может быть полезна. Алгоритмы контекстных бандитов учитывают такую информацию [13], [4]. Чаще всего это нужно для рекомендательных систем, где у контекста есть несколько различных признаков, например, возраст и пол пользователя.

Тензорные бандиты Одним из способов задать контекст в задаче многоруких бандитов является тензор. Каждая ручка задаётся в виде набора значений признаков, который позже

представляется в тензорном виде.

Для работы с тензорными бандитами были представлены следующие алгоритмы [15, 9, 14]. В первой статье на эту тему [15] представлено несколько алгоритмов, использующих разложения Таккера. Наилучшей оценкой на потери среди представленных алгоритмов обладает tensor elimination: $\overline{O}(p^{\frac{d}{2}} + p^{\frac{d-1}{2}} n^{\frac{1}{2}})$, где n – временной горизонт, d – размерность тензора наград, $p = \max(p_1, \dots, p_d)$ – наибольшая размерность. Алгоритм заключается в постепенном вытеснении не оптимальных ручек из множества выборов на основании их наград за предыдущие шаги. В [9] данный алгоритм был обобщён на более общее представление ручек, полученная оценка потерь: $\overline{O}(p^{\lfloor \frac{d}{2} \rfloor} n^{\frac{1}{2}})$, где r – мультилинейный ранг тензора наград. Затем было получено улучшение: $\overline{O}((p/r)^{\lfloor \frac{d}{2} \rfloor} - 2)$ – алгоритм TOFU.

В ноябре 2023 года вышла новая статья о тензорных бандитах [14]. Она существенно расширяет возможности для моделирования связи между признаками. В [15] для этого использовался метод one-hot encoding. Полученный алгоритм G-LowTESTR показывает результаты: $\overline{O}\left(\frac{p^2 \sqrt{\ell n}}{\sqrt{a(1-\gamma)}}\right)$, где γ – мера между определенными сингулярными числами тензора, задающего действие, a и ℓ – коэффициенты линейного преобразования.

3 План дальнейшей работы

В работе уже реализован алгоритм tensor elimination, описанный в [15]. Он продолжает идею алгоритма верхних доверительных границ, таким образом что, для каждой оценивается доверительный интервал наград на основании уже полученных данных, и далее ручки выбираются на основании полученных оценок.

Сейчас производится тестирование алгоритма на искусственных данных следующим образом. Для каждого выбора задаётся распределение. Далее, когда алгоритм обращается к среде с выбранной ручкой, из соответствующего распределения сэмплируется награда. Так алгоритм делает необходимое число шагов и определяет оптимальную ручку.

В процессе работы над данным алгоритмом были также реализованы базовые классы для работы с алгоритмами многоруких бандитов, что должно упростить дальнейшую работу.

Следующим шагом будет реализация жадного алгоритма (epoch-greedy), который также описан в [15], после чего будет произведено сравнение алгоритмов и проверка соответствия результатам статьи. Далее планируется разработать алгоритмы для других тензорных разложений, в частности тензорного поезда. И протестировать их аналогичным образом. Завершением работы должно стать практическое применение полученных алгоритмов для некоторой практической задачи и оценка результатов.

Список литературы

- [1] Peter Auer, Nicolo Cesa-Bianchi и Paul Fischer. “Finite-time analysis of the multiarmed bandit problem”. В: *Machine learning* 47 (2002), с. 235—256.
- [2] Sébastien Bubeck, Nicolo Cesa-Bianchi и др. “Regret analysis of stochastic and nonstochastic multi-armed bandit problems”. В: *Foundations and Trends® in Machine Learning* (2012), с. 1—122.
- [3] Olga Klopp. “Noisy low-rank matrix completion with general sampling distribution”. В: (2014).
- [4] Tor Lattimore и Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.
- [5] Lihong Li, Wei Chu, John Langford и Robert E Schapire. “A contextual-bandit approach to personalized news article recommendation”. В: *Proceedings of the 19th International Conference on the World Wide Web*. 2010, с. 661—670.
- [6] Ivan V Oseledets. “Tensor-train decomposition”. В: *SIAM Journal on Scientific Computing* 33.5 (2011), с. 2295—2317.
- [7] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen и др. “A tutorial on thompson sampling”. В: *Foundations and Trends® in Machine Learning* (2018).
- [8] Daniel J Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, Zheng Wen и др. “A tutorial on thompson sampling”. В: *Foundations and Trends® in Machine Learning* 11.1 (2018), с. 1—96.
- [9] Chengshuai Shi, Cong Shen и Nicholas D Sidiropoulos. “On High-dimensional and Low-rank Tensor Bandits”. В: *arXiv preprint arXiv:2305.03884* (2023).
- [10] Aleksandrs Slivkins и др. “Introduction to multi-armed bandits”. В: *Foundations and Trends® in Machine Learning* 12.1-2 (2019), с. 1—286.
- [11] W. R. Thompson. “On the theory of apportionment”. В: *American Journal of Mathematics*. (1935).
- [12] L. R. Tucker. “Implications of factor analysis of three-way matrices for measurement of change”. В: *Problems in measuring change*. Под ред. C. W. Harris. Madison WI: University of Wisconsin Press, 1963, с. 122—137.
- [13] Chih-Chun Wang, Sanjeev R Kulkarni и H Vincent Poor. “Bandit problems with side observations”. В: *IEEE Transactions on Automatic Control* 50.3 (2005), с. 338—355.

- [14] Qianxin Yi, Yiyang Yang, Yao Wang и Shaojie Tang. “Efficient Generalized Low-Rank Tensor Contextual Bandits”. В: *arXiv preprint arXiv:2311.01771* (2023).
- [15] Jie Zhou, Botao Hao, Zheng Wen, Jingfei Zhang и Will Wei Sun. “Stochastic Low-rank Tensor Bandits for Multi-dimensional Online Decision Making”. В: *arXiv preprint arXiv:2007.1578* (2020).
- [16] Рахуба Максим. *Основы тензорных вычислений*. ФКН НИУ ВШЭ. 2023. URL: [\(http://wiki.cs.hse.ru/%D0%9E%D1%81%D0%BD%D0%BE%D0%B2%D1%8B_%D1%82%D0%B5%D0%BD%D0%B7%D0%BE%D1%80%D0%BD%D1%8B%D1%85_%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B5%D0%BD%D0%B8%D0%B9_\(2023/24\)\)](http://wiki.cs.hse.ru/%D0%9E%D1%81%D0%BD%D0%BE%D0%B2%D1%8B_%D1%82%D0%B5%D0%BD%D0%B7%D0%BE%D1%80%D0%BD%D1%8B%D1%85_%D0%B2%D1%8B%D1%87%D0%B8%D1%81%D0%BB%D0%B5%D0%BD%D0%B8%D0%B9_(2023/24)) (дата обр. 01.09.2023).