Оглавление

[Оглавление 1](#_Toc38989509)

[Введение 3](#_Toc38989510)

[Глава 1. Описание и анализ предметной области 4](#_Toc38989511)

[1.1. Онлайн-реклама 4](#_Toc38989512)

[1.2. Онлайн аукцион 4](#_Toc38989513)

[1.3. Машинное обучение 6](#_Toc38989514)

[1.4. Survival analysis 7](#_Toc38989515)

[1.5. Постановка задачи 8](#_Toc38989516)

[Выводы по главе 1 8](#_Toc38989517)

[Глава 2. Теоретическое исследование 9](#_Toc38989518)

[2.1. Анализ существующих решений 9](#_Toc38989519)

[2.1.1. Аналитическое решение 9](#_Toc38989520)

[2.1.2. Решение с использованием решающих деревьев 10](#_Toc38989521)

[2.2. Решение с использованием глубокого обучения. 11](#_Toc38989522)

[2.2.1. Сведение Survival Analysis к задаче предсказания вероятностей 11](#_Toc38989523)

[2.2.2. Описание работы нейронной сети 12](#_Toc38989524)

[2.2.3. Обучение нейронной сети 13](#_Toc38989525)

[2.3. Подходы к улучшению показателей нейронное сети 15](#_Toc38989526)

[2.3.1. Применение механизма внимания 15](#_Toc38989527)

[Выводы по главе 2 15](#_Toc38989528)

[Глава 3. Практическое исследование 16](#_Toc38989529)

[3.1. Реализация чего-то, предложенного в главе 2 16](#_Toc38989530)

[3.2. Сравнение с аналогами 16](#_Toc38989531)

[3.3. Описание внедрения 16](#_Toc38989532)

[Выводы по главе 3 16](#_Toc38989533)

[Заключение 17](#_Toc38989534)

[Библиографический список 18](#_Toc38989535)

[Приложения 19](#_Toc38989536)

[Приложение 1. Пример 19](#_Toc38989537)

[Приложение 2. Еще пример 19](#_Toc38989538)

Введение

# Описание и анализ предметной области

Введем основные понятия и определения из предметной области, которые необходимы для описания постановленной задачи.

## Онлайн-реклама

Определение 1. Онлайн-реклама – это форма маркетинга, которая использует интернет, чтобы доставить маркетинговое сообщения до потенциального покупателя.

(отсюда https://en.wikipedia.org/wiki/Online\_advertising)

Онлайн-рекламу можно увидеть на большинстве интернет-страницах в виде баннера (что приведено на рисунке ниже). Контекст и тематика рекламы чаще всего совпадает со сферой интересов посетителей сайта, чтобы увеличить показатели эффективности объявления.

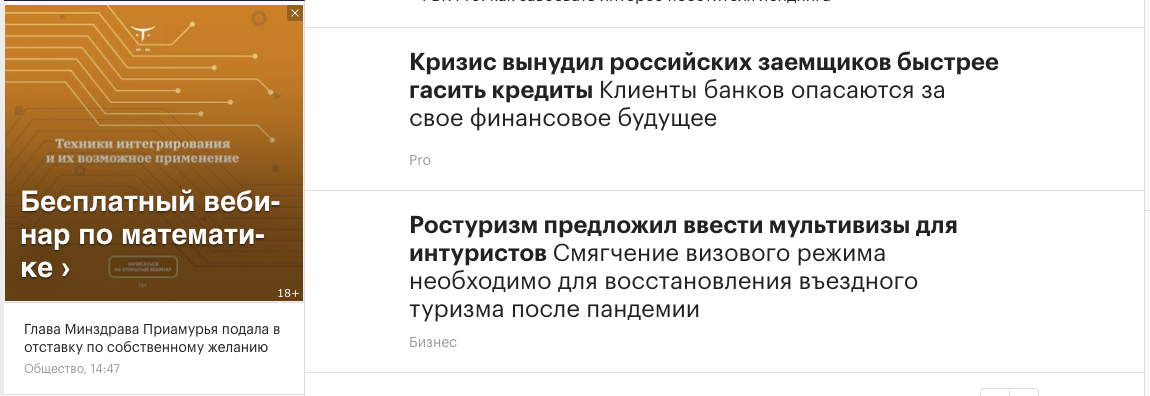


Рисунок 1 – Пример рекламного баннера на интернет-странице rbc.ru

## Онлайн аукцион

Определение 2. Онлайн-аукцион – процесс выбора рекламного объявления для показа пользователю. Кандидаты на показ предлагают свою ставку и выигрывает участник с наибольшей ценой.

Аукционы могут быть открытыми или закрытыми. В закрытом аукционе участники не видят ставку своих оппонентов и не могут изменять свои ставки, в отличии от открытого аукциона, где все участники видят ставки друг друга.

(отсюда https://ru.wikipedia.org/wiki/Аукцион)

Определение 3. Аукцион первой цены – закрытый аукцион, при котором победителем является участник с самой высокой ценой и именно эта цена подлежит уплате.

Определение 4. Аукцион второй цены – закрытый аукцион, при котором победителем является участник с самой высокой ценой, но уплатить он должен «вторую цену», то есть цену своего ближайшего конкурента.

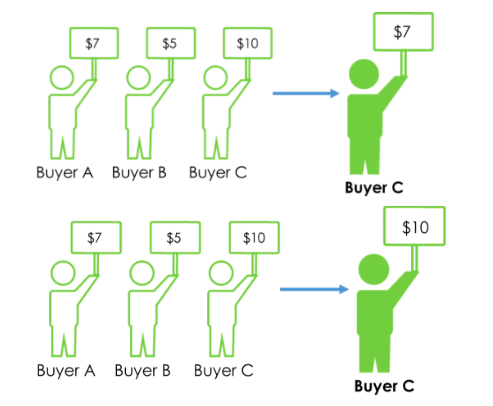


Рисунок 2 – Пример аукциона второй и первой цены

(отсюда <http://www.adnews.com.au/opinion/how-to-better-understand-auction-dynamics-for-video-ad-campaigns>)

Определение 5. Рыночная цена – цена, которую платит победитель аукциона.

Определение 6. RTB (Real Time Bidding) – технология закупки медийной рекламы посредством программируемых онлайн-аукционов.

RTB фокусируется непосредственно на показах целевым посетителям, а не планированию резервов рекламных площадей на определенных сайтах. Каждый показ выкупается за доли секунды – во время загрузки страницы – система. RTB мгновенно проводит аукцион. В результате лучшее предложение от рекламодателей появляется на глазах пользователя, которому оно наиболее интересно.

(отсюда <http://rtb-media.ru/wiki/>)



Рисунок 3 – Механика работы RTB

(отсюда <http://rtb-media.ru/wiki/>)

Ставка в аукционе основывается на информации о пользователе, который заходит на веб-страницу, и потенциальной выгоде, которую он может принести.

О пользователе рекламодатели знают информацию о дате и времени захода на сайт, регион, URL сайта, где будет выставлена реклама, размер рекламного баннера и интересы пользователя (таргеринг). Далее эту информацию будем называть как запрос для ставки.

Выгода от рекламного объявления является позитивным ответом пользователя, показатели CTR и CVR.

Определение. CTR (англ. Click-through rate) – метрика в интернет-маркетинге, которая определяется как отношение числа кликов на баннер или рекламное объявление к числу показов.

(отсюда <https://ru.wikipedia.org/wiki/CTR_(Интернет)>)

Определение. CVR (англ. Conversion rate) – конверсия.

## Машинное обучение

(отсюда https://ru.wikipedia.org/wiki/Обучение\_с\_учителем)

Определение 7. Машинное обучение – подраздел искусственного интеллекта, изучающий обучающиеся методы построения алгоритмов.

Определение 8. Обучение с учителем (англ. Supervised learning) – один из способов машинного обучения, в ходе которого испытуемая система обучается с помощью пары «объект, ответ». Цель обучения в восстановлении зависимости между множеством «объектов» и «ответов».

TODO написать про Нейронные сети, RNN, внимание, кластеризацию, KL-divergence, метрики (точность и т.д.)

## Survival analysis

(отсюда <https://towardsdatascience.com/survival-analysis-part-a-70213df21c2e> и https://ru.wikipedia.org/wiki/Анализ\_выживаемости)

Survival analysis – это класс статических моделей, позволяющий оценить вероятность наступления событий.

Цель анализа в оценке времени, когда произойдет интересующее событие.

Функция выживания (англ. Survival Function) определяет вероятность того, что интересующее событие не произойдет в момент времени .

Функция – плотность распределения вероятности наступления события в момент времени . В Survival analysis называется, как rate of death или failure events per unit time.

Функция условной вероятности определяет, что событие произойдет в рассматриваемый интервал времени, при условии, что оно не произошло до этого интервала.

Определение. Оценка Каплана-Мейера – непараметрическая функция, используется для оценки Функции выживания **вставить номер**.

(отсюда <https://en.wikipedia.org/wiki/Kaplan–Meier_estimator>)

## Постановка задачи

Цель данной работы улучшить предсказание плотности распределения рыночной цены в аукционе второй цены.

Более формально. По запросу (информации о пользователе) получить распределение рыночной цены .

Для решения будет использоваться истории ставок в аукционе, каждый пример можно представить как тройку значений , где – информация о пользователе, – рыночная цена аукциона, b – ставка рекламодателя.

Аукцион закрытый, поэтому рыночная цена известна, только если рекламодатель выиграет аукцион. То есть в случае .

С помощью плотности распределения можно будет рассчитывать вероятность выигрыша и проигрыша аукциона для любой интересующей ставки.

Будем решать сведением задачи к Survival analysis и использованием глубоких нейронных сетей.

Выводы по главе 1

В данной главе были определены основные понятия и определения об онлайн-рекламе и онлайн-аукционе, рассмотрена задача Survival analysis и поставлена задача.

TODO нормально напиши!

# Теоретическое исследование

В первой главы были введены основные определения и сформулирована задача данной работы. В этой главе будут рассмотрены существующие решения этой задачи и новые подходы для улучшения существующих результатов.

## Анализ существующих решений

TODO надо тут тоже норм написать

### Аналитическое решение

Рассмотрим статью 2016 года, которая была опубликована авторами Weinan Zhang, Tianxiong Zhou, Jun Wang, Jian Xu (эта статья <http://wnzhang.net/papers/unbias-kdd.pdf> **надо найти где именно это придумали, там смысл статьи в другом!!**). В этой статье не использовалась информация о пользователе, а учитывалась только ставка рекламодателя и рыночная цена аукциона.

В этой статье решили задачу нахождения вероятности проигрыша/выигрыша аукциона с заданной ставкой с помощью сведения к Survival analysis.

Переведем тройку значений , где – ставка, – рыночная цена, – выигрыш/проигрыш аукциона, в формат . Где – число выигранных аукционов с рыночной ценой равной (по аналогии, интересующее нам событие происходит в день ). – число аукционов, которые не могут быть выиграны со ставкой , то есть число выигранных аукционов с рыночной ценой не менее и число проигранных, в которых ставка была не менее . А саму ставку можно рассматривать как день наблюдения.

Тогда вероятность проигрыша аукциона со ставкой , используя формулу (**номер формы КМ**), будет равна:

Что в Survival Analysis аналогично вероятности того, что интересующее событие не произойдет в интервале от 1 до .

Вероятность выигрыша будет равна:

Плотность распределения рыночной цены:

### Решение с использованием решающих деревьев

Большой недостаток решения, описанного в прошлом разделе, в том, что не используется информация о пользователе. В статье, опубликованной в 2016 году (эта статья <http://apex.sjtu.edu.cn/public/files/members/20160929/functional-bid-lands.pdf>), для это была использована задача кластеризации.

Для получения распределение рыночной цены используется бинарное дерево решений . Главная идея в кластеризации данных, используя информацию о пользователях. Для этого используется метод k-средних и расстояние Кульбака-Лейблера.

При построении дерева в каждом узле будет выбираться разбиение с максимальным расстоянием Кульбака-Лейблера . Математически алгоритм разбиения будет выглядеть следующим образом.

Где и – это вероятность распределения разбиения, – наибольшая рыночная цена, – наибольшее расстояние Кульбака-Лейблера среди всех признаков в -ом узле, - количество признаков, – количество разбиений.

Тогда для предсказание будет выбираться кластер, проходом по дереву, и в листе считаться .Сама плотность распределения считается аналогично статье **(номер статьи)**, также задача сводиться к Survival analysis, только статистические данные считаются для каждого кластера отдельно.

## Решение с использованием глубокого обучения.

Главная цель этой работы состоит в улучшении показателей решения, которого будет описано в этом разделе. (ссылка на общую статью <https://www.aaai.org/ojs/index.php/AAAI/article/view/4407>, ссылка на конкретно про рекламу https://arxiv.org/pdf/1905.03028.pdf). В 2019 году студенты Китайского университета Шанхая опубликовали статью про использование глубоких рекуррентных нейронных сетей в Survival Analysis, и также сделали отдельную публикацию использования этого решения в предсказании рыночной цены в аукционах.

В первой статье были опубликованы показатели применения нового решения в разных областях: в медицине, музыкальной индустрии и предсказании ставок в аукционе. По результатам эксперимента это решение выдала лучшие показатели ROC-кривой (как бинарного классификатора) и ANLP (average negative log probability) среди всех известных на тот момент решений.

### Сведение Survival Analysis к задаче предсказания вероятностей

Во второй статье было описано как задачу Survival Analysis свели к нахождению распределения вероятности рыночной цены и поставленную задачу решили с помощью глубоких нейронных сетей.

Для введения вероятностей выигрыша и проигрыша в аукционе с заданной ставкой была использована формула **(номер формулы)** Survival function.

Чтобы нейронная сеть смогла обучаться предсказывать плотность распределения, необходимо перейти из непрерывного пространства в дискретное.

Для этого представим ставки как последовательность от наименьшей до наибольшей и выберем блоков таких, что для блок .

Все ставки принадлежал множеству натуральных чисел и для лучшего предсказания .

Тогда формулы вероятностей будут иметь следующий вид.

Введем формулу условной вероятности выигрыша в аукционе для ставки . Это значение будет предсказывать нейронная сеть.

### Описание работы нейронной сети

Для предсказания условной вероятности используется – функция рекуррентной нейронной сети, которая приминается пару значений , где – информация о пользователе, а – ставка.

В выходе нейронной сети блоков – условная вероятность выигрыша аукциона для пользователя с информацией и ставкой . Используя значения об условной вероятности, можно найти вероятности выигрыша, проигрыша и плотность распределения рыночной цены.

Формула условной вероятности для и блока . – скрытый вектор из предыдущего рекуррентного блока.

Вероятность проигрыша и выигрыша в аукционе для примера , со ставкой будет равны:

И плотность распределения рыночной цены будет равна:

### Обучение нейронной сети

Для обучения используется две функции ошибки. Первая основана на плотности распределения и ее основная цель минимизировать ANLP для всех выигрышных случаев (обозначим их как ).

Где – номер интервала, в который входит рыночная цена для -ого примера.

Вторая ошибка основывается на функции распределения и разбита на две части. В случае выигрыша аукциона необходимо «занижать» вероятность выигрыша в интервале , и «завышаем» в интервале .

В случае проигрыша рыночная цена не известна, поэтому можно только «занизить» вероятность выигрыша в интервале .

Чтобы использовать обе ошибки вместе введем простую формулу , которая зависит только от ставки и рыночной цены в аукционе.

Тогда вторую функцию ошибки можно расписать как сумму и .

Для обучения модели будет использоваться комбинация двух формул.

Где и гиперпараметры, которые контролируют величину градиента для стабилизации обучения модели.

## Подходы к улучшению показателей нейронное сети

### Применение механизма внимания

Выводы по главе 2

1. Разделы выводов не нумеруются.

# Практическое исследование

Не забудьте переименовать название главы в соответствии с реальным содержанием работы.

## Реализация чего-то, предложенного в главе 2

Для автоматизации построения модели данных требуется разработать формальный подход, позволяющий построить модель данных по реализации алгоритма.

## Сравнение с аналогами

## Описание внедрения

Выводы по главе 3

1. Разделы выводов не нумеруются.

Заключение

Разделы «Введение» и «Заключение» не нумеруются.

Библиографический список

1. Ссылка 1
2. Ссылка 2

Приложения

Если приложение одно, то оно включается в раздел «Приложение» без подразделов.

Приложение 1. Пример

Если приложений несколько, то для каждого из них делается отдельный раздел.

Приложение 2. Еще пример

Вот так.