



КУМΠΑН ВИКТОР ВИКТОРОВИЧ

Junior/middle ML Engineer/ Data science/ Software engineering

НАВЫКИ

- Владение английским B2 (Upper Intermediate)
- Математический и функциональный анализ, дискретная математика.
- DevOps - CI/CD. Опыт работы с Docker, GIT.
- Алгоритмы и структуры данных (STL)
- Python, C++, ООП,
- PyTorch, TensorFlow, NumPy, cv2, genism
- Pandas, Matplotlib, Seaborn
- SQL и опыт проектирование БД
- AWS, Azure, OpenStack.
- Умение стремительно развиваться в высококвалифицированном окружении
- Понимание принципов ведения проектов Kanban, Scrum, Agile.

ОБРАЗОВАНИЕ

НИУ ИТМО Г. САНКТ-ПЕТЕРБУРГ • 2019-2023 БАКАЛАВРИАТ

Факультет информационных технологий и программирования ФИТиП, направление подготовки прикладная математика и программирование. Средний балл 4.9. В рейтинге студентов в 5 лучших (за весь период обучения). **Первое место** на **Всероссийском конгрессе** молодых ученых.

ОПЫТ ПРОДУКТОВОЙ И ПРОЕКТНОЙ РАБОТЫ

SAMSUNG COMPUTER VISION ENGINEER • IMAGE PROCESSING TEAM

Задачи: Оптимизация процесса тренировки и инференса моделей (Pipeline). Улучшение SOTA CV моделей (VAE, GAN, AE). Пре и пост обработка данных. Разработка функций оптимизации улучшающие продуктовые метрики.

Результат:

- Решены задачи с оптимизацией pipeline – тренировка ускорена на 40%. Добавлен функционал для валидации моделей и сделан более юзабельный pipeline.
- Выведены функции оптимизации улучшающие продуктовые метрики и их применение подавило артефакты на изображениях.
- Выявлены артефакты в данных, не позволяющие обучать модель без дефектов. Помогло подавить большинство дефектов.
- Кастомизированы SOTA CV модели (VAE, GAN, AE) – улучшило продуктовые метрики.

YANDEXCUP COMPUTER VISION • СОПОСТАВЛЕНИЕ ИЗОБРАЖЕНИЙ К НАИБОЛЕЕ РЕЛЕВАНТНОМУ ТЕКСТОВОМУ ОПИСАНИЮ / ZERO-SHOT.

В рамках турнира предоставлен датасет из ссылок на изображения (5 млн) и соответствующим им поисковым запросам(>20млн). **Реализовано простое MLOps решение для:**

- Многопоточного скачивания, обработки(аугментации) данных, анализа текстовых данных, кластеризации текстовых описаний и визуализации embeddings
- Обучения моделей I2T zero-shot classification, создание моделей кластеризации (Train: BoW -> SentenceTransformer -> AgglomerativeClustering -> KNeighborsClassifier; Inf: SentenceTransformer -> KNeighborsClassifier)
- Оценка качества моделей, логирование (tensorboard)



@SKYLARPRO



SKYLARPRO



+79508998389



ITMO

○ Inference моделей

За baseline взят CLIP by Sber (visual_encoder – VIT-B32; text_encoder – GPT3Small) так как он влез в ограничения по весу модели. **Начальный результат accuracy на private 42%** (30 классов), пробовал заменить GPT3Small -> RuBert ; GPT3Small -> BPE embeddings (предоставленные организаторами) + 4 transformers block. **Финальным решением** было приделать к baseline параллельный классификатор [1, N], класс которого соответствовал номеру projections стоявшего на выходе text_encoder (тренировка End-to-End). Это позволило поднять точность на семантически близких описаниях. **Финальный accuracy на private 58%** (30 классов).

СТАРТАП (БИЗНЕС КЕЙС) • “ПРОДВИНУТЫЙ АНАЛИЗ ТОНАЛЬНОСТИ” ПОСТОВ В СОЦИАЛЬНЫХ СЕТЯХ

Задачи и реализация:

- Поставлена бизнес-задача, выделены ключевые особенности решений и сопоставлены бизнес и ML метрики. Коммуникация с инвесторами.
- Размечен датасет совместно с экспертами. Произведен preprocessing, анализ данных
- Классический ML: Gradient Boosting, Bayes classifier с различными vectorizers TF-IDF, HashVect, word2vec. Результат **f-мера 75 %**.
- В качестве DL baseline взята архитектуры BERT, GPTs. Проблема Малых данных решена аугментацией и подбором pre-train моделей для смежных задач. Результат **f-мера 85%**.
- Проведение A/B тестирований, оценка гипотез, production.

AI JOURNEY 16 МЕСТО (БРОНЗА) • AI TRAIN ДЕТЕКЦИЯ, СЕГМЕНТАЦИЯ ОБЪЕКТОВ НА ЖД-ИНФРАСТРУКТУРЕ

В рамках турнира предоставлен датасет из 11 классов детекции и 3-х классов для сегментации.

- Проведен статистический анализ данных. Для детекции объектов был взят YOLOV5 (bb – EfficientNet; Neck – Bi-FPN; Head - Conv), выполнена легкая аугментация данных, так как некоторые классы оказались чувствительны.
- Выполнена тонкая настройка NN (freeze bb) и подбор гиперпараметров. В начале тренировок loss взял CE, далее перешел к FL, это помогло поднять f-score для отстающих классов. Для сегментации использовал в качестве encoder ResNet-18, decoder – Unet тренировал с помощью CE, но после оптимизации перешел на JaccardLoss.



- **Качество** оценивалось как $\text{competition_metric} = 0.7 * \text{mAP@.5} + 0.3 * \text{meanIoU}$. на **baseline** было **0.41**, удалось **повысить до 0.47**.

НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА • ВЛИЯНИЕ ADVERSARIAL ATTACKS НА СНИМКИ МРТ МОЗГА

Работа была проведена на датасете Brats (3d снимки мозга), выполнен анализ данных, за основу была взята CNN ResNet-18.

- Произведена кастомизация архитектуры под 3d изображения с применением transfer learning. Выполнена тренировка CNN с различными функциями активации ReLu и BReLu и ее тонкая настройка.
- Для проверки гипотез была произведена аугментация данных, тренировка CNN на FGSM атаках. **Поднять F-меру до 80%**.

НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА • СЕГМЕНТАЦИЯ ДЕФЕКТОВ СТАЛИ

Работа проведена на датасете "Северсталь".

- За основу был взят U-net с encoder'ом Res-Net 18, выполнена тонкая настройка NN. Применены методы борьбы с overfitting. Выполнена визуализация работы NN.
- Произведен анализ датасета на возможные статистики в данных. Выполнена аугментация данных и их предобработка.
- **Качество** модели оценивалось IoU **baseline - 0.5** **поднял до 0.75** на валидации.

ПРИКЛАДНОЕ ПРОГРАММИРОВАНИЕ • C ++

1. Реализация шаблонного вектора со всеми основными операциями и с динамическим освобождением памяти.
2. Реализация нагруженного дерева (trie, бор) с упором на время выполнения и потребляемую память.
3. Реализация игры морской бой на основе принципа ООП и написание умного бота для игры. Виртуализация.
4. Реализация умной телефонной книги на основе STL, оптимизированной по времени работы и по памяти.
5. Реализация решателя судоки (количество решений, решение) оптимизировано по времени и по памяти.

ПРИКЛАДНОЕ ПРОГРАММИРОВАНИЕ • PYTHON

1. Первое место на Всероссийском конгрессе молодых ученых. Софт для автоматизированного поиска в открытых научных источниках информации для исследования предметной области.



2. Реализация игры жизнь, на основе ООП, работа с GUI и пользовательским интерфейсом.
3. Реализация системы контроля версии с основными командами и принципами работы.
4. Реализация пула процессов с управлением памятью и рабочими, если память во время вычислений выходит за пределы выделенной, то этот расчет отправляется и обрабатывается во втором раунде.
5. Реализация веб-сервиса "записной книжки" на Django, с реализацией REST API и упаковкой в Docker контейнер и покрытием Unit tests .

МАТЕМАТИЧЕСКИЕ МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ • PYTHON

1. Реализация **DecisionTree** с возможность выбора основных гиперпараметров, на основе этого дерева реализация **RandomForest** с возможность выбора основных гиперпараметров.
2. Реализация **Polynomial regression, Linear regression** с возможность как аналитического решения, так и итерационного. Визуализация решений и применение на реальных данных.
3. Реализация нейрона с различными функциями активации Sigmoid(LogLoss), ReLu, LeakyReLu, ELU (MSELoss).
4. Реализация **Naive Bayes Classifier**.
5. Работа с основными библиотеками Pandas, NumPy, PyTorch.

