

# Причинные нейронные сети “тензорного поезда”

Causal Tensor Train Neural Networks

Сергей А. Терехов, к.ф.-м.н.

РАНИ, РАИИ

Машинное обучение: эффективный инструмент для познания физической картины мира  
г. Москва, МГУ ○ НИИЯФ им Д.В. Скобелецкина ○ 18 ноября 2021



- Много-размерные цензурированные данные
- Причинно-следственные последовательности
- Нейронная сеть “тензорный поезд” (TTNN - Tensor Train Neural Network)
- Приложения: от логистики ресурсов до персональных коллекций текстов



# Часть I

## Мотивирующие проблемные ситуации

# Ситуация 1: Ресурсная логистика сети



Источник: [www.sandia.gov](http://www.sandia.gov)

- Крупная торговая сеть
  - 5,000+ торговых точек
  - Обширный динамичный ассортимент товаров
  - 30,000+ торговых работников
  - 2М посетителей ежедневно
  - 900 населенных пунктов
  - Цены, акции, логистика, запасы, поставщики, конкуренты...
- Сотни процессов, связанных с **принятием решений**

- Наблюдаются продажи в сети, описываемой множествами *склад*, *товар*, *дата*. Число продаж известно только для случаев, когда данный товар продавался на данном складе.
- Продажи не могут превышать имеющиеся остатки товара, поэтому часть наблюдений **цензурирована**. Данные представляют собой смеси не-цензурированных и цензурированных справа Пуассоновых отсчётов.
- Требуется получить оценку интенсивности продаж  $\beta$  для всех товаров на всех складах в течение всего времени наблюдений.
- **Почему важны статистические оценки**
  - Сами исходные наблюдения редко пригодны для анализа, т.к. являются реализациями случайных величин. Данные фрагментарны, искажены, получены с ошибками.
  - Оценки устойчивы и могут использовать априорную информацию (Т. Bayes).
  - Оценки могут быть получены для условий, в которых нет наблюдений.

- **Более общий случай.** Проблемная область описывается набором словарей («контекстов»). Каждый словарь состоит из знаков («слов»). Значимыми являются только сочетания знаков из разных контекстов.
- 6W: Who does What by What for What, Where and When.
- **Подход.** Знаки из контекста  $d$  кодируются латентными векторами  $U_d$ , а общая **ассоциативная** значимость контекстной ситуации описывается тензорной декомпозицией:

$$\beta \approx B(i_1, i_2, \dots, i_d) = \sum_{m=1}^M U_1(i_1, m) \cdot U_2(i_2, m) \cdot \dots \cdot U_d(i_d, m)$$

- Это решает описанную проблему товарной логистики, но не описывает ситуацию, когда значения одних переменных являются причинами для других.

## Ситуация 2: Персональные научные архивы

- За “цифровые” годы: 500+ книг, 10K+ статей
  - “Важное”, “Прочитать”, “Прочитать *срочно!*”, ...
  - Узкий тематический круг, оказывается, весьма обширен, отражает динамику интересов за годы научной и практической работы.
- Потребности едва формализуемы
  - Материалы для обзора по смежной теме. Цель - **не пропустить**.
  - Подготовить раздел учебной лекции типа “SVM как задача линейного программирования”. Цель - **правильно преподать** конкретной аудитории.
  - Экспертиза научного проекта, одно из направлений - не в круге текущей активной тематики эксперта. Цель - **проблемные места, открытые вопросы, риски**.
- Глобальные технологии: “Машина Тьюринга *купить в Москве...*”



Источник: pinterest.ru

- “Естественный” персональный отбор тематик и материала.
  - Только научная литература, произведённая учеными и специалистами, отражающая логику их мыслительного процесса. *Causal-by-selection*
  - Не энциклопедические статьи, отражающие состояние дел в справочной форме.
- Базовый смысловой элемент, обладающий элементарной целостностью, это **предложение** из научного текста (не слово!). *Causal-by-structure*.
- Инструмент - нейронная сеть тензорного произведения (TTNN, 2017) с **направленной** обработкой предложений из текстов на фоне последовательностей из “мешка слов”. *Causal-by-design*.
- Цель - получить модель персонального ассистента, используя вычислительные возможности массовой доступности (без зависимостей от облачных ресурсов, специализированного оборудования и масштабных данных).



## Часть II

# Тензорные нейронные сети

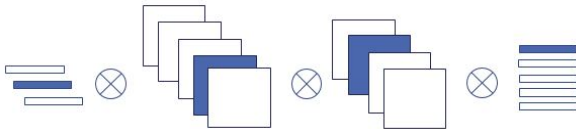
# Симметричные и направленные тензорные декомпозиции

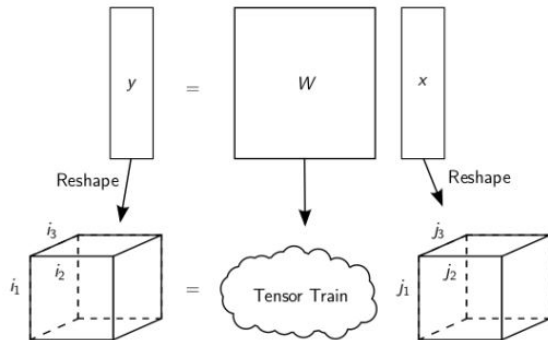
- Симметричные ассоциативные связи в канонической тензорной декомпозиции:

$$\beta \approx B(i_1, i_2, \dots, i_d) = \sum_{m=1}^M U_1(i_1, m) \cdot U_2(i_2, m) \cdot \dots \cdot U_d(i_d, m)$$



- Умножаемые матрицы могут быть любыми: Tensor Train (И. Оселедец, акад. Е.Е. Тыртышников, ИВМ, 2011)



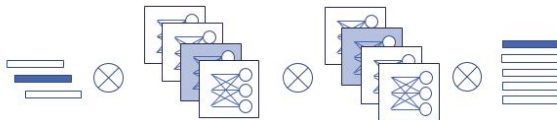


- Тензорное представление больших матриц в обучаемых моделях. А.Новиков, А.Родоманов, А.Осокин, Д.Ветров (ICML 2014), E.Stoudenmire, D.Schwab (NIPS 2016)

- Тензорный поезд является декомпозицией числового тензора (алгебраическая конструкция). Прототип для обучающейся статистической модели.
- Важные обобщения:
  - Во-первых, для наблюдений доступны не все возможные комбинации переменных, многомерный массив данных  $D$  обладает значительной разреженностью.
  - Во-вторых, наблюдения в  $D$  являются случайными, и цель состоит не аппроксимация этих величин, а в оценивании параметров порождающих их распределений.
  - И в-третьих, используемые в декомпозициях многомерных массивов линейно-матричные операции обобщаются на нелинейные нейронные операции, что позволяет описывать сложные взаимодействия между признаками.

# Модель Tensor Train Neural Network

- Для моделирования направленных (упорядоченных) ассоциаций предложена нейронная сеть тензорного поезда (TTNN - tensor train neural network, С. Терехов, Нейроинформатика 2017).



- Обучение большого числа маленьких, динамически формируемых нейронных сетей, слои которых комбинируются для каждого тензорного элемента  $\beta(i_1, i_2, \dots, i_N)$ . Без супер-вычислений.
- Последовательность  $i_1, i_2, \dots, i_N$  может иметь переменную длину и строго упорядочена, что принципиально для моделирования причинно-следственных связей.
- Важнейшее преимущество - открытая система с независимыми структурными элементами.

- Задача формулируется в виде набора контекстных словарей, сочетания элементов которых определяют конкретную ситуацию, оценка которой вычисляется тензорной машиной.
- Алгоритм обучается на основе примеров:
  - **с учителем**, если имеются примеры ситуаций с оценками или измерениями,
  - **без учителя**, когда примеры имевших место ситуаций отличаются от ситуаций со случайными сочетаниями контекстов,
  - **с подкреплением**, с использованием одного из контекстов для кодирования управляющих решений в цепочках (сценариях) из ситуаций.
- ... а также **из строгих логических ситуаций**, которые декларируются как невозможные или достоверные.

# Как это работает?

Тензорная машина пригодна для следующих типовых задач:

- Оценить заданную ситуацию (сочетание контекстов). Сравнить две ситуации.
- Заданы все переменные контекстов, кроме одной, управляющей. Выбрать значение для управляющего сигнала с максимальной ожидаемой оценкой.
  - Выбрать набор альтернативных сигналов управления, с оценками не ниже 95% от максимальной
- Задана ситуация. Какую наиболее высокой оценки можно достигнуть, изменив значения 1 или 2 контекстных переменных?
- В ситуации известны значения большинства контекстов. Найти наиболее вероятные значения для остальных.

- Даны две ситуации - начальная и конечная. Продвигаясь по шагам с изменением не более 1 переменной на шаге, достигнуть конечной ситуации с максимальными оценками на промежуточных шагах.
- Где провести следующий эксперимент?
  - Для заданной ситуации какое из значений управляющей переменной имеет наивысшую оптимистичную верхнюю оценку? (UCB-бандит)
  - Для заданной ситуации выбрать случайное управление, с вероятностью того, что оно оптимально? (Томпсоновские выборки)
- Другие задачи оценивания, поиска, управления, оценки риска и принятия решений.

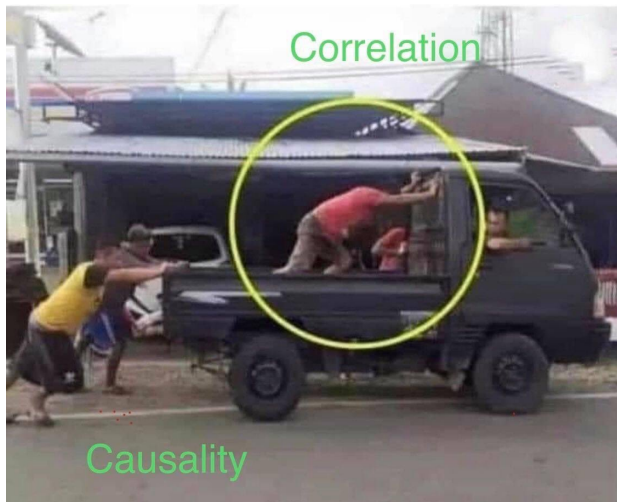


- Социальная функция простейших роботов (при использовании в них тензорных машин) реализуется через передачу ими друг другу причинных (т.е. обученных на множестве сочетаний контекстов из индивидуального опыта в рамках структурных моделей) векторных представлений отдельных символов или целых фрагментов словарей.
- Получив такую порцию информации, робот далее может использовать ее для уточнения своих оценок для контекстов, а также для дальнейшего самообучения (адаптации векторов в новых контекстах).
- Единая нейронная сеть, как конгломерат, не требуется.

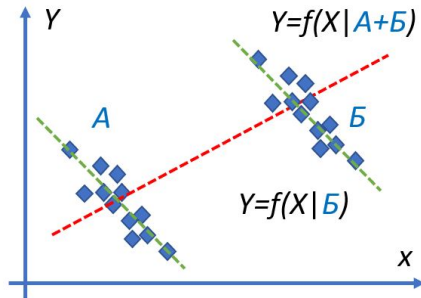
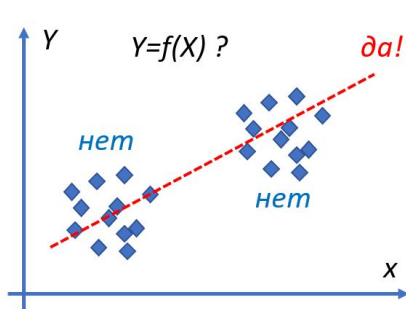
## Часть III

Причины и следствия. Почему это важно?

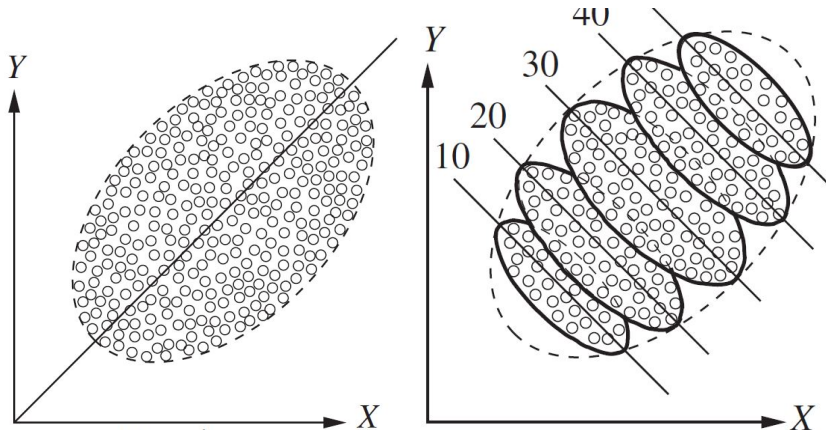
# Корреляция и причина



# «Ложные» корреляции



# Смена тенденции при учёте дополнительных переменных



# Причинно-следственные рассуждения

- Причинно-следственные рассуждения (causal inference) - концептуальный и технический инструмент для ответа на вопросы типа «Какой эффект будет получен от гипотетических (предполагаемых, но ещё не осуществлённых)<sup>1</sup> действий?»
- Если будет понятен чистый эффект от гипотетических воздействий, то мы сможем «обратить» вопрос: какое действие является **причиной** эффекта.
- Это даст формальный язык для анализа причин и следствий.

---

<sup>1</sup>Interventions - «вмешательства» (в реальный мир)

# Об оценке влияний в гипотетических<sup>2</sup> ситуациях

- Контрфактическое утверждение «Если бы **A** было истинным, то **C** могло бы быть истинным». В реальной ситуации **A** не имеет места (ложно). Как оценивать влияния факторов и возможные исходы **C** в гипотетических ситуациях?
- Почему бы просто не ограничиться Байесовыми методами вычисления условных вероятностей?
  - $P(C|\bar{A})$  описывает фактическую ситуацию, вероятность наступления  $C$  при условии не-наступления  $\bar{A}$ , когда вся прочая информация  $K$  зафиксирована.
  - Любые данные и статистики имеются только для фактических ситуаций, наблюдений из «альтернативных миров» нет.
  - Если наступает  $A$ , то необходим детальный учет изменений в контекстах  $K$ . **A** - это решение (воздействие).

---

<sup>2</sup>Контрфактических (counterfactual), J.Pearl,



# Три мыслительных уровня (The Book Of Why)

| Уровень                               | Активность                  | Вопросы                                                                             | Примеры                                                                                                                                                                                 |
|---------------------------------------|-----------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Ассоциации</b><br>$P(y see(x))$    | Смотреть, наблюдать         | Что это? Как наблюдение X изменяет мою уверенность в Y?                             | Что данный симптом говорит мне о возможном заболевании? Что этот анкетный опрос говорит нам о результатах выборов?                                                                      |
| <b>Воздействия</b><br>$P(y do(x), z)$ | Воздействовать, вмешиваться | Что если? Что произойдет при воздействии X?                                         | Если я приму аспирин, вылечит ли это головную боль? Что будет, если запретить сигареты?                                                                                                 |
| <b>Контрфакты</b><br>$P(y x', y')$    | Воображать, ретроспектива   | Почему? Является ли X причиной для Y? Что было бы, если бы я действовал по-другому? | Был ли именно аспирин причиной, по которой ушла головная боль? Стало бы образование лучше, если бы удвоилось число курсов по выбору? Что было бы, если бы я бросил курить 2 года назад? |

Таблица: Ladder of causation - «лестница причинности», J.Pearl

<https://causalai.net/r60.pdf>



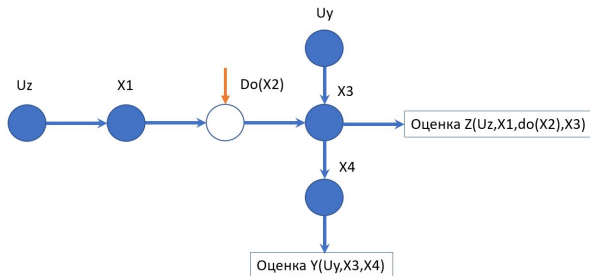
- Структурные причинно-следственные модели<sup>3</sup> - формальный «язык программирования» для описания причинных связей и вероятностных неопределённостей. Эти уравнения (набор инструкций) определяют процесс порождения наблюдаемых данных.
- Процесс начинается с переменных, которым установлены фиксированные значения («входы»), а также с получения реализаций для случайных переменных («неопределённости»).
- В структурных уравнениях используется (несимметричный!) оператор присваивания «:=», отражающий факт проведения вычислений с переменными, значения которых уже известны, и, далее, присваивания результатов другим переменным, по логическим цепочкам.
- При моделировании вычисления повторяются многократно, для различных реализаций случайных переменных, имеющих в модели.

---

<sup>3</sup>Structual causal models, SCM

# Тензорные элементы графов причин и структурных моделей SCM

- Каноническая декомпозиция описывает единичный узел графа с входящими связями  $Z = CP(U_1, U_2, \dots, X_1, X_2, \dots)$ .
- Нейронная сеть TTNN способна представить цепи узлов, а также ветвящиеся структуры с общими узлами.
- Значения дискретных переменных узлов цепи определены как результат  $do()$ -воздействий, а выход TTNN дает оценку цепи или связанных цепей в целом.



## Часть IV

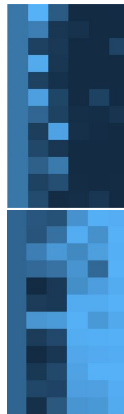
# Причинные представления в текстах

- Ассистент персональных архивов имеет дело с научными текстами, предложения в которых:
  - порождены не статистически, а в процессе цепочек *do()*-решений авторов.
  - структура предложений вторична, первична же специфическая целостность.
  - $\Rightarrow$  *нужен новый подход*.
- Задачи *обратны* задачам, решаемым рекомендательными системами (цель которых - завоевать внимание). Здесь цель иная - сконцентрировать, оградить, освободить от переработки заведомо большого потока (“нейро-ассенизация”, 1999).
- “Принять текст для чтения” - больше похоже на медицинское назначение лекарства или терапии. Такое назначение носит индивидуальный характер, и обязано быть причинным.

- Каждое слово представлено элементом цепи (переменной длины). Выбор слов - это  $do()$ -действия автора текста.
  - Значение переменной слова - вектор активности нейронов слоя нейронной сети - определено контекстной ситуацией.
  - Векторный сигнал о значении слова передаётся последующим словам причинно-структурной модели предложения посредством векторов выходов их нейронных слоёв. Эти выходы отражают причинно-следственные связи между переменными, расположенными ранее в упорядоченном списке всех переменных.
- Предложение оценивается, как целое, суммой выходов  $Z \in [-M, M]$  последнего слоя нейронной сети TTNN. Здесь  $M$  - число нейронов каждого слоя.
- Функция потерь  $L = (M - Z)_+$  для реальных предложений текста, и  $L = (Z + M)_+$  для случайных предложений (bag-of-latest-words) той же длины.

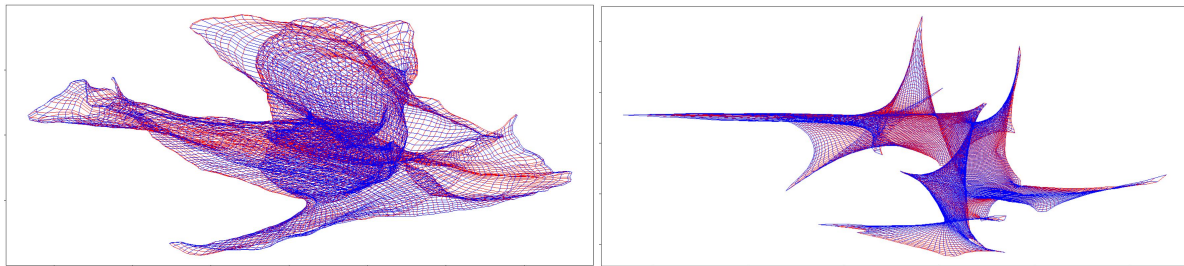
- Обучение предложениям научных текстов не ориентировано на решение какой-либо специфической задачи (классификации, фильтрации, прогнозирования...). Это важно в системах-ассистентах для слабо формализованных постановок.
- Слова представлены двумя неспецифичными “кодами”
  - Статический код - веса и пороги нейронов слоя TTNN. Этот код является структурным, допускает анализ и как единый вектор, и как представления отдельных нейронов (с возможностью их переиспользования).
  - Динамический контекстный код - активности нейронов данного слова в контексте данного предложения. Эти выходы отражают причинно-следственные связи между переменными, расположенными ранее в упорядоченном списке всех переменных.
- Статический код относится к слову, как таковому (обучается как на реальных, так и на случайных предложениях). Динамические коды - проявления слова в его естественных контекстах предложений из научной лексики. Две большие разницы!

- Тензорная система возбуждается входными векторами  $U$ , которые играют роль внешних экзогенных переменных. Для тензорной сети имеется несколько таких векторов, по числу корневых узлов сети.
- Экзогенные переменные могут отвечать внешнему информационному шуму, а также содержательным “модам” функционирования, например, паттерну нейромедиаторов, в контексте которого происходит функционирование нейронной сети
- Переменные  $U$  могут служить также для связи между крупными TTNN-блоками в более глобальной блочной системе.



Активности нейронов (слева направо) для случайного (вверху) и реального предложений.  $M = 12$

# Карты Кохонена вложений из весов нейронов



- Выборочные пары координат вложения векторов нейронных весов,  $M = 10$



- Новые слова - символы возникают из расщепления часто используемых старых обобщённых понятий. Слово копируется, и далее адаптируется в новых контекстах.
- Нейронный слой для каждого слова не имеет информации о происхождении его входов, а также о том, как будут использованы выходы. Это позволяет формировать графовые и гипер-графовые структуры, где целые блоки функционируют равноправно с отдельными словами.
- Устойчивые блоки могут укрупнённо представляться отдельными слоями нейронов, как мета-слова.

- Тензорная нейронная сеть выступает моделью аппарата внутренней функции (априорной, “эмоциональной”?) оценки предложений. Спектр модальностей оценок определяется компонентами входных векторов  $U$ .
- Запрос к ассистенту представляется в виде DAG-графа и причинной структурной модели, элементы которой оцениваются путем как тензорных нейронных вычислений, так и программируемой логики  $do()$ -воздействий.
- Результат также определяется этим графом, в котором уже вычислены оценки для предложений, их последовательностей, и вплоть до фрагментов документов.
- Новые документы “индексируются” совокупностью моделей-запросов, накопленных индивидуальным опытом ассистента.

- Тензорные нейронные сети TTNN представляют широкий спектр возможностей для моделирования как ассоциативных, так и причинно-следственных структур.
- Целостные оценки дискретных последовательностей переменной длины и структурные векторные представления их элементов актуальны для применений в задачах управления и принятия решений.

*Мы находимся на начальном этапе развития причинно-следственных методов искусственного интеллекта в нашу цифровую эпоху (см 2109.00725, 2105.02761).*



[Pearl J, 2009] Pearl J. *Causality*. 2-nd ed. Cambridge University Press, 2009.



[Pearl J, 1988] Pearl J. *Probabilistic reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988, 1997.



[Oseledets I, Tyrtyshnikov E, 2010] Ivan Oseledets, Eugene Tyrtyshnikov. *TT-cross approximation for multidimensional arrays*. Linear Algebra and its Applications 432 (2010) 70–88. [www.mat.uniroma2.it/~tvmscho/papers/Tyrtyshnikov5.pdf](http://www.mat.uniroma2.it/~tvmscho/papers/Tyrtyshnikov5.pdf)



[Kolda T, 2010] Evrim Acar, Daniel M. Dunlavy, Tamara G. Kolda, Morten Morup. *Scalable Tensor Factorizations with Missing Data*. 2010. <http://www.cs.sandia.gov/~dmdunla/publications/AcDuKoMo10.pdf>



[CSE17] *Tensor Decompositions: Applications and Efficient Algorithms* at SIAM CSE17. <http://perso.ens-lyon.fr/bora.ucar/tensors-cse17/index.html>



[Oseledets I, 2011] I.V. Oseledets. *Tensor-Train Decomposition*. SIAM Journal on Scientific Computing, 2011, 33, 2295-2317. <http://epubs.siam.org/doi/abs/10.1137/090752286>



[Feder A et al, 2021] Amir Feder et al. *Causal Inference in Natural Language Processing: Estimation, Prediction, Interpretation and Beyond*. <https://arxiv.org/abs/2109.00725>



[Velickovic P, Blundell C, 2021] Petar Velickovic, Charles Blundell. *Neural Algorithmic Reasoning*. Patterns, <https://arxiv.org/abs/2105.02761>



[Terekhov S, 2020] Serge A. Terekhov. *Tensor train neural networks in retail operations*. In: Advances in Neural Computation, Machine Learning, and Cognitive Research III, pages 17–24, Cham, 2020. Springer.