

Грибов Илья

ML-ENGINEER · DATA SCIENCE

☎ (+7) (985) 141-69-39 | ✉ mathandmlexplorer@yandex.com | 🌐 t.me/ai_data_enjoyer | 📺 UniverselsMyCreation

Опыт Работы

М-Холдинг

Москва

ML-ENGINEER

28 Июл. 2023 - 20 Авг. 2023

- **Задача:** сопоставить разные описания с разных магазинов одного и того же товара
- **Технологии:** PyTorch, Translation api, Git, Sql, Docker, Pandas, Numpy, ClickHouse, Scikit-learn, Imblearn, TSNE, Hyperopt, re, nltk, etc
- **1 этап: Подготовка данных:** Был реализованы и протестированы различные способы перевода текста с английского на русский (изначально все описания была на английском): создание словаря уникальных слов и дальнейших их перевод с помощью api переводчиков, использование и fine-tuning различных seq2seq моделей глубокого обучения, а также готовых трансформеров с HuggingFace. Кроме того была проведена первоначальная предобработка текста с помощью библиотек re, nltk, и так далее
- **2 этап: Получение качественных эмбеддингов:** Используя предобученный BERT на задачи максимизации, loss функции метрического обучения (Triplet Loss, ArcFace Loss, Contractive Loss, SoftMax Loss) и kNN, была получена такая модель, которая хорошо группировала выходные фичи описаний товаров по классам в некотором признаковом пространстве
- **3 этап: Постобработка:** После получения N наиболее похожих описаний товаров на заданное описание необходимо было наилучшим образом сматчить два похожих описания. На данном этапе была обучена ранжирующая модель, а также использовались различные метрики и регуляризаторы (BLEU, Rouge, и так далее) для решения данной подзадачи

МГУ им. М.В.Ломоносова

Москва

NLP RESEARCHER

1 Июл. 2023 - 20 Июл. 2023

- Исследовались различные причины и проблемы в сфере интерпретируемости нейронных сетей в задаче обработки естественного языка и способы их решения
- Исследовались методы построения качественных состязательных примеров для текстовых данных, которые можно было бы использовать, например, для аугментации данных или исследования поведения нейронных сетей

Учебные проекты

RecSys approaches

Реализация и исследование основных подходов рекомендательных систем

- Коллаборативная фильтрация
- ALS и deep learning
- user2user, item2item, user2item
- Матричная факторизация

Реализация алгоритма решающего дерева

Реализация алгоритма решающего дерева на языке C++

- Был реализован и протестирован алгоритм решающего дерева

Hard Skills

Linux/Terminal, ssh, scp, grep, ls, cd, pwd, wget, pip, etc

C/C++, синтаксис, ООП, системные вызовы

Python, синтаксис, asyncio, ООП, декораторы, метаклассы

Git, знание основного синтаксиса

SQL, join, select, where, over, groupby, having, etc

Computer Vision, pytorch, tensorflow, opencv, torchvision, hugging face, albumentations, etc

NLP, pytorch, tensorflow, torchvocab, natasha, hugging face, nltk, re, etc

Machine Learning, scikit-learn, hyperopt, imblearn, numpy, pandas, catboost, etc

Docker, beginner

Algorithms and data structures, two pointers, hashmap, stack, bin search, sortings, dfs, trees and graphs, etc

Maths, math statistics, linear algebra, math analysis, probability theory, etc

English, Upper-Intermediate - Advanced (B2-C1)

Germany, Elementary (A1)

Образование

МГУ им. М.В.Ломоносова

Бакалавр; Средний балл: 4.94

Москва

Май 2020 - Июнь 2024

- Степень бакалавра по программе "Прикладная математика и информатика" (Ситсемное программирование)

Дополнительное образование

МФТИ Deep Learning School

Москва

Участие в школе глубокого обучения по CV и NLP

Янв. 2023 - Авг. 2023

- Изучались способы решения, подходы и архитектуры основных задач компьютерного зрения: Сегментация, Детекция, Классификация, Генерация, Вариационное автокодирование изображений
- Изучались способы решения, подходы и архитектуры основных задач обработки естественного языка: Суммаризация, Языковое моделирование, Классификация, NER, TF-IDF, и так далее текстов, - а также основные способы обработки и парсинга текстов: re, spacy, nltk, torchvocab, natasha, pymorphy2, и так далее

Курс ШАД по компьютерному зрению

Москва

Прохождение курса ШАД по cv на ВМК

Дек. 2023 - Сеп. 2023

- Изучались способы решения, подходы и архитектуры основных задач компьютерного зрения: Сегментация, Детекция, Классификация, Генерация, Вариационное автокодирование изображений, а также Motion Detection, Metric Learning и ViT (Vision in Transformers)
- Изучались классические способы решения и подходы задач компьютерного зрения, основанные на классическом машинном обучении (SVM, Decision Tree, etc) и генерации фичей (HOG, Ransac, etc), а также различные классические свертки и способы обработки изображений

Machine Learning

Москва

Прохождение основного курса ВМК по машинному обучению

Сеп. 2022 - Май 2023

- Изучались классические алгоритмы машинного обучения (SVM, Gradient Boosting, AdaBoost), метрики (MAPE, SMAPE, RMSE, ROC AUC, etc) и задачи (PCA, TSNE, Active Learning, ImbLearning, SemiLearning, etc)

Соревнование и Хакатоны

VL TLS-emb

Москва

соревнование по построению качественного TLS эмбединга

Май 2023

- В процессе соревнования использовались основные подходы для токенизации текста (fasttext, nltk, etc) и обучения векторных представлений (word2vec, bert, rnn, cnn, etc)

Kaggle

Москва

Участие в соревнованиях на KAGGLE

2023

- Участие в соревновании на Kaggle по предсказанию цены поездки такси, в котором активно использовался EDA (exploratory data analysis), и грамотный feature engineering
- Участие в соревновании на Kaggle по предсказанию активности Steam, в котором активно использовался feature engineering и Stacking алгоритмов машинного обучения