

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

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Исследовательский проект на тему:

Сжатие словарей для нейросетевого анализа исходных кодов программ

Выполнил студент:

группы #БПМИ171, 4 курса

Гусев Андрей Алексеевич

Принял руководитель ВКР:

Чиркова Надежда Александровна

Научный сотрудник

Факультет компьютерных наук НИУ ВШЭ

Москва 2024

Содержание

Аннотация	3
1 Примеры	7
1.1 Ссылки на статьи	7
1.2 Рисунки	7
1.3 Таблицы	8
1.4 Формулы	8
А Пример секции аппендикса	10

Аннотация

Ваша аннотация на русском языке.

Ключевые слова

Глубинное обучение, разреживание моделей, рекуррентные нейронные сети

Введение

В современном мире маркетплейсы стали неотъемлемой частью электронной коммерции, предоставляя платформы для продажи товаров и услуг различным производителям и ритейлерам. В 2023 году маркетплейсы продолжили быть главной движущей силой российской онлайн-торговли. Рост объема трат на маркетплейсах в 1,5 раза по сравнению с предыдущим годом свидетельствует о том, что интерес потребителей к онлайн-покупкам только укрепляется. На это влияют общерыночные факторы: продолжают развиваться альтернативные каналы поставок продукции ушедших брендов, улучшаются условия доставки, повышается удобство пользования платформами, расширяется сеть пунктов выдачи.

Согласно исследованиям Tinkoff eCommerce, количество транзакций на маркетплейсах за год выросло на 63% (см. рисунок 0.1). Лидерами по росту количества покупок стали Мегамаркет (число транзакций выросло в 4,3 раза), Wildberries (в 2 раза) и Ozon (в 1,6 раза).



Рис. 0.1: Динамика покупок на маркетплейсах в регионах.

Появился тренд на рост популярности маркетплейсов в российских регионах. В 2023 году жители российских городов стали значительно активнее совершать покупки на маркетплейсах: выросло как количество транзакций на онлайн-площадках, так и их сумма (см. рисунок 0.2). По количеству совершенных транзакций особенно заметен рост в таких городах, как Омск (+91%), Красноярск (+88%), Новосибирск (+79%), Челябинск (+79%) и Волгоград (+75%). В Москве зафиксирован наименьший прирост числа транзакций (+41%). Увеличение интереса жителей регионов к маркетплейсам объясняется рядом причин: расширением географии присутствия площадок, развитием сетей ПВЗ и логистических сервисов,

улучшением условий доставки.



Рис. 0.2: Динамика покупок на маркетплейсах в регионах.

Вместе с тем выросло количество селлеров на 8%. Рынок становится более зрелым: место неопытных продавцов занимают более профессиональные. Они ведут бизнес более уверенно, укрепляют свои позиции на площадках и торгуют на нескольких платформах одновременно. Количество селлеров, ведущих торговлю на двух и более маркетплейсах, за год увеличилось на 17%. Самой привлекательной платформой для старта бизнеса является Wildberries: 63% продавцов в конце 2023 года выбирали ее в качестве первой площадки (см. рисунок 0.3).

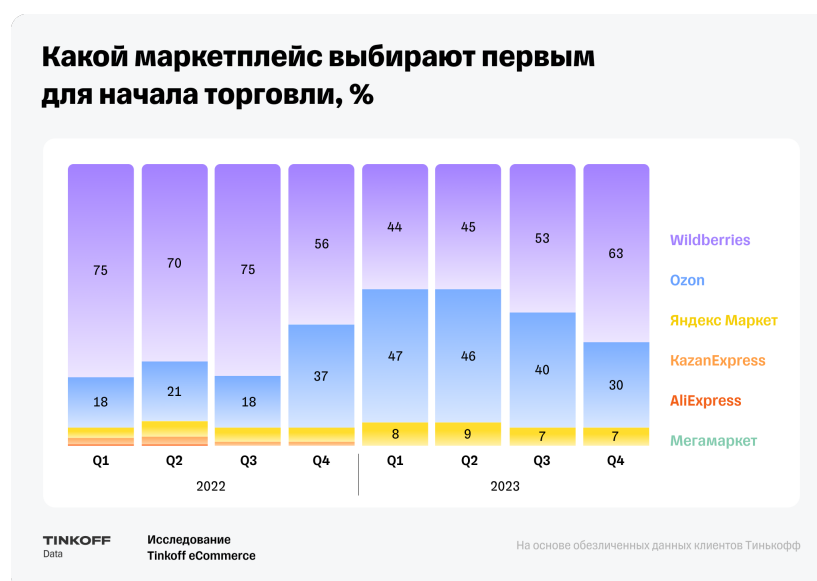


Рис. 0.3: Популярность маркетплейсов за 2023 год.

С увеличением конкуренции продавцы все чаще обращаются к системам, которые могут помочь им в продажах, а также автоматизировать процесс работы. Одним из ключевых

факторов успешной продажи становится эффективная SEO-оптимизация карточек товаров. Подбор наиболее подходящей категории и создание продаваемого описания, содержащего ключевые слова, позволяет улучшить видимость товаров в результатах поиска как на самом маркетплейсе, так и в поисковых системах, что напрямую влияет на увеличение продаж.

На данный момент существуют несколько сервисов (например, TurboTextPro, CopyMonkey, Gerwin), позволяющих сгенерировать описание товаров по характеристикам, ключевым словам или фото. Однако, качество сгенерированных описаний не всегда позволяют использовать их в системах автономного управления. В настоящее время российский рынок не предлагает специальных технологий и решений для подбора наиболее подходящей категории товара на маркетплейсе. Таким образом, задача создания качественного инструмента для эффективной SEO-оптимизации карточек товаров является актуальной.

Цель данной работы — разработать сервис, в основе которого будет реализован функционал, способный классифицировать товары на маркетплейсе на основе их фотографий и генерировать соответствующие к ним описания. Сервис реализуется для маркетплейса Wildberries, так как он представляют наибольшую популярность у продавцов и покупателей. В качестве дальнейшего развития проекта можно рассмотреть другие площадки.

Для достижения поставленной цели необходимо решить следующие задачи:

- подготовить набор данных, содержащий изображения товаров и соответствующие им категории и текстовые описания;
- провести исследование текущих архитектур нейронных сетей, используемых для классификации изображений и генерации текста, и на основе этого исследования выбрать наиболее подходящую архитектуру или их комбинацию для решения поставленной задачи;
- обучить выбранную нейронную сеть на подготовленных данных, оптимизировать и настроить параметры модели для повышения её производительности и качества результатов, а затем оценить эффективность реализованной архитектуры нейронной сети;
- интегрировать модель в программное обеспечение.

1 Примеры

1.1 Ссылки на статьи

Ссылки на статьи оформляются с помощью пакета `biblatex`, например `[chirkova18]`. В описании статье в `bib` файле нужно обязательно указывать место публикации работы (журнал или конференцию) и год. Обратите внимание, что для описания статей из разных источников в списке литературы используются разные команды в `bib` файле: статья из журнала `[ctan]`, статья с конференции `[chirkova18]`, книга `[knuth-acp]`, глава книги `[knuth-fa]`. Если статья еще не опубликована нигде, а только выложена на arXiv, то на нее тоже можно сослаться `[chirkova18_arxiv]`, но предпочтительно сослаться на опубликованную версию, если она уже существует. Если вы хотите сослаться на сайт, то можно либо так же внести его в список литературы `[knuthwebsite]` (рекомендуется, если таких ссылок у вас много из-за особенностей темы вашего проекта), либо использовать ссылку внизу страницы¹. При работе с онлайн ресурсами не забывайте указывать дату обращения к этому ресурсу, так как в отличие от опубликованных статей, эти ресурсы могут измениться в любой момент.

1.2 Рисунки

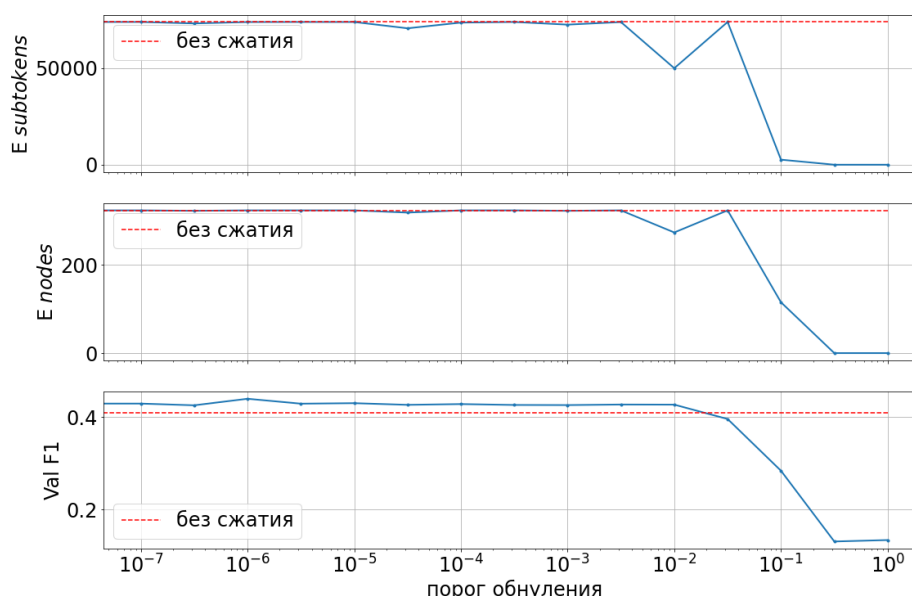


Рис. 1.1: Пример графика. Тут должна быть подпись, поясняющая что происходит на рисунке (краткая, но достаточная для понимания основной идеи графика).

¹Книги доступны по ссылке: <http://www-cs-faculty.stanford.edu/~uno/abcde.html>, дата обр. 16.05.2013

Все рисунки в тексте должны иметь подписи и вы на них должны ссылаться в тексте. Например, на Рисунке 1.1 изображен пример графика. Не забывайте подписывать все оси на графиках, добавлять легенду и пояснять все обозначения, а также используйте адекватного размера шрифты и толщину линий на графиках (все должно быть видно и понятно без многократного увеличения). На рисунке из примера явно не хватает обозначения синей линии в легенде.

1.3 Таблицы

Все таблицы в тексте тоже должны иметь подписи и вы на них должны ссылаться в тексте. Например, в Таблице 1.1 показаны результаты примерного эксперимента.

Таблица 1.1: Пример таблички. Тут должна быть подпись, поясняющая что происходит в таблице (краткая, но по делу).

	Val			Test			nodes	subtokens
	Prec	Rec	F1	Prec	Rec	F1		
запуск 1	0.4894	0.3775	0.4263	0.4824	0.3683	0.4177	10029	179
запуск 2	0.4887	0.3739	0.4237	0.4891	0.3724	0.4228	10039	177
запуск 3	0.4820	0.3751	0.4219	0.4838	0.3677	0.4178	10037	180
среднее	0.4867	0.3755	0.4239	0.4851	0.3695	0.4195		
дисперсия	0.0041	0.0019	0.0022	0.0036	0.0025	0.0029		

1.4 Формулы

Формулы стоит центрировать, а также нумеровать, если вы ссылаете на них в тексте. Также не забывайте пояснять все обозначения в формулах. Например, запишем следующую задачу оптимизации:

$$\theta^* = \min_{\theta} F(\theta), \quad (1)$$

где F – квадратичная функция от параметра θ . При необходимости, далее в тексте можно сослаться на формулу (1). При этом, в зависимости от конкретных формул, можно использовать разные слова: формула, уравнение, задача оптимизации и т.п.

Список литературы

- [1] Nadezhda Chirkova, Ekaterina Lobacheva, Dmitry Vetrov. Bayesian Compression for Natural Language Processing. In EMNLP 2018.

А Пример секции аппендикса