

Абакумов А.В., Еремеев С.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264 г. Муром, Владимирской обл., ул. Орловская, д. 23
artem210966@yandex.ru*

Классификация и кластеризация изображений на основе топологической декомпозиции

Персистентная топология [1] позволяет извлечь из изображения не только топологические свойства, полезные при обработки изображений, но и для классификации и кластеризации.

Изображение разбивается на компоненты, которые можно упорядочить в виде дерева по порядку объединения. Итоговая структура, топологическое дерево, можно использовать в том числе для сравнения изображений. Так как каждая компонента — это отдельная область на изображении, сравнение по дереву позволит выявить точенные изменения, когда изображения слабо изменяется. Например, таким свойствам обладают скриншоты приложений. Их распознавание и классификация используется в задаче автоматического тестирования графического интерфейса [2]. Другая схожая задача – документирование. Каждый активный элемент на скриншоте репрезентуется компонентой связности из топологического дерева, что даёт возможность сегментировать изображение.

Компоненты связности дают по-новому взглянуть на многие классические алгоритмы и методы. Обычно они применяется ко всему изображению. Но, если рассматривать компоненты как независимые зоны, возможно их применение на каждую такую зону отдельно, или только на часть зон (т.е. с фильтрацией). Подобное внедрение компонент в процесс обработки даёт возможность исследовать новые свойства, такие, как ресурс пикселя [3], который отображает «важность» конкретной точки, что позволяет выполнять операции, схожие с детектором границ Канни.

Классифицировать и кластеризировать можно как объекты на одном изображении, так и множество изображений.

В первом случае создаются компоненты связности, а затем они уже классифицируются по топологическим свойствам. Сравнение двух компонент возможно либо эксклюзивно (без дочерних компонент), либо инклюзивно (с дочерними компонентами). Например, при сравнении двух домов алгоритм рассматривает либо только компоненту самого дома, либо включает также компоненты всех элементов дома.

В задачах классификации и кластеризации часто прибегают к понятию «расстояния» между объектами – численному показателю, означающему насколько они похожи. Топологическое сравнение изображений подразумевает сравнение всех компонент, но даже в изображении размером 256 на 256 пикселях может быть 2000-3000 компонент, что приводит к долгому времени выполнения, особенно, когда количество изображений велико. Для ускорения можно применять фильтрацию, но это может привести к потере точности. Вместо этого существует более топологически-надёжный метод – перевод обычных компонент связности в Гильбертово пространство [4], что позволит получить персистентные ландшафты (persistence landscape). Они добавляют новые свойства, позволяющие корректно сравнивать их, а также сократить количество компонент [5], т.е. ускорить обработку. Минусом является сложность подбора функции расстояния. Эксперименты показали, что лучше всего подходит норма по бесконечности с кластеризацией через k-means. При малом числе кластеров (2-3) точность достигает 95%, но с

большим числом классов точность может деградировать. Дополнительным фактором является топология самих изображений. Если разные классы топологически схожи, то алгоритм не сможет их разделить.

Литература

1. Robert Ghrist. Barcodes: The Persistent Topology Of Data. 2008. DOI: 10.1090/S0273-0979-07-01191-3
2. Abakumov A.V., Eremeev S.V. Segmentation of Graphical User Interface Elements Based on Topological Decomposition for GUI Testing Tasks. Data Analytics and Management in Data Intensive Domains. Vol 2086. DOI: 10.1007/978-3-031-67826-4_14
3. Абакумов А.В. Обработка изображений на основе клеточных автоматов и компонент связности. Радиотехнические и телекоммуникационные системы. 2024. № 1 (53). С. 61-68.
4. Чуканов С.Н. Формирование признаков на основе методов вычислительной топологии. Компьютерная оптика. 2023. Т. 47, № 3. С. 482-490. DOI: 10.18287/2412-6179-СО-1190.
5. Абакумов А.В., Еремеев С.В. Оптимизация декомпозиции изображений по топологическим признакам с использованием персистентных ландшафтов. 15-я Международная конференция «Интеллектуализация обработки информации» ИОИ-2024. 2024. Гродно, Беларусь.

Абрамова Е.С., Абрамов О.А., Орлов А.А.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264 г. Муром, Владимирской обл., ул. Орловская, д. 23
elena.tarantova@yandex.ru*

Исследование инкрементной машины экстремального обучения

Машина экстремального обучения (Extreme Learning Machine, ELM) представляет собой алгоритм обучения нейронной сети с одним скрытым слоем, характеризующийся высокой скоростью обучения и простотой реализации. ELM решает задачи регрессии и классификации за счет случайной инициализации весов входного слоя и вычисления весов выходного слоя с использованием аналитического решения через псевдообратную матрицу [1]. Однако классическая версия ELM не поддерживает инкрементное обучение, поскольку требует полной переобучения на всем наборе данных при поступлении новых данных.

Инкрементная машина экстремального обучения (Incremental Extreme Learning Machine, IELM) решает эту проблему, позволяя обновлять модель в условиях поступления новых данных без необходимости полного переобучения [2]. Это достигается за счет применения формулы Шермана-Моррисона-Вудбери для эффективного вычисления обновленной псевдообратной матрицы.

Для исследования инкрементного обучения была создана синтетическая выборка, моделирующая задачи регрессии. Для имитации сценария инкрементного обучения на модель последовательно подавались небольшие партии новых данных, состоящие из 10 объектов за одну итерацию обновления. Это позволило проанализировать, как модель адаптируется к новым данным и как изменяется качество ее предсказаний со временем.

Для оценки качества инкрементного обучения были использованы следующие параметры:

- Среднеквадратичная ошибка (MSE). MSE измерялась на тестовых данных до и после каждого инкрементного обновления, что позволило отслеживать динамику качества предсказаний.
- Прирост ошибки (Δ MSE). Δ MSE определялся как разница между MSE после и до очередного обновления. Это позволило анализировать, улучшается ли модель на каждом шаге или начинает деградировать.
- Изменение весов модели. Вычислялась норма разницы между весами выходного слоя до и после обновления для оценки степени влияния новых данных на модель.
- Время обновления. Измерялось для каждой итерации инкрементного обновления, чтобы оценить вычислительную эффективность по сравнению с полным переобучением.
- Стабильность модели. Анализовалась устойчивость MSE на протяжении нескольких итераций, чтобы выявить признаки переобучения или деградации качества.

Модель продемонстрировала стабильность на начальных этапах инкрементного обучения; однако с увеличением отклонения новых данных от распределения старых наблюдались значительные колебания в значениях MSE.

Переобучение проявлялось в виде резкого роста MSE, что подтверждалось значительными изменениями в весах. Это указывает на то, что модель начинает терять информацию о старых данных и адаптируется к новым, шумным данным.

Норма изменения весов, рассчитываемая как евклидова норма вектора разности между весами модели до и после обновления, отражает степень адаптации весов в процессе обучения.

Значительное изменение весов свидетельствует о неоптимальной адаптации к новым данным.1830. Т. 4, 5, 6, 7, 8, 12, 18, 19, 20.

Асанов А.М.

*Федеральное государственное автономное образовательное учреждение высшего образования «Российский университет дружбы народов имени Патриса Лумумбы»
117198, г. Москва, ул. Миклухо-Маклая, 6
asanov.acxat@gmail.com*

Классификация текстурных изображений макропористых криогелей

Классификация текстурных изображений является важной задачей в ряде областей науки и промышленности, таких как металлургия (оценка микроструктур), медицина (анализ тканей), сельское хозяйство (контроль структуры почвы) и других отраслях. В данной работе рассматривается задача классификации изображений макропористых криогелей на основе поливинилового спирта (ПВС), приготовленных с различными добавками и температурными условиями [1].

Термин «текстура» в изображениях часто определяется как присутствие повторяющихся базовых элементов с различными ориентациями в пространстве, при этом эти элементы обладают недостаточно выраженными контурами или границами. Она несёт информацию о поверхностных свойствах объектов, таких как пористость, шероховатость, регулярность структуры. Текстуры могут быть однородными или сложными, а их анализ требует применения специальных методов, позволяющих учитывать мелкомасштабные особенности изображения.

Обучение нейронных сетей на датасетах текстурных изображений требует применения специализированных архитектур. Отличительные характеристики текстур, такие как низкий контраст, неоднородное распределение яркости и акцент на мелких деталях, делают традиционные архитектуры, ориентированные на обработку изображений с ярко выраженной пространственной структурой, менее подходящими. Текстуры требуют моделей, способных улавливать абстрактные, повторяющиеся паттерны и игнорировать пространственные искажения. В связи с этим для решения задачи классификации была разработана свёрточная нейронная сеть, основанная на Texture CNN (T-CNN) [2], особенностью которой является использование энергетического слоя для агрегации плотных неупорядоченных текстурных признаков из активаций последнего свёрточного слоя. В отличие от традиционных подходов, где используется пространственная информация, энергетический слой объединяет значения по всей карте признаков, что позволяет извлекать глобальные текстурные характеристики.

Помимо выбора архитектуры, важнейшую роль играет предварительная обработка данных. Текстуры часто нуждаются в адаптивной нормализации, которая позволяет выравнивать гистограммы изображений и подчеркивать детали, не нарушая общей текстурной структуры. Подобная нормализация помогает сети лучше воспринимать различия между текстурами и улучшает качество обучения. В рамках данной работы рассматривается техника адаптивного выравнивания гистограммы с ограничением контраста (Contrast Limited Adaptive Histogram Equalization, CLAHE) [3]. В дополнение к контрастному преобразованию была применена нормализация на основе статистики датасета ImageNet, что позволило привести интенсивности каналов изображений к значениям с заданными средними и стандартными отклонениями.

Метод CLAHE позволяет улучшить локальный контраст и детализацию текстур за счёт адаптивного перераспределения яркости в пределах локальных областей изображения.

Математически процесс применения CLAHE включает в себя следующие шаги:

Изображение делится на тайлы. Затем для каждого тайла:

- Вычисляется локальная гистограмма контрастности $H_t(r)$.
- Производится нормализация локальной гистограммы: $P_t(r) = \frac{H_t(r)}{N_t}$, где N_t – общее количество пикселей в тайле.

- Вычисляется локальная кумулятивная функция распределения (CDF):

$$CDF_t(r) = \sum_{j=0}^r P_t(j)$$

- Производится преобразование значений пикселей внутри тайла $I' = CDF_t(I) \cdot (L - 1)$

После всех итераций результаты обработки тайлов интегрируются для получения конечного изображения. На рисунке 1 приведено сравнение одного случайного изображения макропористого криогеля ПВС в оригинальном виде и после обработки методом CLANE.

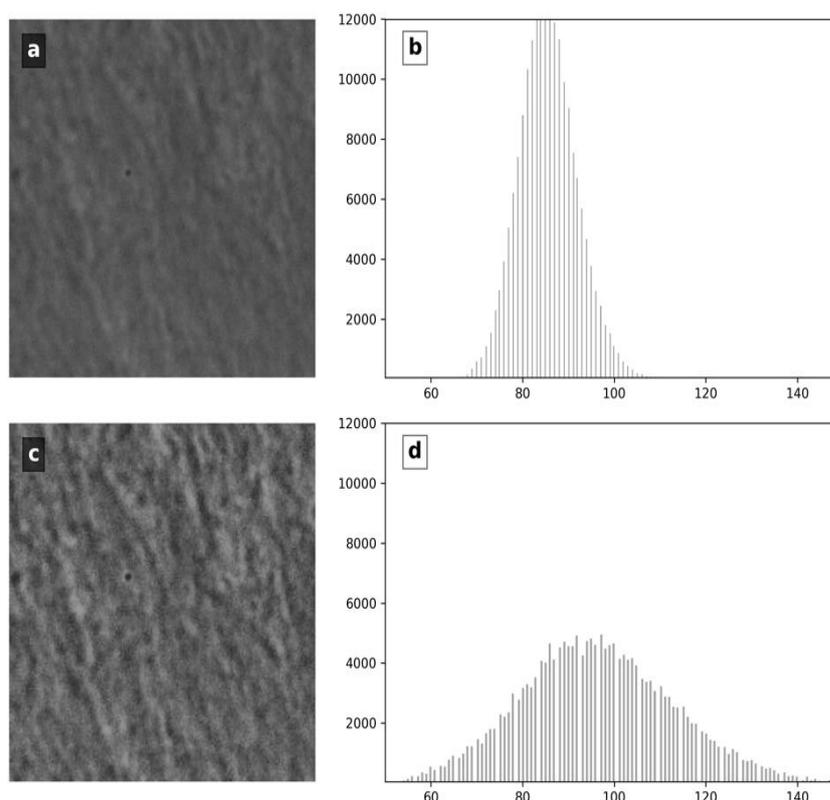


Рис. 1. Сравнение исходного изображения и его гистограммы с результатами обработки методом CLANE: *a)* исходное изображение, *b)* гистограмма исходного изображения, *c)* изображение после обработки, *d)* гистограмма изображения после

В исследовании использовался набор данных изображений макропористых криогелей ПВС [4], который был разделен на 20 классов, представляющих различные комбинации типов добавок (пять видов) и температур приготовления (четыре уровня). Макропористые криогели ПВС представляют собой физические гели, формирующиеся при охлаждении растворов полимеров до криогенных температур [1]. Они обладают анизотропной пористой структурой и формируются благодаря кристаллизации растворителя внутри геля.

Разработанная модель нейронной сети содержала 4 блока свёрток. Каждый блок включает в себя свёрточный слой (32, 64, 96, и 128 фильтров соответственно) с ядром 3×3 , с применением зеркального отражения входного тензора, и функцией активации ReLU, за которым следует слой максимального пулинга (pooling) с ядром 2×2 . Далее для соединения с полносвязными слоями

использовался пулинг с глобальным усреднением. Полученные в результате пулинга векторные представления подавались на вход полносвязной сети, состоящей из двух скрытых слоев с 1024 нейронами в каждом и выходного слоя, размерность которого соответствовала количеству классов. Также для предотвращения переобучения применялся слой Dropout с вероятностью отключения нейронов 0.5. Общая архитектура состояла из 1 387 700 обучаемых параметров. Модель обучалась в течение 30 эпох.

Результаты эксперимента продемонстрировали высокую точность классификации на валидационной выборке, которые приведены в таблице 1.

Таблица 1. Метрики качества работы модели.

Accuracy	Precision	Recall	F1-measure
0.8962	0.8981	0.8962	0.8960

Литература

1. Kurochkin, I.I. et al. (2020). "Cryostructuring of Polymeric Systems †: Application of Deep Neural Networks for the Classification of Structural Features Peculiar to Macroporous Poly(vinyl alcohol) Cryogels Prepared without and with the Additives of Chaotropes or Kosmotropes". *Molecules*, 25(19), p. 4480. doi:10.3390/molecules25194480.
2. Andrearczyk, V. and Whelan, P.F. (2016). "Using filter banks in Convolutional Neural Networks for texture classification". *Pattern Recognition Letters*, 84, pp. 63–69. doi:10.1016/j.patrec.2016.08.016.
3. Zuiderveld, K. (1994). "Contrast Limited Adaptive Histogram Equalization". *Graphics Gems*, pp. 474–485. doi:10.1016/b978-0-12-336156-1.50061-6.
4. Okunev, D.A. (2024). "Lozinsky Texture Dataset". *Kaggle*. <https://www.kaggle.com/datasets/okunevda/lozinsliy-texture-dataset/data>.

Астафьев А.В., Астафьева О.С.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, Владимирская область, г. Муром, ул. Орловская, д.23
E-mail: Alexandr.Astafiev@mail.ru*

Разработка систем распознавания человеческой деятельности на основе информации о состоянии канала связи

Направление обработки одномерных и многомерных сигналов является одним из наиболее обширных направлений научных исследований для решения повседневных прикладных задач. Одной из таких задач является построение систем распознавания человеческой активности. Работа направлена на решение фундаментальной научной проблемы – отсутствие высокоточных и универсальных методов и алгоритмов автоматического распознавания деятельности людей на сложных наблюдаемых территориях с использованием радиочастотных устройств. Результат научных исследований могут быть использованы для решения таких практических задач как: автоматический контроль соблюдения мер промышленной безопасности, координирование действий при возникновении чрезвычайных ситуаций, разработка охранных систем с сохранением конфиденциальности: определения факта присутствия и детектирование аномалий.

Исторически основными анализируемыми параметрами при работе с радиоустройствами (WiFi, Bluetooth и прочие) были время и уровень принимаемого сигнала (RSSI). Использование уровня принимаемого сигнала имеет большую популярность среди научных групп в виду того, что показания можно получить практически на любом устройстве, а шкала измерения стандартизирована. Однако, основными недостатками показателя уровня принимаемого сигнала является большая зашумленность и малая точность, в виду того, что значение является целочисленным. Есть ряд научных работ, которые показали высокую точность при работе с параметром уровня принимаемого сигнала, но при решении практических задач они также используются редко.

В настоящее время у прикладных разработчиков появилась возможность получать данные с физического уровня радиоустройств. Эта информация намного подробнее описывает сигнал и называется информация о состоянии канала связи (channel state information, CSI). Информация о состоянии канала связи описывает как сигнал распространяется от передатчика до приемника и включает в себя информацию о фазе и амплитуде на каждой из поднесущих. Так, за одно измерение по стандарту IEEE 802.11n можно получить информацию о фазе и амплитуде на 56 поднесущих на двух парах антенн. Общее количество параметров составляет 448. Этот факт побудил большое количество научных групп применять эту информацию для решения собственных практических задач.

В работе предлагается использование нейросетевых технологий в виду того, что явная аналитическая зависимость между активностями и многомерными канальными матрицами неизвестна. Так, разработаны четыре типа моделей нейронных сетей: LSTM, 1D-CNN, 1D-LSTM, BLSTM. Для обучения моделей была сформирована обучающая выборка, включающая в себя информацию об активностях пяти испытуемых. В экспериментальных исследованиях также была использована тестовая выборка, включающая в себя канальные матрицы, собранные с другом, схожем помещении, также при участии пяти испытуемых. Предложены четыре модели нейронных сетей, точность которых достигает 98,56% на этапе обучения. Самой точной моделью, по результатам исследований, стала модель BLSTM со значением точности в 86,26%.

При чем, точность всех проведенных испытаний не на много, но выше конкурентов. BLSTM более точно распознает факт того, что человек сидит или лежит. Стоящего человека более точно распознает модель 1D-LSTM. А вот ходьбу и бег на месте лучше распознают модели LSTM и 1D-CNN соответственно. В заключении можно сделать вывод, что использование канальных матриц для решения задачи распознавания человеческой активности позволяет достичь высокоточных результатов с сохранением возможности масштабирования.

Бейлекчи Д.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, Владимирская область, г. Муром, ул. Орловская, д.23
dmibei@yandex.ru*

Применение моделей рекуррентных нейронных сетей для фильтрации аддитивных помех в системах громкоговорящей связи

Наличие акустических помех и шумов в речевых сигналах, поступающих с микрофона в телекоммуникационных системах громкоговорящей связи, работающих в сложных акустических условиях, таких как производственные помещения, объекты транспортной инфраструктуры, является одной из основных проблем, влияющих на качество связи, которые приводят к потере информации или ее искажению при передаче данных через каналы связи по протоколам IP-телефонии [1].

Существуют различные алгоритмы адаптивной фильтрации акустических помех в речевых сигналах [2], но зачастую они требуют высокой вычислительной мощности обрабатывающих устройств, что повышает стоимость окончательного оборудования. Для реализации таких способов часто требуется или наличие локального дорогостоящего серверного оборудования и постоянный канал связи с сервером, или, если применяются облачные технологии, постоянное подключения к сети Интернет, что во многих случаях применения громкоговорящих систем сложно выполнить, и при этом ухудшается общая надежность системы.

Основной задачей данной работы является построение программной системы для исследования и оптимизации моделей, которые позволят обеспечить качественную фильтрацию на конечных устройствах с учетом ограниченной вычислительной мощности. Исследование проводится для реализации алгоритмов подавления помех в устройствах систем диспетчерской и оперативно-командной громкоговорящей связи.

Рекуррентные нейронные сети [3] являются эффективным инструментом для обработки временных последовательностей, таких как речевой сигнал, и фильтрации в нем акустических помех. Для повышения эффективности в задачах фильтрации речевых сигналов используются их более сложные вариации, такие как LSTM-сети, которые помогают преодолеть проблему исчезающего градиента, которая может возникать в классических рекуррентных нейронных сетях при работе с длинными последовательностями данных, делая обучение более стабильным и эффективным.

Существуют предложенные схемы применения моделей рекуррентных нейронных сетей, таких как RNNNoise [4], но для них отсутствуют результаты по оценке качества фильтрации для условий применения в системах громкоговорящей связи и оценки требуемой производительности для использования в архитектурах микроконтроллеров, применяемых в конечных устройствах систем громкоговорящей связи, таких как ARM Cortex-M.

Исследуемая нейронная сеть построена на основе управляемых рекуррентных блоков (GRU-сеть) и состоит из трех скрытых слоев с общей архитектурой 42-24-48-96-22 (всего 232 нейрона). Управляемый рекуррентный блок является более современной и упрощенной вариацией рекуррентных нейронных сетей по сравнению LSTM. GRU обладает схожей структурой, но с меньшим количеством управляющих ключей, при этом сохраняя эффективность в моделировании долгосрочных зависимостей и требуя меньшего количества параметров для

обучения. Алгоритм обучения сети использует метод рекурсивного спуска и может потребовать высоких вычислительных затрат, но выполняется один раз при настройке модели.

Для исследования характеристик модели фильтрации была разработана программная система, структурная схема которой приведена на рисунке 1.



Рис. 1 – Функциональная схема системы для исследования модели обработки сигнала.

Речевые сигналы с естественным или наложенным шумом подаются на вход модели нейронной сети, что позволяет проводить комплексную оценку качества фильтрации модели в различных шумовых условиях, анализировать ее эффективность и устойчивость к различным типам помех.

Для оценки эффективности подавления помех в системе предусмотрены как визуальные, так и количественные методы анализа. Для визуальной оценки применяются спектрограммы, отображающие частотно-временную структуру как исходного сигнала, так и сигнала после фильтрации. В качестве количественных метрик используются отношение сигнал/шум (SNR) и показатель субъективного качества речи PESQ ITU-T P.862.

Разработанная система исследования моделей фильтрации построена по модульному принципу, при этом за фильтрацию сигнала отвечает модуль нейросети, написанный на языке C, который может быть заменен на другой модуль, реализующий другой алгоритм фильтрации речевого сигнала, в том числе другой модели нейросети. Оптимизированный и протестированный модуль фильтрации речевого сигнала может быть применен в качестве модуля в программном обеспечении устройства для фильтрации речевого сигнала в реальном времени.

Применение программного модуля исследуемой модели нейросети при использовании вычислений с плавающей точкой на ряде одноплатных компьютеров с процессором архитектуры ARM Cortex-A53 при тактовой частоте 1.2 ГГц, показывает нагрузку на одно ядро процессора до 20% при обработке одного потока речевого сигнала, что требует оптимизации реализации алгоритма для снижения до приемлемой нагрузки в 5-10%.

Разработанная система позволяет выполнить оптимизацию количества формируемых признаков для обработки нейронной сетью, а также использовать целочисленные операции и операции с фиксированной точкой при вычислении признаков и расчетах выходов сети с контролем требуемого качества речевого сигнала, что необходимо для реализации данного алгоритма на микроконтроллерах архитектур ARM Cortex-M4F и Cortex-M7 и применения модуля в оконечных устройствах громкоговорящей связи.

Литература

1. Столбов М.Б. Основы анализа и обработки речевых сигналов: Учебное пособие. – Санкт-Петербург: Университет ИТМО, 2021. – 101 с.
2. Бейлекчи Д.В. Обработка акустических сигналов методами локального анализа в телекоммуникационных системах / Д.В. Бейлекчи, В.А. Ермолаев, Ю.А. Кропотов, // Радиотехнические и телекоммуникационные системы. 2015.- № 1 (17). - С. 49-56.
3. Гудфеллоу Я., Бенджио И., Курвилль А. «Глубокое обучение»/ пер. с англ. А.А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018 – 652 с.
4. Valin, Jean-Marc. (2018). A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement. DOI:10.1109/MMSP.2018.8547084.

Белов А.А.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264 г. Муром, Владимирской обл., ул. Орловская, д. 23
e-mail: aleks.murom@mail.ru*

Предварительная обработка спутниковых снимков для создания растровой подложки ГИС автоматизированной системы экологического мониторинга

При решении задач автоматизированного экологического мониторинга выбросов газообразных загрязняющих веществ, на промышленно развитых территориях возникает проблема предварительной обработки фотографических данных, полученных со спутников. Данные фотографические снимки используются для создания растровой картографической подложки геоинформационной подсистемы системы мониторинга выбросов и служат важнейшим источником пространственно-координированных данных об источниках загрязняющих выбросов и других объектах экологического мониторинга. Также на растровой фотографической подложке, наиболее удобно визуализировать зоны распространения опасных, газообразных, загрязняющих веществ, согласно выбранных расчетных методик. Растровая подложка также служит базой для процедуры создания векторной карты ГИС, на которой каждый из объектов в рамках тематических слоев будет подвергнут векторизации.

Качественные характеристики находящихся в свободном доступе спутниковых снимков промышленно развитых территорий зачастую не удовлетворяют заданным параметрам, снимки могут содержать области цветового несоответствия, разнообразные искажения, артефакты, присущие спутниковым фотографическим системам.

Обнаруженные дефекты растровой подложки ГИС требуется выделить и минимизировать. Для работы с картографическими данными максимально удобным является применение подходов пространственно-частотной обработки, базирующейся на применении аппарата вейвлет-преобразования. При этом, как стандарт в области анализа и обработки фотографических изображений, будут применяться методы двумерного вейвлет-разложения.

С применением вейвлетов будет произведена первичная обработка спутниковых фотографических снимков, удалены разнообразные «компоненты шума на изображениях», произведена адаптивная фильтрация, обнаружены и устранены разнообразные дефекты спутниковых снимков. Вейвлеты позволяют выделить важнейшие структурные признаки спутниковых снимков, включая сегментацию объектов на изображениях, повышение четкости и контрастности контуров объектов, минимизация цветовых перепадов и прочее. При этом ставится задача сжатия изображений, что также с легкостью достигается на основе двумерного вейвлет-преобразования. Практически минимизируя потерю значимой части информации, вейвлет-преобразование, осуществляет сжатие данных с высоким коэффициентом. Интерес вызывают применение вейвлет-анализа для решения задачи поиска картографических данных в БД на основе нечетких запросов.

Подход к анализу и обработке данных спутниковых снимков в экологической ГИС-подсистеме включает в себя ряд этапов:

- ввод цифровых картографических снимков из различных источников в систему мониторинга опасных, газообразных выбросов;
- предварительную цифровую обработку изображений на основе вейвлет-разложения;

- получение сведений о расположении и пространственных отношениях объектов мониторинга на основе обработки и анализа структурных признаков изображений;
- компактное хранение и вывод обработанных картографических изображений.

Белякова А.С., Сенечкин Д.О.

Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264 г. Муром, Владимирской обл., ул. Орловская, д. 23
asbelyakova@rambler.ru

Перспективы развития программных средств контроля за состоянием здоровья при сахарном диабете

По данным Всемирной организации здравоохранения, в мире 422 миллиона человек имеют диагнозом «сахарный диабет». По расчетам Минздрава, в нашей стране это заболевание диагностировано примерно у 10,5 млн человек — у каждого 15-го россиянина. Около 95 % из них имеют сахарный диабет 2 типа, а остальные – сахарный диабет 1 типа [1]. Пациенты с СД 1 типа для регуляции углеводного обмена подкожно вводят инсулин, а пациенты с СД 2 в основном используют диету и таблетированные препараты. Это объясняется различными особенностями углеводного обмена при этих заболеваниях. Важным и обязательным при этом является контроль показателей глюкозы крови.

Системы непрерывного мониторинга глюкозы (НМГ) крови представляют собой комплекс программно-технических средств: датчик содержит электрохимический сенсор, осуществляющий измерение уровня глюкозы в крови и связанные с ним электронные компоненты. Сенсор находится под кожей в тканевой жидкости, откуда в клетки поступает кислород и полезные вещества, включая глюкозу. Датчик (трансмисмиттер) посредством bluetooth сигнала взаимодействует с программным обеспечением, установленном на смартфоне пациента и позволяющим следить за уровнем глюкозы непрерывно (рис. 1). В зависимости от тенденций изменения показателей глюкозы крови пациент регулирует объем инсулинотерапии, что позволяет достигать более качественное компенсации сахарного диабета[2].

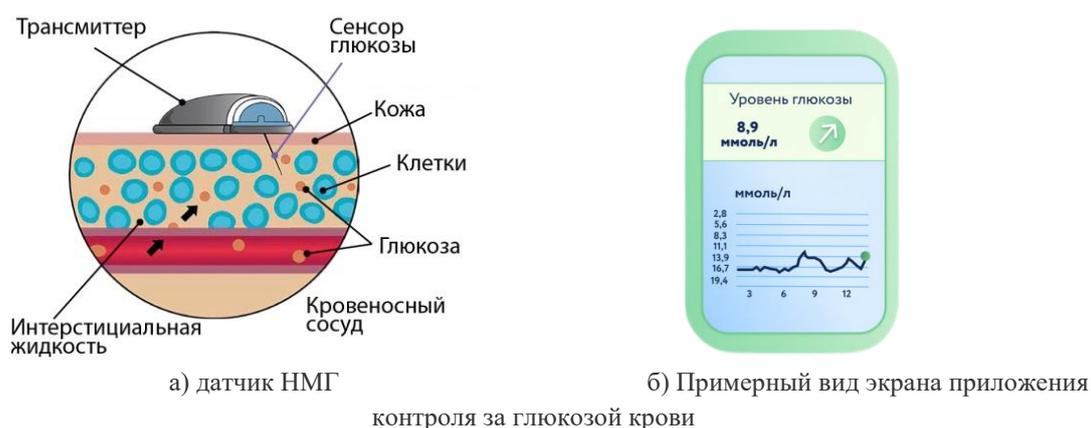


Рис. 1 – Система непрерывного мониторинга глюкозы крови

Кривая глюкозы крови представляет собой одномерный массив данных, значения в котором зависят множества параметров: количества и состава потребляемой пищи, корректности настроек базисной и болюсной инсулинотерапии, уровня и вида активности пациента, его антропометрических параметров, а также особенностей работы организма конкретного пациента. Несмотря на то, что существуют общие подходы к компенсации сахарного диабета, нет единого правила для всех пациентов, по которому можно было бы однозначно определить

как компенсировать сахарный диабет так, чтобы держать глюкозу крови в диапазоне нормы [3]. Поэтому для корректного подбора доз базисной и болюсной инсулинотерапии важно вести подробный дневник с указанием объема потребляемой пищи, активности, состояния здоровья и т.д. Это процесс требует времени, терпения и усидчивости, поэтому сложно сохранять всю эту информацию ежедневно в течении жизни.

Современные программные средства контроля за глюкозой крови предоставляют функционал, позволяющий хранить данные об объеме базисной и болюсной терапии, количестве принятой пищи в граммах углеводов или в хлебных единицах (1 хлебная единица = 10-12 граммов углеводов). Такими программными средствами являются XDrip+, DiaBox, LibreLink (для сенсоров FreeStyle Libre)[3] и другие. Однако, для корректного прогнозирования изменений глюкозы крови этих данных недостаточно: не учитывается белково-жировая составляющая, способ приготовления пищи, ее консистенция, сочетаемость различных продуктов питания и их гликемический индекс.

Также существует большое разнообразие программ контроля за калорийностью питания таких как FatSecret, калькулятор калорий ХиКи, EatFit, Eatsum, Easyfit, Nutrilio и других.

Перспективным является объединение возможностей программных приложений для контроля за глюкозой крови и приложений для контроля за калорийностью и составом пищи, обладающее следующими возможностями:

1. Получение данных НМГ в реальном режиме времени
2. Организация возможности хранения, добавления, изменения данных о потребляемых моно продуктах с учетом их количества, пищевой ценности, консистенции, гликемического индекса.
3. Организация возможности хранения, добавления, изменения данных о потребляемых сложных блюдах с учетом их состава, количества, пищевой ценности, консистенции, гликемического индекса.
4. Автоматический расчет пищевой ценности сложных блюд.
5. Организация возможности хранения, добавления, изменения данных о двигательной активности и ее продолжительности. Перспективным является получение этой информации с помощью фитнес-браслета.
6. Организация возможности хранения, добавления, изменения данных о состоянии здоровья: признаки заболеваний и уровень их интенсивности.
7. Автоматический расчет углеводного коэффициента в зависимости от времени суток.

После накопления максимально возможного объема информации за несколько месяцев непрерывного мониторинга планируется разработка методики и алгоритмов краткосрочного прогнозирования глюкозы крови и коррекции базисно-болюсной инсулинотерапии в соответствии с углеводными коэффициентами конкретного пациента и его чувствительностью к инсулину в различное время суток.

Данное направление в области развития средств контроля за состоянием здоровья за состоянием здоровья является перспективным и предназначено не только для лиц с сахарным диабетом, но и для каждого человека, следящего за состоянием своего организма.

Литература

1. Дедов И.И., Шестакова М.В., Викулова О.К., Железнякова А.В., Исаков М.А., Сазонова Д.В., Мокрышева Н.Г. Сахарный диабет в Российской Федерации: динамика эпидемиологических показателей по данным Федерального регистра сахарного диабета за период 2010 – 2022 гг. Сахарный диабет. 2023;26(2):104-123. <https://doi.org/10.14341/DM13035>
2. Петеркова В.А., Аметов А.С., Майоров А.Ю., Галстян Г.Р., Лаптев Д.Н., Черникова Н.А. Резолюция научно-консультативного совета «Применение технологии непрерывного

мониторинга глюкозы с периодическим сканированием в достижении гликемического контроля». Сахарный диабет. 2021;24(2):185-192. <https://doi.org/10.14341/DM12753>

3. Белякова А.С., Михеева А.М. Оценка изменений колебаний глюкозы крови по данным непрерывного мониторинга посредством использования статистических характеристик. Сборник тезисов докладов всероссийской научной конференции. 2024.-с.280-281. https://www.mivlgu.ru/conf/zvorykin2024/pdf/sec08_full.pdf

Болгак А.В., Ватулин Э.И.
Юго-Западный государственный университет
305040, г. Курск, ул. 50 лет Октября, 94

Устройство для умножения квадратных бинарных матриц с конвейеризацией операции чтения данных из специализированной многопортовой памяти

Высокопроизводительные вычисления стали неотъемлемой частью повседневной жизни человека, влияя на множество ее аспектов, от прогнозирования погоды и разработки новых лекарственных препаратов до обеспечения национальной безопасности. В основе многих вычислительно сложных задач лежит именно матричное умножение. Эффективность его реализации определяет скорость решения огромного спектра задач. К областям применения высокопроизводительных вычислений, основанных на матричных операциях, можно отнести: моделирование баллистических траекторий, обработка спутниковых изображений, создание систем раннего предупреждения, молекулярное моделирование, поиск новых лекарственных препаратов, анализ генома, моделирование поведения сложных механизмов (САЕ), томография (КТ, МРТ), обработка медицинских изображений, анализ данных электроэнцефалографии и электрокардиографии, моделирование климата. Многие алгоритмы, используемые в перечисленных областях, сводятся именно к матричным операциям, например, нейронные сети, лежащие в основе современных систем искусственного интеллекта, выполняют миллиарды матричных умножений за секунду. Эффективность этих операций напрямую влияет на время обучения модели и её точность.

Актуальность оптимизации матричного умножения обусловлена стремлением максимально использовать вычислительные ресурсы современных процессоров и графических ускорителей. Различные модификации и гибридные подходы позволяют адаптировать алгоритм к конкретной аппаратной платформе и размеру матриц: минимизация числа обращений к оперативной памяти, использование кэширования данных, параллельная обработка данных, использование технологий OpenMP, CUDA, OpenCL позволяет значительно ускорить вычисления.

Эффективное умножение матриц является ключевым фактором в достижении высоких скоростей вычислений в самых разных областях. Непрерывное развитие алгоритмов и аппаратных средств, а также интеграция различных методов оптимизации позволяют решать все более сложные задачи, открывая новые горизонты в науке и технологиях. Дальнейшее развитие в этой области будет определять темпы прогресса во многих критически важных сферах человеческой деятельности [2].

Одно из направлений в умножении матриц представлено задачами умножения бинарных матриц, которые возникают в области теории графов и дискретной математики в целом (хранение и обработка информации о бинарных отношениях) [1]. Целью данной работы является оценка временных затрат на обработку квадратных бинарных матриц размером $n \times n$ устройства с конвейеризацией операции чтения данных из специализированной многопортовой памяти и ее сравнение с временными затратами прототипа.

Известно устройство с систолической структурой (прототип) для умножения произвольных квадратных матриц, содержащее матрицу из $n \times n$ операционных блоков для выполнения элементарных операций умножения с накоплением и блоки коэффициентов матриц (см. рис. 1) [3]. Недостатком прототипа являются большие временные затраты при умножении бинарных матриц и, как следствие, низкое быстродействие, которое лимитируется доступом в специализированную многопортовую память при обработке матриц размером $n > 64$. Технической задачей предложенного устройства является уменьшение временных затрат на обработку квадратных бинарных матриц за счет конвейеризации операции чтения данных из

специализированной многопортовой памяти. Это достигается добавлением в каждую ячейку хранения групп двухступенчатых триггеров (см. рис. 2), что позволяет значительно уменьшить временные затраты на этапе работы устройства путем снижения величины конвейерного такта.

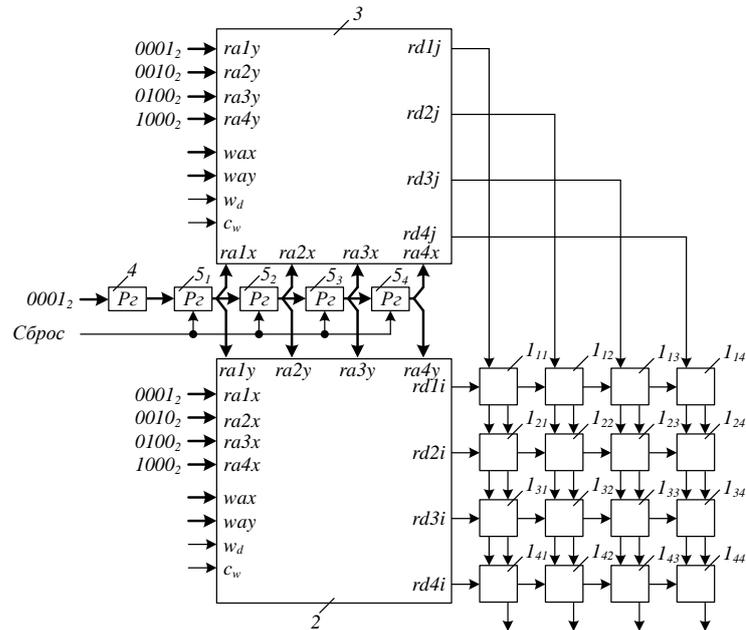


Рис. 1. Функциональная схема устройства

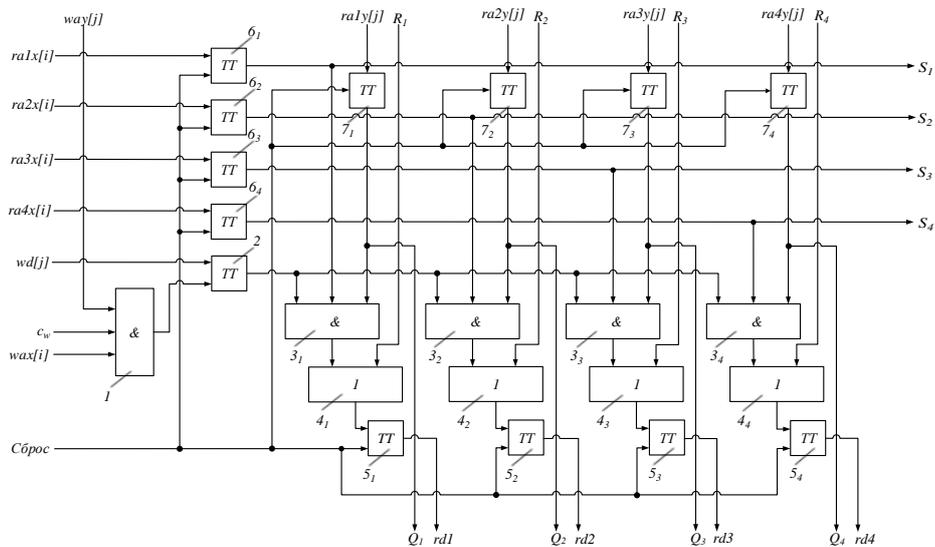


Рис. 2. Схема ячейки блока хранения предлагаемого устройства

На основании расчетов можно сделать вывод, что предлагаемое устройство для умножения квадратных бинарных матриц с конвейеризацией операции чтения из специализированной многопортовой памяти позволяет снизить время обработки квадратных бинарных матриц размером $n < 2000$ до 200 раз.

В перспективе дальнейших исследований планируется детальная оценка временных затрат на обработку квадратных бинарных матриц, оценка получаемого выигрыша в зависимости от n , а также оценка аппаратной сложности предлагаемого устройства.

Литература

1. Ватутин Э.И., Зотов И.В. Построение матрицы отношений в задаче оптимального разбиения параллельных управляющих алгоритмов // Известия Курского государственного технического университета. Курск, 2004. № 2. С. 85–89.
2. Штейнберг Б.Я. Блочно-рекурсивное параллельное перемножение матриц // Известия высших учебных заведений. Приборостроение. 2009. Т. 52. № 10. С. 33–41.
3. Патент РФ на полезную модель № 193927. Устройство для умножения бинарных матриц. Заявл. 26.06.2019, опубл. 21.11.2019, бюл. № 33.

Борданов И.А., Щаников С.А.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
E-mail: bordanov2011@yandex.ru*

Оценка точности работы искусственных нейронных сетей на базе мемристивных устройств

Человеческий мозг способен обрабатывать данные биологическими нейронными сетями при этом потребляя всего 1 фДж на синаптическое событие [1]. Вдохновленные эффективностью вычисления в мозге, искусственные нейронные сети (ИНС) добились больших успехов в задачах искусственного интеллекта, таких как распознавание визуальной и аудио информации, анализ больших объемов данных, принятии решений и т.д [2]. Однако в связи с растущим спросом на вычислительные мощности обычные компьютеры на архитектуре фон Неймана, которые в большинстве своём используются для реализации алгоритмов работы нейронных сетей, демонстрируют серьезную временную задержку и высокое энергопотребление [3].

Одним из перспективных решений сложившейся проблемы может быть переход от классической архитектуры фон Неймана к более биоподобным, таким как вычисления в памяти, которые базируются на принципах того, что и вычислитель, и память являются одним устройством. Такая реализация позволит избавиться от накладных затрат связанных с пересылкой данных от устройства хранения к устройству обработки и соответственно повысить энергоэффективность ИНС [4].

Для осуществления вычислений в памяти ИНС могут быть аппаратно реализованы на базе мемристивных устройств (ИНСМ), резисторов с эффектом памяти. Использование мемристивных устройств, в качестве синаптических элементов, позволяет осуществлять энергонезависимое хранение весов, что также положительно сказывается на энергоэффективности.

Несмотря на то, что такая реализация обладает рядом преимуществ, ее основным недостатком является подверженность вариациям характеристик мемристивных элементов. Эти вариации, в свою очередь, могут вносить искажения в весовые коэффициенты синапсов, снижая точность функционирования ИНС [5]. Следовательно, оценка качества функционирования таких ИНСМ является важным этапом на стадии проектирования.

Для решения данной проблемы в работе предложен подход к оценке точности работы ИНСМ с применением эмпирических моделей мемристивных устройств, а также методологии имитационного моделирования. Применение данного подхода было апробировано на примере ИНСМ обученной решению задачи классификации данных.

Литература

1. Laughlin S. B., de Ruyter van Steveninck R. R., Anderson J. C. The metabolic cost of neural information // Nature neuroscience. 1998. Vol. 1. №. 1. С. 36-41.
2. Rahman F. Nonlinear Dynamics of Neural Networks: Applications in Pattern Recognition // Communications on Applied Nonlinear Analysis. 2023. Vol. 30. №. 3. С. 16-31.
3. Ielmini D., Wong H. S. P. In-memory computing with resistive switching devices // Nature electronics. 2018. Vol. 1. №. 6. С. 333-343.

4. Amirsoleimani A. et al. In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal Oxide Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives //Advanced Intelligent Systems. 2020. Vol. 2. №. 11. С. 2000115.

5. Bordanov I., Antonov A., Korolev L. Simulation of calculation errors in memristive crossbars for artificial neural networks // 2023 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE, 2023. С. 1008-1012.

Булаев А.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
bulaevav.work@yandex.ru*

Обзор методов обнаружения и идентификации объектов на картографических изображениях

Задача обнаружения объектов интереса и аномалий на видеопоследовательностях и отдельных изображениях остается актуальной уже не одно десятилетие, и число приложений, где она возникает, только возрастает. Среди них приложения, связанные с дистанционным зондированием Земли, наземным мониторингом, медицинскими исследованиями, радиолокацией, сельским хозяйством, рентгеновским сканированием багажа и многие другие. Для обнаружения объектов интереса используется различная априорная информация вплоть до заданной библиотеки эталонов, а под аномалией (злокачественные образования, появление артефактов на видеопоследовательности и пр.), как правило, понимается некоторая область изображения, характеристики которой отличаются от прогнозных, сформированных в процессе обработки изображений [1].

Обнаружение объекта является результатом его распознавания на картографическом изображении, т.е. выявление положения объекта интереса и его идентификация или отнесение к какому-либо классу объектов (заранее предопределенном).

Для распознавания образов выделяют три типа методов, отличающихся принципами своей работы:

- метод перебора;
- анализ характеристик образов;
- обучение искусственных нейронных сетей.

Используемые методы распознавания образов напрямую связаны с представлением описания объектов рассматриваемой предметной области. По этому критерию методы относятся к интенциональным (методы распознавания, основанные на операциях с признаками объектов) и экстенциональным (методы распознавания, основанные на операциях с объектами).

Каждая группа методов включает несколько подгрупп [2]:

1) интенциональные методы: методы, основанные на оценках плотностей распределения значений признаков; методы, основанные на предположениях о классе решающих функций; лингвистические методы; логические методы;

2) экстенциональные методы: метод сравнения с прототипом; метод k-ближайших соседей; алгоритмы вычисления оценок; коллективы решающих правил.

Каждая группа методов позволяет распознавать объекты с разной степенью точности, по разному набору исходных данных, но в условиях, когда образы объектов искомым и объектов на анализируемом изображении равноценны по своему представлению. Что не всегда применимо к картографическим изображениям, поскольку они могут формироваться в условиях, которые могут дополнительно вносить погрешность в распознавание.

Так при анализе снимков с беспилотного летательного аппарата (БПЛА) на качество распознавания объектов может влиять время суток, в которое делается снимок (за счет влияния теней отбрасываемых объектами), степень контрастности распознаваемых объектов на фоне

подстилающей поверхности, время года, перекрытие искомым объектам другими объектами (естественного и техногенного происхождения) и другие факторы.

Все указанные выше явления могут снижать качество распознавания объектов при применении классических методов обработки изображений. Зачастую для достижения лучших результатов используются комбинированные решения, в которых, например, поиск объектов на изображениях оригиналах ведется с применением методов обработки изображений, а их классификация реализуется с использованием нейросетевых подходов.

В МИ ВлГУ на протяжении уже многих лет ведутся исследования в области применения топологических признаков (характеристик) изображения для его анализа.

Исследования применимости топологических методов обработки информации показали их эффективность при описании векторных структуры городских объектов и решении задач, основанных на анализе их взаимного расположения [3-5], применимость топологического подхода для представления n-мерных объектов в картографии [6].

Авторами статей [3, 5, 6] ведутся исследования в области применимости топологического подхода к описанию растровых картографических сцен с применением баркодов [7].

В настоящий момент доказана применимость подхода использования баркодов к решению практических задач по оценке подобия объектов, по их сопоставлению на разномасштабных картах [8], по обнаружению объектов [9].

В дальнейшем предполагается проводить исследования применимости методов топологического представления картографических сцен, распознавания и идентификации объектов с применением баркодов по видеопоследовательности получаемой БПЛА и применения полученных данных для автоматической корректировки маршрута автономного устройства на основе изображений подстилающей поверхности.

Особый интерес представляет распознавание объектов по частичному совпадению баркодов самого объекта и картографической сцены в условиях динамического изменения картографируемых участков местности (например, при съемке с БПЛА).

Литература

1. Андриянов, Н.А. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet / Н.А. Андриянов, В.Е. Дементьев, А.Г. Ташлинский // Компьютерная оптика. – 2022. – Т. 46, № 1. – С. 139-159. – DOI: 10.18287/2412-6179-СО-922.
2. Калинина, Н.Д. Анализ методов распознавания и поиска образов на космических снимках / Н.Д. Калинина, А.В. Куров // Инженерный журнал: наука и инновации. – 2012 - №1. – С. 174-188.
3. Еремеев, С.В. Алгоритмы формирования графовой модели городской территории в ГИС / С.В. Еремеев, Д.Е. Андрианов, В.А. Комков // Геоинформатика. – 2013. – № 4. – С. 19-24.
4. Булаев, А.В. Формальная модель установления топологических отношений с объектами, содержащими криволинейные сегменты / А.В. Булаев // Алгоритмы, методы и системы обработки данных. – 2008. – № 13. – С. 16-24.
5. Еремеев, С.В. Алгоритм кодирования пространственных идентификаторов в иерархических топологических системах / С.В. Еремеев, М.М. Филимонов // Журнал «АМиСОД». – 2014. – Т. 4, № 29. – С. 50-58.
6. Андрианов, Д.Е. Алгоритм кодирования nD пространственных объектов в ГИС / Д.Е. Андрианов, С.В. Еремеев, Ю.А. Ковалев, К.В. Купцов // Сборник трудов ИТНТ-2018 IV Международная конференция и молодёжная школа «Информационные технологии и нанотехнологии» (ИТНТ-2018) – 2018. – С. 800-805.

7. Модели и алгоритмы обработки и анализа данных в геоинформационных системах / К.В. Купцов, С.В. Еремеев, Д.Е. Андрианов. – М. : ВНИИГеосистем, 2020. – 98 с. : ил.

8. Еремеев, С.В. Алгоритм совмещения пространственных объектов разномасштабных карт на основе топологического анализа данных / С.В. Еремеев, Д.Е. Андрианов, В.С. Титов // Компьютерная оптика. – 2019. – Т. 43, № 6. – С. 1021-1029. – DOI: 10.18287/2412-6179-2019-43-6-1021-1029.

9. Еремеев С.В. Обнаружение многолетних бугров пучения с использованием декомпозиции цифровых моделей рельефа по топологическим признакам / С.В. Еремеев, А.В. Абакумов, Д.Е. Андрианов, И.В. Богоявленский, Р.А. Никонов // Современные проблемы дистанционного зондирования Земли из космоса. - 2023. - Т. 20. № 6. - С. 129–143.

Варламов А.Д., Фомин В.Е.

Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

602264 г. Муром, Владимирской обл., ул. Орловская, д. 23

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

г. Томск, пр. Ленина, 36

varlamov_aleks@mail.ru; vadim.fomin.2001@bk.ru

Разработка алгоритмической модели работы виртуального помощника ученика для изучения иностранных слов на образовательной платформе

В последнее время преподаватели иностранных языков проявляют большой интерес к использованию технологий искусственного интеллекта для решения целого ряда учебных задач, в том числе для изучения и повторения иностранных слов [1]. С этой целью разработчики цифровых образовательных платформ стремятся внедрить в работу электронных курсов чат-боты и виртуальные ассистенты [2].

Заметим, что виртуальный (цифровой) ассистент в числе технологий искусственного интеллекта использует обработку естественного языка, имитируя живой диалог с пользователем [3]. Соответственно, ввиду значимости и востребованности внедрения цифрового ассистента с технологиями искусственного интеллекта в образовании [4-8], данная работа является актуальной.

С учетом специфики задачи, целевой аудитории пользователей образовательной платформы (ученики 5-9 классов), возможностей методов искусственного интеллекта и представленных на рынке средств их реализации была разработана алгоритмическая модель работы виртуального помощника ученика, которая включает в себя следующие этапы:

1. Вызов помощника учеником.
2. Формирование словаря для текущей работы.
3. Приветствие ученика и призыв к началу тренинга.
4. Обработка ответа ученика с помощью интентов.
5. Анализ результата ответа ученика. Если ответ непонятный или некорректный, то выполняется отправка сообщения пользователю с просьбой повторить ответ и переход к этапу 4. Если ученик ответил отказом, то выполняется завершающий этап 12.
6. Выбор случайного слова из словаря.
7. Формирование просьбы к ученику перевести выбранное слово.
8. Обработка ответа пользователя.
9. Анализ результата ответа ученика. Если система посчитала ответ непонятным или некорректным, то пользователю отправляется просьба повторить ответ и происходит переход на этап 8. Если ответ правильный, то происходит переход на этап 10, иначе на этап 11.
10. Если ученику были предложены все слова из текущего словаря, то происходит переход на этап 12, в противном случае на этап 6.
11. Если перевод слова есть в общем словаре, то дается определение перевода слова и подсказка с переводом слова ученика. После этого выполняется переход на этап 7.
12. Отправляется сообщение с прощанием и призывом к возвращению.

Данная модель предполагает обязательное использование методов искусственного интеллекта на этапах 4-6 и 8-9. Предполагается, что на этих ответственных задачах будут задействованы большие языковые модели [9].

Таким образом, применение виртуального помощника ученика на базе языковых моделей позволяет персонализировать обучение и сократить нагрузку на преподавателей. Дальнейшее развитие подобных технологий направлено на улучшение распознавания естественного языка и адаптацию образовательного контента под индивидуальные потребности пользователей.

Литература

1. Brown T., Mann B. AI-Powered Language Learning: Opportunities and Challenges // *Advances in NLP*. 2023. №2. С. 25-33.
2. Johnson K. Chatbots in Education: A New Era of Interactive Learning // *EdTech Review*. 2023. №1. С. 12-18.
3. OpenAI Research Team. Natural Language Processing in Digital Assistants // *White Paper*. 2024.
4. Smith J. The Role of AI in Personalized Education // *AI and Education*. 2023. №4. С. 45-53.
5. Patel R. Building Effective Vocabulary Training Systems with AI // *Journal of AI in Learning*. 2023. №3. С. 30-38.
6. Lee A., Zhao M. Intent Recognition in Conversational AI Systems // *Computational Linguistics Journal*. 2022. №5. С. 22-29.
7. Gartner Research. Enhancing AI-Based Language Learning through NLP // *Tech Report*. 2023.
8. Microsoft AI Lab. Lexical Analysis and Translation Models for AI Tutors // *White Paper*. 2023.
9. FutureEd Initiative. Large Language Models in Educational Applications // *Policy Brief*. 2024.

Eduard Vatutin^{1,2}, Jia Wei-Ting², Jun Chi Ma², Qiang Miao²¹ *Southwest State University**305040, Russia, Kursk, 50 let Oktyabrya st., 94*² *Daqing Normal University**163712, China, Daqing, Xiyuan st., 85*

Using spectra of numerical characteristics of Latin squares during getting spectra of diagonal Latin squares

One of the well-known types of combinatorial objects are Latin and diagonal Latin squares (abbr. LS and DLS respectively). A large number of scientific publications are devoted to the study of their properties and the solution of a number of related combinatorial problems [1]. One of the directions is the construction of spectra of the selected numerical characteristic for a given type of LS/DLS. In this publication, the main attention is aimed to the construction of spectra of the number of intercalates in the DLS of order N (numerical series A345760 in OEIS¹). This problem was solved by Brute Force approach for small orders $1J \ NJ \ 7$, Brute Force approach with separating of the search space into equivalence classes and using parallel computing systems for orders $8J \ NJ \ 9$ [2], and cannot be solved exactly for larger orders due to the extremely high computational complexity, in view of which, to solve it, it is possible to use heuristic methods [3], which ensure the receipt of heuristic approximations of the desired spectra in a reasonable time.

The method for constructing heuristic approximations of spectra proposed in [4] is based on constructing a base spectrum from some source of initial DLSs with its subsequent expansion by traversing the neighborhoods obtained by rotating intercalates and short loops for the DLSs in the current spectrum being processed. This method shows good results and provides significantly more powerful spectra compared to, for example, the random (RS) or limited Brute Force (LBF) methods. Information on the spectral approximations obtained with its usage is given in [5] and is presented in the left part of the table.

The key distinguishing feature of DLS from LS by definition is the prohibition of duplicate values appearing on the diagonals of the square. Taking this feature into account leads to the fact that rotations of some intercalates and loops are prohibited, since they may result in duplicate values appearing on the diagonals, which is unacceptable for DLS. The consequence of this is a narrowing of the search space (decreasing the cardinalities of the neighborhoods of squares in the processed spectrum). There are no such restrictions when processing the LS, and it is expected that more powerful spectra² and more stricter restrictions on the minimum and maximum values in their composition will be obtained in comparison with the previously constructed DLS spectra.

In order to test this assumption, similar to those considered earlier [5], a series of computational experiments were organized with the aim of constructing spectra of the number of intercalates in the LS. Computational experiments were performed in single-threaded mode on the Core i7 4770 CPU, the computational time costs amounted up to 2 weeks to one experiment for orders $NJ \ 20$ (processing larger orders without using of parallel computing systems is difficult), their results are shown in the middle part of the table. Analysis of the obtained results allows us to conclude that for all orders $10J \ NJ \ 20$ it was possible to obtain more powerful LS spectra compared to similar DLS spectra (for example, $93 \rightarrow 103$ for order $N = 10$, $100 \rightarrow 145$ for order $N = 11$, etc.).

¹ <https://oeis.org/A345760>

² <https://oeis.org/A368182>

New elements of the LS spectra that absent in the DLS spectra do not allow the original DLS spectra to be expanded directly. However, the diagonalization procedure [6] can be applied to them in order to obtain DLSs from given LSs that are isomorphic to them as LS and, accordingly, have the same number of intercalates, which is an invariant of the corresponding main class of LS. In some cases this can be done (not all LSs can be diagonalized to DLSs), which ultimately allows us to obtain more powerful DLS spectra (for example, the cardinality of the spectrum of the number of intercalates in the DLS was increased using a similar technique from 93 to 98 items for order $N = 10$ (5 LSs can't be diagonalized to DLS), from 100 to 145 items for order $N = 11$, etc.). Starting from order $N = 16$ the diagonalization procedure becomes computationally complex due to the significant increase of the number of transversals in the diagonalized LSs, which does not allow its usage in practice without a significant (by several orders of magnitude) increase in the necessary computational costs and parallelization of the corresponding computational code.

Thus, in the result of the computational experiments performed, the most powerful heuristic approximations of the spectra of the number of intercalates in the LS and DLS known at the moment were constructed for orders 10J NJ 20, obtained results have been tested and are currently in the process of being added to the OEIS encyclopedia.

Table. Parameters of the spectra of the number of intercalates in LS and DLS of order N

Order N	DLS spectrum S without transition to LS spectrum			LS spectrum S			DLS spectrum S after transition to LS spectrum		
	inf S	sup S	$ S $	inf S	sup S	$ S $	inf S	sup S	$ S $
10	0	101	93	0	125	103	0	109	98
11	0	100	100	0	172	145	0	172	145
12	2	252	210	0	324	259	0	324	259
13	0	180	176	0	208	200	0	208	200
14	3	353	337	0	391	362	0	391	362
15	5	319	305	0	630	536	0	630	536
16	2	960	792	0	960	794	–	–	–
17	0	736	685	0	736	705	–	–	–
18	9	431	421	0	729	655	–	–	–
19	0	457	447	0	472	469	–	–	–
20	3	1100	1009	0	1500	1362	–	–	–
21	11	785	772	–	–	–	–	–	–
22	9	747	737	–	–	–	–	–	–
23	0	899	885	–	–	–	–	–	–
24	16	1680	1610	–	–	–	–	–	–
25	0	1700	1677	–	–	–	–	–	–

Bibliography

1. Keedwell A.D., Dénes J. Latin Squares and their Applications. Elsevier, 2015. 438 p. DOI: 10.1016/C2014-0-03412-0.
2. Vatutin E.I., Nikitina N.N., Manzuk M.O., Albertyan A.M., Kurochkin I.I. On the construction of spectra of fast-computable numerical characteristics for diagonal Latin squares of small order (in Russian) // Intellectual and Information Systems (Intellect – 2021). Tula, 2021. pp. 7–17.

3. Vatutin E.I., Titov V.S., Emelyanov S.G. Basics of discrete combinatorial optimization (in Russian). M.: Argamac-media, 2016. 270 p.

4. Vatutin E.I., Titov V.S., Pykhtin A.I., Kripachev A.V., Nikitina N.N., Manzuk M.O., Albertyan A.M., Kurochkin I.I. Heuristic method for getting approximations of spectra of numerical characteristics for diagonal Latin squares (in Russian) // Intellectual information systems: trends, problems, prospects. Kursk: SWSU, 2022. pp. 35–41.

5. Eduard Vatutin, Jia Wei-Ting, Jun Chi Ma, Qiang Miao. Construction of the heuristic approximations of the spectra of the number of intercalates in diagonal Latin squares of orders 10–25 // National Supercomputing Forum (NSCF – 2024). Pereslavl-Zalessky, 2024. Accepted for publication

6. Vatutin E., Belyshev A., Nikitina N., Manzuk M., Albertian A., Kurochkin I., Kripachev A., Pykhtin A. Diagonalization and Canonization of Latin Squares // Lecture Notes in Computer Science. Vol. 14389. Springer, Cham., 2023. pp. 48–61. DOI: 10.1007/978-3-031-49435-2_4.

Иванов А.М.¹, Курочкин И.И.¹²
¹НИТУ МИСИС
²ИППИ РАН
kurochkin@iitp.ru

Применение алгоритмов агрегации ответов на краудсорсинг-платформе

Как понятие, краудсорсинг (crowdsourcing) появился около 20 лет назад в 2006 году [1]. Хотя один из подвидов краудсорсинга – добровольные распределенные вычисления был известен еще в 1994-1997 годах [2]. В современном мире краудсорсинг стал одним из инструментов для обработки больших объемов данных и выполнения задач с высокой степенью параллелизма, к примеру, Toloka, Amazon Mechanical Turk, reCaptcha, CrowdFlower. Краудсорсинг-платформы предоставляют возможность распределять задачи между большим количеством участников, решения которых затем объединяются для получения финального результата. Процесс объединения и обработки результатов работы исполнителей называется агрегацией ответов. Методы и алгоритмы распределения заданий [3] и агрегации ответов [4] позволяют улучшить эффективность работы краудсорсинг-платформы.

С ростом масштабов задач, решаемых с помощью краудсорсинга, возрастает потребность в разработке методов анализа и повышения эффективности работы платформ. Использование имитационного моделирования становится одним из основных инструментов для оценки и сравнения различных подходов к организации работы платформ. Имитационное моделирование позволяет проводить эксперименты без необходимости внедрять изменения непосредственно в действующие системы. Это снижает риски и затраты, а также позволяет тестировать различные сценарии, а также методы распределения задач и алгоритмы агрегации результатов.

Имитационное моделирование также помогает анализировать влияние различных факторов на работу платформы, такие как компетентность исполнителей, сложность задач и количество выставленных меток. Это особенно важно для принятия решений в условиях неопределенности или ограниченных данных, позволяя предвидеть возможные проблемы и их решения до реализации на реальной платформе. Данный инструмент обеспечивает глубокий анализ и обоснованные выводы, которые могут быть использованы для улучшения эффективности и качества работы краудсорсинг-платформ в реальных условиях.

Цель данной работы заключается в повышении эффективности работы краудсорсинг-платформы и увеличения качества получаемых в ходе ее работы данных. Агрегация ответов играет ключевую роль в обеспечении качества и надежности решений, полученных от множества исполнителей. Алгоритмы агрегации необходимы для того, чтобы учесть различия в компетентности исполнителей и сложности задач, а также минимизировать влияние ошибок на конечный результат.

На краудсорсинг-платформах можно выделить три основных типа исполнителей:

- неопытные исполнители;
- опытные исполнители;
- фродеры (frauders) или мошенники.

Фродеры намеренно допускают большое количество ошибок, неопытные исполнители чаще допускают ошибки при решении сложных задач, а опытные исполнители почти не допускают ошибок. Различие в типах исполнителей требует использования алгоритмов, способных адаптироваться к различным уровням компетентности и минимизировать влияние недобросовестных действий фродеров [5].

Алгоритмы, учитывающие или рассчитывающие навык каждого исполнителя, такие как Honeypot и метод Дэвида-Скина (EM-алгоритм), позволяют оценивать компетентность исполнителей и корректировать вес их ответов в зависимости от их предыдущих действий на платформе. Задачи могут значительно различаться по уровню сложности. Более сложные задачи требуют более внимательного подхода, высокая сложность задач приводит к большему числу ошибок даже среди опытных исполнителей. Алгоритмы, такие как GLAD, учитывают не только компетентность исполнителей, но и сложность вопросов. Это позволяет более точно оценивать вероятность правильного ответа и улучшать общее качество агрегации.

Существуют открытые наборы данных, которые были собраны на краудсорсинг-платформе Toloka (Яндекс.Толока): relevance-2, relevance-5, nist-trec-relevance. Предварительные эксперименты на этих трех открытых наборах данных показали, что алгоритмы, учитывающие или рассчитывающие навык каждого исполнителя, такие как метод Дэвида-Скина и Honeypot, дают лучшее качество классификации по метрике F1-score. Анализ характеристик открытых наборов данных подтвердил, что количество выставленных меток и качество исполнителей оказывают значительное влияние на качество работы алгоритмов агрегации.

Анализ результатов экспериментов показал существенное влияние алгоритмов агрегации ответов на работу платформы. Что подтверждает необходимость проведения имитационного моделирования, а также сбора наборов данных для повышения эффективности работы краудсорсинг-платформ. Различные факторы, такие как компетентность исполнителей, сложность задач, количество выставленных меток, тип исполнителей и характеристики исходных данных, существенно влияют на эффективность работы алгоритмов агрегации. Предложенный подход позволяет оценить влияние различных факторов на качество и скорость выполнения задач, а также выбрать наиболее подходящие методы распределения заданий и агрегации ответов для конкретного проекта.

В ходе работы было установлено, что выбор подходящего алгоритма агрегации ответов является важным фактором, влияющим на качество выполнения задач на краудсорсинг-платформах. Использование имитационного моделирования позволяет проводить анализ и сравнение различных подходов в контролируемых условиях, приближенных к реальным. Это помогает выбирать наиболее подходящие методы для конкретного проекта, а также способно существенно упростить процесс принятия решений при проектировании и улучшении краудсорсинг-платформ.

Литература

1. Howe J. et al. The rise of crowdsourcing //Wired magazine. – 2006. – Т. 14. – №. 6. – С. 176-183.
2. Anderson D. et al. SETI@ home: internet distributed computing for SETI //Bioastronomy 99. – 2000. – Т. 213.
3. Difallah D. et al. Deadline-aware fair scheduling for multi-tenant crowd-powered systems //ACM Transactions on Social Computing. – 2019. – Т. 2. – №. 1. – С. 1-29.
4. Wang W., Zhou Z. H. Crowdsourcing label quality: a theoretical analysis //Sci. China Inf. Sci. – 2015. – Т. 58. – №. 11. – С. 1-12.
5. Raykar V. C. et al. Supervised learning from multiple experts: whom to trust when everyone lies a bit //Proceedings of the 26th Annual international conference on machine learning. – 2009. – С. 889-896.

Исенбаев А.А., Быков А.А.
*АО «Международный университет информационных технологий»
Республика Казахстан, г. Алматы, ул. Манаса, 34/1
aldiyar.issenbayev@gmail.com, bykov_a_a@list.ru*

Концептуальный анализ гибридного подхода к оптимизации транзакционной обработки потоков данных на многоядерных процессорах

Системы транзакционной обработки потоков данных (TSP) имеют критическое значение для управления данными в реальном времени и широко используются в таких областях, как IoT-аналитика, финансы и мониторинг. Это особенно заметно в «умных» энергетических сетях (smart grids), медицинских системах мониторинга и высокочастотном трейдинге, где любая задержка или несогласованность данных может привести к критическим сбоям либо финансовым потерям.

Главной проблемой здесь является обеспечение согласованности и масштабируемости при интенсивных и динамично меняющихся нагрузках. Традиционные подходы — изоляция снимков (snapshot isolation) и безблокировочные механизмы (lock-free) — лишь частично решают эти задачи. Изоляция снимков успешно поддерживает согласованность в условиях высокой частоты операций чтения, но накладывает значительные накладные расходы на управление версиями, что ограничивает её масштабируемость при высокой частоте записей [1]. Безблокировочные механизмы демонстрируют высокую производительность в условиях высокой частоты операций записи, но могут терять эффективность при преобладании операций чтения из-за увеличения затрат на повторные попытки операций [2]. Кроме того, в литературе широко обсуждаются оптимистические схемы управления конкурентным доступом [3] и мультиверсионные структуры [4], но и они зачастую сталкиваются с повышенным уровнем откатов транзакций либо большим расходом ресурсов при резких изменениях соотношения чтения и записи.

Цель данного исследования — предложить гибридный подход к управлению транзакционной обработкой потоков данных, сочетающий изоляцию снимков и безблокировочные механизмы с модулем профилирования состояния в реальном времени. Задача такого подхода — повысить производительность и масштабируемость TSP-систем на многоядерных процессорах без потери транзакционной целостности, учитывая при этом динамические колебания нагрузки (соотношение чтения/записи, всплески трафика и т.д.).

Предлагается использовать модуль профилирования, который анализирует такие ключевые параметры нагрузки, как: доля операций записи P , число конфликтов (либо откатов) в транзакциях N , средняя задержка транзакций T . Пороговое значение W указывает критическую долю записей, при превышении которой включается безблокировочный режим.

Система инициализируется с порогом W . Если $P > W$, то система переключается в безблокировочный режим (lock-free), а при меньших значениях ($P \leq W$) остаётся в режиме изоляции снимков (snapshot isolation). Так как «snapshot isolation» в большей степени ориентирован на чтение, а «lock-free» — на запись, условно их можно понимать как «чтение-ориентированные» и «запись-ориентированные» режимы.

Для минимизации затрат на переключение все активные транзакции, начатые в предыдущем режиме, завершаются по его правилам. Новые транзакции запускаются в выбранном режиме. Во время переключения возможна кратковременная задержка, которую стараются сгладить буферизацией или краткой приостановкой новых транзакций.

Модуль профилирования запускается через заданные интервалы: каждые τ секунд или после M транзакций. В этот момент система повторно измеряет долю операций записи P , число конфликтов N и другие метрики. Если $P > W$, система переключается на lock-free; в противном

случае остаётся в snapshot isolation. Такой подход обеспечивает динамическую адаптацию к нагрузкам, что особенно актуально для высококонкурентных систем, работающих в реальном времени.

Для оценки эффективности предложенного подхода можно использовать популярные тестовые стенды:

- Yahoo Streaming Benchmark (YSB): моделирует переменные нагрузки с разным соотношением операций чтения и записи. Это позволяет проверить, насколько система способна адаптироваться к изменяющимся условиям.

- TPCx-IoT: тестирует масштабируемость в условиях высокой частоты транзакций, типичной для IoT-приложений. Данные этого стенда удобны для анализа пропускной способности и откликов системы при потоковой обработке больших объёмов.

Критические метрики включают:

- Пропускная способность (throughput), отражающая, сколько транзакций обрабатывается в секунду;

- Задержки (latency), измеряющие, сколько времени уходит на завершение отдельной транзакции;

- Адаптивность, показывающую, насколько быстро система реагирует на изменения соотношения чтения/записи и перестраивается на другой режим.

Предварительный анализ свидетельствует, что гибридный подход способен удачно комбинировать сильные стороны изоляции снимков и безблокировочных механизмов, обеспечивая баланс между согласованностью и масштабируемостью. Однако остаётся ряд открытых вопросов, связанных с накладными расходами при профилировании и временем переключения между режимами. В дальнейшем планируется исследовать аспекты энергоэффективности и применения методов машинного обучения для предсказания изменений нагрузки.

Литература

1. J. Verheijde, Vassilios Karakoidas, Marios Fragkoulis, and Asterios Katsifodimos, “S-QUERY: Opening the Black Box of Internal Stream Processor State,” 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 1314–1327, May 2022, doi: 10.1109/icde53745.2022.00103.
2. G. Mencagli, Massimo Torquati, A. Cardaci, A. Fais, L. Rinaldi, and M. Danelutto, “WindFlow: High-Speed Continuous Stream Processing With Parallel Building Blocks,” IEEE transactions on parallel and distributed systems, vol. 32, no. 11, pp. 2748–2763, Nov. 2021, doi: 10.1109/tpds.2021.3073970.
3. M. J. Cahill, U. Röhm, and A. D. Fekete, “Serializable isolation for snapshot databases,” ACM Transactions on Database Systems, vol. 34, no. 4, pp. 1–42, Dec. 2009, doi: 10.1145/1620585.1620587.
4. P.-Å. Larson, S. Blanas, C. Diaconu, C. Freedman, J. M. Patel, and M. Zwillig, “High-performance concurrency control mechanisms for main-memory databases,” Proceedings of the VLDB Endowment, vol. 5, no. 4, pp. 298–309, Dec. 2011, doi: 10.14778/2095686.2095689.

Канунова Е.Е.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
E-mail: kanunovae@list.ru,*

Сравнительный анализ различных нотаций для моделирования информационных систем

Проектирование информационных систем (ИС) представляет собой важный этап в создании современных цифровых решений, обеспечивающий как их функциональную пригодность, так и соответствие бизнес-требованиям. Этот процесс включает моделирование архитектуры системы, потоков данных и бизнес-процессов, что делает его неотъемлемой частью успешной реализации ИС. В условиях стремительно развивающихся технологий, роста объёмов обрабатываемых данных и всё большей зависимости бизнеса от цифровых инструментов, актуальность темы проектирования и выбора эффективных подходов становится крайне высокой.

В докладе представлен сравнительный анализ различных нотаций для моделирования ИС с целью определения их эффективности и удобства в проектировании систем.

Выделены три наиболее популярные и востребованные в современной практике нотации, которые демонстрируют широкие возможности применения: IDEF0, BPMN, eEPC.

Основное назначение IDEF0 — представить систему в виде иерархической структуры, в которой каждый компонент связан с другими через входы, выходы, управляющие воздействия и механизмы. Это делает нотацию особенно востребованной в технических и инженерных проектах, а также в крупных предприятиях, где важно описывать и оптимизировать метапроцессы. Применение IDEF0 особенно важно в ситуациях, где необходимо четко представить структуру системы и ее основные компоненты, а также показать связи между ними.

Основная цель BPMN — это создание четких и наглядных моделей бизнес-процессов, которые могут быть использованы для анализа, улучшения и автоматизации процессов. Одной из ключевых особенностей BPMN является способность отображать логику процессов, включая развилки и параллельные ветви, что дает возможность моделировать альтернативные сценарии и варианты исполнения.

Основная цель eEPC — это создание точных и наглядных моделей, которые помогают визуализировать, как различные события влияют на выполнение действий, и как они взаимодействуют друг с другом. В отличие от других нотаций, которые могут не уделять достаточного внимания связям между событиями и действиями, eEPC делает акцент именно на этом взаимодействии, что позволяет получить более глубокое понимание бизнес-процессов.

В докладе описываются достоинства и недостатки каждой из нотаций, а также приводится визуальное представление отображений событий и действий. Представлены результаты смоделированных процессов в каждой из нотаций.

По результатам сравнительного анализа сделаны следующие выводы. Нотация BPMN является наиболее универсальной и гибкой для моделирования сложных процессов с большим количеством участников. Её интуитивная понятность и широкий выбор инструментов делают её оптимальным выбором для большинства бизнес-задач. Нотация eEPC показывает высокую эффективность в моделировании процессов, ориентированных на события, благодаря глубокой детализации взаимодействий. Однако её применение ограничивается меньшей гибкостью и сложностью освоения для пользователей с недостаточным опытом. Нотация IDEF0 сохраняет

свою актуальность в техническом и инженерном проектировании благодаря акценту на функциональной декомпозиции, но уступает современным подходам по удобству использования и поддержке динамичных процессов. Эти различия позволяют определить области применения каждой из нотаций в зависимости от специфики задач и требований проекта.

Литература

1. IDEF, EPC и BPMN: как выбрать нотацию для моделирования бизнес процессов [Электронный ресурс]. – URL: <https://education.yandex.ru/knowledge/idef-epc-i-bpmn-kak-vibrat-notatsiiu-dliamodelirovaniia-biznes-protssessov>

2. Графические нотации описания бизнес-процессов. Построение IDEF0, BPMN, EPC [Электронный ресурс]. – URL: <https://www.koderline.ru/expert/eksperty-o-produktakh-1s/article-graficheskie-notatsii-opisaniya-biznes-protssessov-postroenie-idef0-bpmn-epc/>

Канунова Е.Е., Бобров И.С.

Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
E-mail: kanunovae@list.ru,

Обучение нейросети для определения опухолей на маммографических снимках

В современном мире машинное обучение – это мощный инструмент, который помогает значительно повысить точность диагностики заболеваний и обнаружить патологии, которые могут быть пропущены при использовании традиционных методов. Оно автоматизирует рутинные процессы, освобождая время врачей для более сложных задач, и улучшает качество ухода за пациентами.

Благодаря способности обрабатывать большие объемы информации, алгоритмы машинного обучения улучшают исследовательскую работу и клиническую практику, а также позволяют разработать индивидуальные планы лечения на основе данных о каждом пациенте. С помощью анализа данных и прогнозирования, машинное обучение способствует раннему выявлению заболеваний, что повышает шансы на успешное лечение. Оно так же может снизить затраты на здравоохранение за счет оптимизации диагностических процессов, делая медицинские услуги более доступными. Кроме того, непрерывное обучение и адаптация алгоритмов машинного обучения к новым данным обеспечивает постоянное улучшение их эффективности и соответствие актуальным медицинским стандартам.

В докладе представлены несколько основных этапов обучения модели на маммографических снимках для выявления опухолей.

Сначала необходимо собрать и подготовить данные, состоящие из большого количества маммографий с соответствующими метками, которые указывают наличие или отсутствие опухолей;

Следующим шагом является предобработка изображений, включающая в себя изменение размера, нормализацию и сегментацию (рисунки 1);

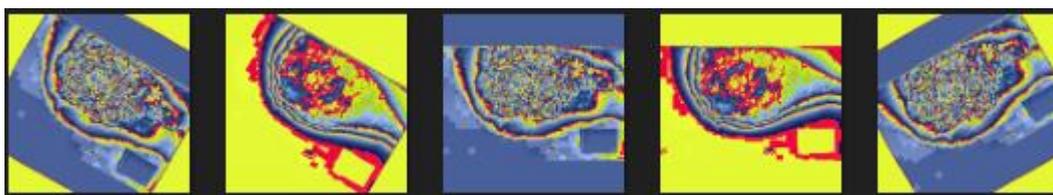


Рис. 1 – Результаты предварительной обработки снимков

Разработка модели. Для задач классификации изображений, включая обнаружение опухолей, чаще всего используются сверточные нейронные сети. Эти сети хороши для работы с 2D-данными, такими как изображения маммографических снимков. Описание архитектуры нейронной сети:

1. Входной слой:

- Нормализует изображения, преобразуя значения пикселей из диапазона [0, 255] в [0, 1] для улучшения сходимости модели.

2. Сверточные слои:

- Первый сверточный слой извлекает базовые признаки (например, края) с помощью 32 фильтров.

- Слой подвыборки уменьшает размер карты признаков.
- Второй сверточный слой с 64 фильтрами для выявления более сложных признаков.
- Еще один слой подвыборки.
- Третий сверточный слой с 128 фильтрами для извлечения высокоуровневых признаков.
- Последний слой подвыборки перед полносвязными слоями.

3. Полносвязные слои:

- Преобразует двумерные данные в одномерный вектор для дальнейшей обработки.
- Полносвязный слой с 128 нейронами и функцией активации, добавляющей нелинейность.

4. Выходной слой:

- Один нейрон с функцией активации sigmoid, который выдает вероятность принадлежности к положительному классу, что подходит для бинарной классификации.

5. Компиляция модели:

- Использует оптимизатор Adam и функцию потерь бинарной кросс-энтропии, что делает модель подходящей для бинарной классификации.

На следующем этапе происходит обучение модели. Важно разделить данные на обучающую, валидационную и тестовую выборки.

Заключительным шагом является оценка модели. После обучения важно провести тестирование и оценку производительности метриками оценки: accuracy, precision, recall и F1-score.

На рисунке 2 представлены результаты работы нейронной сети.

```
Accuracy: 0.5451
Precision: 0.5407
Recall: 0.9858
F1-score: 0.6983
```

Рис. 2 – Результат работы нейронной сети

Высокий Recall (0.9858) в контексте оценки моделей машинного обучения означает, что модель хорошо справляется с задачей выявления всех положительных примеров из набора данных. Recall измеряет, какую долю из всех реальных положительных случаев модель правильно классифицировала как положительные.

Машинное обучение представляет собой современный инструмент в области медицины, который способен не только повысить точность диагностики заболеваний, но и оптимизировать рабочие процессы в сфере здравоохранения. Особенное внимание стоит уделить процессу предварительной обработки данных, поскольку грамотно подготовленные маммографические снимки являются ключом к успешному обучению модели. Этапы, начиная от сбора и подготовки данных и заканчивая обучением и оценкой модели, формируют основу для надежной и эффективной диагностики опухолей. В итоге, внедрение технологий машинного обучения обещает значительные изменения в подходах к выявлению и лечению заболеваний, улучшая качество медицинского обслуживания и снижая на него затраты.

Литература

1. Автоматизированная обработка и анализ маммографических снимков: монография/ С.С. Садыков, Ю.А. Буланова, Е.А. Захарова; Владим. Гос. Ун-т имени Александра Григорьевича и Николая Григорьевича Столетовых. – Владимир: Изд-во ВлГУ, 2014.

2. Извлечение знаний методами машинного обучения : учебное пособие по курсам «Модели и методы инженерии знаний», «Методы машинного обучения» / А. Н. Целых, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2022. — 105 с. — ISBN 978-5-9275-4215-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/131448.html>

3. Цифровая обработка изображений: учебное пособие / Е. А. Шефер. — Санкт-Петербург: Санкт-Петербургский государственный университет промышленных технологий и дизайна, 2019. — 100 с. - <https://www.iprbookshop.ru/102493.html>

4. Цифровая обработка изображений / Гонсалес Рафаэл, Вудс Ричард ; перевод Л. И. Рубанов, П. А. Чочиа ; под редакцией П. А. Чочиа. — Москва : Техносфера, 2012. — 1104 с. — ISBN 978-5-94836-331-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/26905.html>

Колпаков А.А., Гавриков А.П., Бейлекчи Д.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
desT.087@gmail.com*

Результаты исследования производительности программной системы фильтрации акустических помех в речевом сигнале на основе рекуррентной нейронной сети

Система фильтрации речевого сигнала и обработки нейросетевых моделей является важным инструментом для исследования эффективности алгоритмов подавления акустических помех и обработки речевых данных с помощью нейронных сетей. Такая система играет крайне важную роль в самых различных приложениях и областях, включая, но не ограничиваясь следующим:

1. Распознавание речи: Системы, использующие нейросетевые модели, могут значительно улучшить точность распознавания речи, даже в условиях шума и помех. Это позволяет создавать более надежные голосовые ассистенты и повышать удобство использования голосовых интерфейсов в различных устройствах.

2. Улучшение качества звонков: В телекоммуникации такие системы используются для фильтрации фонового шума, обеспечивая чистоту и ясность голосовой передачи даже в шумных условиях, что особенно важно для мобильных звонков и конференц-связи.

3. Аудио-форензика: В области правоохранительных органов и судебных разбирательств системы фильтрации и обработки могут играть ключевую роль в идентификации и очистке записей речи для получения доказательств.

Для этого данная система должна обладать следующими функциональными возможностями:

1. Управление речевым сигналом:

- Наложение различных типов акустических помех на речевой сигнал с возможностью регулировки уровня их интенсивности;

- Применение алгоритма фильтрации для подавления наложенных помех и восстановления исходного сигнала;

- Сохранение зашумленного и отфильтрованного речевых сигналов для дальнейшего анализа.

2. Управление нейросетевыми моделями:

- Выбор и загрузка предобученных нейросетевых моделей для различных задач обработки речевых данных;

- Возможность обучения нейросетевых моделей на новых данных;

- Сохранение обученных моделей для последующего использования.

3. Мониторинг и аналитика:

- Визуализация исходных, зашумленных и отфильтрованных речевых сигналов в виде спектрограм;

- Вычисление объективных метрик качества для оценки эффективности алгоритмов фильтрации и нейросетевых моделей;

- Сохранение результатов обработки и вычисленных метрик для последующего анализа.

4. Интеграция в другие программные продукты:

- Предоставление программной библиотеки для интеграции системы фильтрации и обработки нейросетевых моделей в другие программные продукты.

Далее приведены результаты исследования производительности библиотеки Vosk для маленькой (рисунок 1) и большой (рисунок 2) моделей.

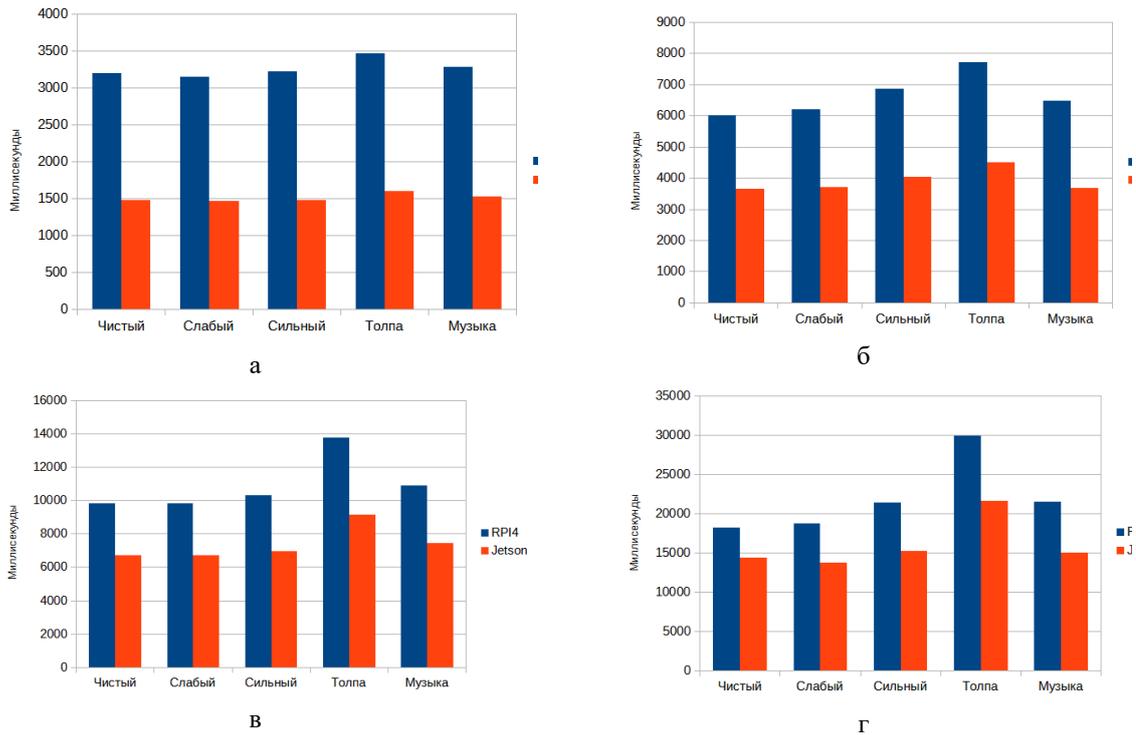


Рис. 1 – Гистограмма показателей производительности одноплатных компьютеров для библиотеки Vosk с маленькой моделью, а – аудио-семпл 5 секунд, б – аудио-семпл 15 секунд, в – аудио-семпл 30 секунд, г – аудио-семпл 60 секунд.

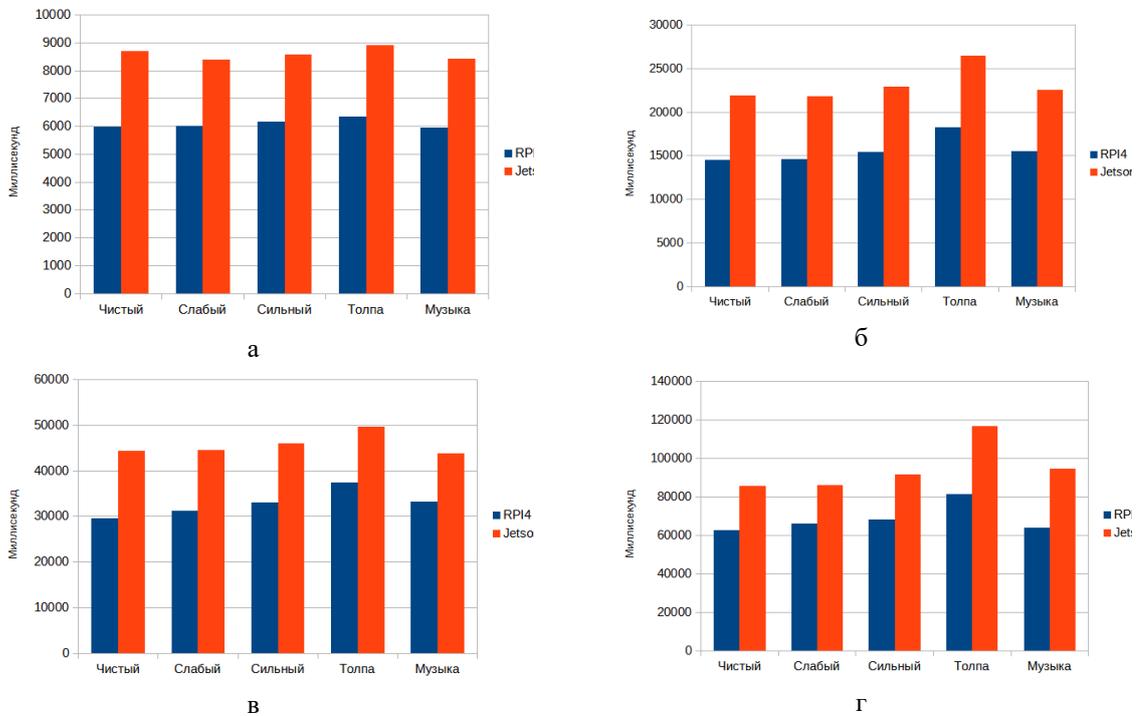


Рис. 2 – Гистограмма показателей производительности одноплатных компьютеров для библиотеки Vosk с большой моделью, а – аудио-семпл 5 секунд, б – аудио-семпл 15 секунд, в – аудио-семпл 30 секунд, г – аудио-семпл 60 секунд.

Исходя из вышеперечисленного можно сделать вывод, что при использовании библиотеки Vosk на Raspberry Pi 4 лучше загружает маленькой модель на 40% но медленнее распознаёт речь из аудио семпла, примерно, на 115% при работе с пятисекундными семплами, на 69% при работе с пятнадцатисекундными семплами, на 47% при работе с тридцатисекундными семплами и на 37% при работе с шестидесятисекундными семплами в сравнении с Nvidia Jetson Nano. В то же время, при работе с большой моделью Raspberry Pi 4 быстрее загружает модель на 12% и быстрее распознаёт речь, примерно, на 30% при работе с пятисекундными семплами, на 33% при работе с пятнадцатисекундными семплами, на 29% при работе с тридцатисекундными семплами и на 28% при работе с шестидесятисекундными семплами в сравнении с Nvidia Jetson Nano.

Литература

1. Турчин В.Н., Турчин Е.В. «Марковские цепи. Основные понятия, примеры, задачи» - Днепропетровск: ЛизуновПресс, 2021г. – 192 с.
2. Тампель И.Б., Карпов А.А. Автоматическое распознавание речи. Учебное пособие. – СПб: Университет ИТМО, 2016. – 138 с.
3. Vishniakou, V. A. Implementation of the internet of things network for monitoring audio information on a microprocessor and controller / V. A. Vishniakou, B. H. Shaya // Системный анализ и прикладная информатика. – 2022. – №. 1. – Р. 34-38. – DOI 10.21122/2309-4923-2022-1-34-38. – EDN HTZTIK.
4. Ермолаев В.А., Кропотов Ю.А., Бейлекчи Д.В. Обработка акустических сигналов методами локального анализа в телекоммуникационных системах / В.А. Ермолаев, Ю.А. Кропотов, Д.В. Бейлекчи // Радиотехнические и телекоммуникационные системы. – 2015. – № 1 (17). – С. 49-56.
5. Kropotov Yu.A. Mathematical models of telecommunication systems with acoustic feedback / Yu.A. Kropotov, A.A. Belov, A.Y. Proskuryakov, A.A. Kolpakov // 2019 International Multi-conference on Industrial Engineering and Modern Technologies, FAREASTCON 2019. – 2019. – pp. 8934203.
6. Кропотов Ю.А., Бейлекчи Д.В., Белов А.А., Ермолаев В.А., Карасев О.Е., Колпаков А.А., Коноплев А.Н., Проскуряков А.Ю. Информационно - управляющие телекоммуникационные системы аудиообмена и автоматизированного мониторинга // Радиотехнические и телекоммуникационные системы – 2014. – № 1 (13). – С. 99-113.

Комкова С.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
savicheva.svetlana@gmail.com*

Алгоритм обнаружения глаукомы на изображениях сетчатки глаза человека

В докладе рассматривается алгоритм, который позволяет выделять области, соответствующие оптическому диску и оптической чаше на сетчатке и производить на их основе диагностику наличия или отсутствия глаукомы у человека. Работа алгоритма подробно описана и проиллюстрирована, также представлены результаты проведенных экспериментов.

Глаукома является одной из основных причин потери зрения у людей. Это заболевание связано с повреждением зрительного нерва, происходящим внутри глаза [1, 2]. Потеря зрения при глаукоме развивается постепенно, зачастую пациент не осознает проблему до тех пор, пока она не перейдет в запущенную стадию. Глаукома является необратимым заболеванием, поэтому ранняя диагностика имеет критическое значение, так как при своевременном и адекватном лечении можно остановить прогрессирование болезни.

На изображении сетчатки глаза можно выделить: оптический диск, оптическую чашу и нейроретинальный край [2]. У здорового человека оптический диск имеет розово-оранжевый цвет. При появлении глаукомы он приобретает форму чаши и начинает расширяться. Оптическая чаша расположена в центре оптического диска и представляет собой выемку в зрительном нерве. Нейроретинальный край расположен между оптическим диском и оптической чашей и его форма у здорового человека не должна меняться. В результате увеличения размеров оптического диска нейроретинальный край становится тоньше. Таким образом, основными признаками глаукомы являются увеличенные размеры зрительного диска и истончение нейроретинального пояса. Поэтому точная идентификация зрительного диска и зрительной чаши играет ключевую роль в диагностике этого заболевания.

Предложенный алгоритм включает в себя несколько этапов:

1. Подготовка изображения
2. Фильтрация изображения
3. Удаление кровеносных сосудов
4. Сегментация оптического диска методом наращивания области.
5. Сегментация оптической чаши
6. Объединение результатов сегментации
7. Выявление глаукомы

Проверка работы предложенного алгоритма производилась на основе изображений, взятых из 4 баз: HRF, DIARETDB1, DRIONS-DB, Messidor. В ходе эксперимента изображения классифицировались на «здоровый» и «глаукома». Алгоритм прост с точки зрения вычислений, поэтому его можно использовать в качестве удобного инструмента в приложениях для обнаружения глаукомы.

Литература

1. Комкова С.В. Предварительная обработка изображений сетчатки глаза // Естественные и технические науки. 2016, №6(96), С. 144-146.
2. Комкова С.В. Методика маркировки кровеносных сосудов на изображениях сетчатки глаза человека // Естественные и технические науки. 2019, №10, С. 250-252.

3. Комкова С.В. Методика формирования вектора признаков по изображениям сетчатки глаза. Естественные и технические науки. 2021, №4, С. 250-252.

4. Комкова С. В. Алгоритм идентификации экссудатов на изображениях сетчатки глаза человека // Вестник компьютерных и информационных технологий. 2022. Т. 19, № 1. С. 47 – 51. DOI 10.14489/vkit.2022.01. pp.047-051 .

Кульков Я.Ю., Емельянов В.А.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
y_mail@mail.ru*

Алгоритм работы модуля распознавания углеродных цепочек на изображениях

Данные о распознавании карбина находят свое применение в различных научных и технических областях. Эти данные могут быть использованы для создания новых материалов, улучшения существующих технологий и разработки инновационных методов исследований. Кроме того, они способствуют изучению свойств карбина и его взаимодействия с другими веществами. [1-2]

Это создает необходимость разработки отдельной программы или плагина, который будет использовать алгоритмы машинного обучения и компьютерного зрения для автоматического распознавания и классификации наночастиц на изображениях, полученных с помощью электронных микроскопов. Плагин должен обеспечивать высокую точность распознавания и быть адаптированным к различным условиям и типам наноматериалов. [3-5]

Для решения этой задачи была выбрана нейронная сеть, которая выполняет как обнаружение объектов, так и их сегментацию. В процессе исследования была выбрана архитектура YOLO, признанная одной из самых эффективных на сегодняшний день. В качестве обучающих данных использовались изображения, полученные с электронного микроскопа, которые были разделены на категории для обучения обнаружению частиц и связей между ними.

Каждая категория делилась на обучающий и тестовый наборы. Затем на обучающих наборах были выделены объекты, необходимые для обнаружения, и результаты разметки были загружены в нейронную сеть. Обучение проводилось с использованием языка Python и библиотеки ultralytics на облачных вычислительных ресурсах. Для проверки использовалась встроенная функция predict. Полученная модель смогла выделить значительную часть частиц, хотя возникли проблемы с ложными срабатываниями и выделением частиц диаметром до 10 пикселей, что можно исправить, увеличив количество таких изображений в обучающем наборе.

Разработанный плагин включает два основных метода — для поиска частиц и связей между ними и может использоваться через графический интерфейс программы ImageJ. Обмен данными об изображении осуществляется с помощью классов ImagePlus, Roi и RoiManager из библиотеки ImageJ API. Для вывода пользовательского интерфейса используется класс GenericDialog. Библиотека ONNX Runtime отвечает за работу с нейросетевыми моделями, конвертированными из формата YOLO в формат ONNX, так как для использования оригинального формата нет библиотек для языка Java. Для вычисления контуров масок связей применяется библиотека OpenCV. Если программа не обнаружит объектов, будет выведено соответствующее сообщение, а в случае обнаружения объектов появится окно «ROI Manager» с перечислением всех объектов по вероятности соответствия. При поиске связей для некоторых объектов могут быть предложены несколько вариантов масок.

Литература

1. Yan He, Huakai Xu, Gang Ouyang, Guowei Yang. Thermal properties of carbyne nanostructures, Results in Physics, Volume 34, 2022, 105311, ISSN 2211-3797.

2. Guowei Yang. Synthesis, properties, and applications of carbyne nanocrystals. *Materials Science and Engineering: R: Reports*, Volume 151, 2022, 100692, ISSN 0927-796X.

3. Автоматизированный метод распознавания наночастиц алюминия, осажденных в вакууме на кремниевые подложки / В. К. Гончаров, А. А. Пехота, М. В. Пузырев, В. Ю. Ступакевич // *Материалы и структуры современной электроники : Материалы VIII Международной научной конференции*, Минск, 10–12 октября 2018 года / Редколлегия В.Б. Оджаев [и др.]. – Минск: Белорусский государственный университет, 2018. – С. 331-334.

4. L. Zhao, R. He, K. T. P. Wang, and K. S. Choi, “Automated recognition of carbon nanotubes in transmission electron microscopy images,” *Nano Research* 3, 764-771 (2010).

5. H. Zheng, Z. Wang, Y. Liu, L. Yang, and H. Shen, “A novel method for automated recognition of carbon nanostructures in transmission electron microscope images,” *Carbon* 50, 181-192 (2012).

Куручкин И.И., Прун. А.И., Путилина Е.В.
*Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации им.А.А. Харкевича
Российской академии наук
127051, г. Москва, Большой Каретный переулок, д.19, стр.1
kurochkin@iitp.ru*

Децентрализованное распределенное обучение глубоких нейронных сетей с динамическими графами обмена данными

Решение различных прикладных задач с помощью современных архитектур глубоких нейронных сетей в последние годы является актуальным и эффективным подходом. Обучение глубоких нейронных сетей на больших датасетах с использованием одного вычислительного устройства, даже с хорошим графическим ускорителем, далеко не всегда возможно за адекватное время. Это означает, что для обучения глубоких нейронных сетей требуются высокопроизводительные многопроцессорные вычислительные системы. Не смотря на увеличение количества кластерных систем и систем с архитектурой MPP (Massive Parallel Processing), дефицит высокопроизводительных систем только увеличивается. В этом случае можно использовать распределенные системы.

Распределенные вычислительные системы позволяют объединить для вычислений разнородные вычислительные ресурсы. К основным особенностям распределенных вычислительных систем можно отнести: большие возможности по масштабированию, устойчивость к ошибкам и выходу из строя вычислительных узлов и каналов передачи данных. Для обеспечения данных особенностей требуется особая организация процесса передачи данных и вычислений.

С помощью распределенных вычислительных систем достаточно эффективно можно решать задачи, которые можно разделить на множество небольших автономных подзадач. Такой тип задач называют задачами с разделением по данным или «bag of tasks». Для успешного решения некоторых задач необходимо гарантированное вычисление всего множества подзадач. В этом случае, для увеличения надежности каждая подзадача реплицируется на несколько вычислительных узлов распределенной системы. С одной стороны существенно повышается надежность вычисления подзадачи и вероятность верификации результата, но с другой стороны уменьшается вычислительные возможности распределенной системы. Некоторые типы задач, такие как распределенное глубокое обучение, могут быть разбиты, с некоторыми ограничениями, на независимые подзадачи и при этом не требуют решения всего множества подзадач. Для успешного решения таких задач достаточно 50-90% полученных результатов. Это позволяет убрать репликацию подзадач и не снижать вычислительную возможность распределенной системы.

В настоящее время существует множество подходов и методов распределенного глубокого обучения [1], [2]. С помощью некоторых методов можно достаточно эффективно реализовать распределенное обучение только на гомогенных вычислительных системах, к примеру, на ресурсах приватного облака [1]. Но в случае, когда распределенная вычислительная система имеет высокую степень гетерогенности необходимо не только использовать определенные методы распределенного обучения, но и адаптировать их к особенностям распределенной системы.

К основным способам адаптации можно отнести:

- Формирование локальных датасетов в соответствии с ресурсами вычислительных узлов;

- Аугментация «на лету» на стороне вычислительного узла для минимизации передачи данных;
- Многократное использование данных, которые уже есть на узле;
- Настройка гиперпараметров локального глубокого обучения на вычислительном узле в рамках одной эпохи распределенного обучения;
- Методика синхронной или асинхронной агрегации результатов с учетом возможного отсутствия части результатов;
- Настройка системы выдачи заданий (балансировки нагрузки)[3].

Распределенное глубокое обучение является итеративным процессом. Одна итерация или эпоха распределенного обучения с параметрическим сервером (глобальной моделью) состоит из нескольких этапов:

1. Формирование локальных датасетов для вычислительных узлов распределенной системы;
2. Передача актуальных (текущих) весов глобальной модели и локальных датасетов на вычислительные узлы;
3. Проведение одной или нескольких итераций локального глубокого обучения на вычислительных узлах;
4. Передача градиентов весов обратно на параметрический сервер;
5. Обновление весов глобальной модели на основе полученных агрегированных результатов.

В случае децентрализованного обучения (без параметрического сервера) некоторые этапы изменяются, но остаются одинаковыми для каждого вычислительного узла:

1. Передача/запрос данных для формирования локального датасета;
2. Формирование локального датасета на вычислительном узле распределенной системы;
3. Проведение одной или нескольких итераций локального глубокого обучения на вычислительном узле;
4. Передача градиентов весов другим узлам децентрализованной распределенной системы;
5. Обновление весов локальной модели, на основе переданных на этот узел градиентов.

При передаче градиентов другим узлам в рамках децентрализованного распределенного обучения формируется граф передачи данных. Граф передачи данных может иметь постоянную сетевую топологию, к примеру, дерево или кольцо. Но для децентрализованной распределенной системы может быть характерно временное незапланированное отключение некоторых вычислительных узлов, а также нарушение работы каналов передачи данных. В этом случае граф передачи данных будет постоянно изменяться. Однако можно определить основные принципы его построения для своевременного обмена текущими результатами в децентрализованной распределенной системе [4]. Если для организации распределенной системы используется платформа BOINC [5], то можно использовать BOINC-сервер, как параметрический сервер или как транзитный узел для передачи градиентов от одного узла распределенной системы другому. Одной из главных причин наличия одного или нескольких транзитных узлов является возможное отсутствие непосредственной связи между различными узлами децентрализованной распределенной системы. Использование динамического графа передачи данных для распределенной системы на платформе BOINC обусловлено особенностью системы выдачи заданий – задания запрашиваются узлами, а не рассылаются сервером. Использование динамического графа передачи данных позволяет уменьшить время пересылки градиентов через транзитные узлы и дает возможность использовать актуальные градиенты без большой задержки.

Графы передачи данных являются ориентированными графами. Одним из важных свойств ориентированного графа при решении задачи децентрализованного глубокого обучения в ориентированном графе передачи данных является наличие циклов. Для эффективного обмена градиентами в процессе распределенного обучения каждая вершина (узел) должна быть включена в один или несколько циклов, при этом необходимым условием является достижимость любой вершины графа из произвольной вершины. Следует учесть, что наличие длинных циклов может увеличить количество эпох распределенного обучения при использовании больших датасетов.

Литература

5. Ben-Nun, T. and Hoefler, T. (2019). "Demystifying Parallel and Distributed Deep Learning: An In-depth Concurrency Analysis". *ACM Computing Surveys*, 52(4), pp. 1–43.
6. Langer, M. et al. (2020). "Distributed Training of Deep Learning Models: A Taxonomic Perspective". *IEEE Transactions on Parallel and Distributed Systems*, 31(12), pp. 2802–2818.
7. Kurochkin I., Kondrashov N. Comparison of Various Algorithms for Scheduling Tasks in a Desktop Grid System Using a ComBoS Simulator // *Communications in Computer and Information Science*, Vol. 1304. Springer, Cham., 2020. – С. 29-40.
8. Раппопорт А.М., Гнеденко Л.С., Курочкин И.И. Построение коммуникационной сети с заданным доминирующим множеством и максимальной кратчайшей цепью // *Высокопроизводительные вычислительные системы и технологии*, Т.4, №2, 2020. С. 53-58.
9. Anderson D. P. BOINC: a platform for volunteer computing // *Journal of Grid Computing*. Т. 18, №1, 2020. – С. 99-122.

Мареев А.В.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирской обл., ул. Орловская, 23
a.v.mareev@yandex.ru*

Применение самообучающейся нейронной сети для локализации железнодорожных колёс в видеопотоке

В современном мире технологии искусственного интеллекта (ИИ) находят применение в различных сферах деятельности человека. Одной из таких сфер является железнодорожный транспорт, где использование нейронных сетей может значительно повысить эффективность и безопасность работы.

Одной из задач, решаемых с помощью нейронных сетей, является локализация железнодорожных колёс в видеопотоке [1]. Это может быть полезно для контроля качества на производстве.

Для решения этой задачи можно использовать самообучающуюся нейронную сеть. Такая сеть способна адаптироваться к изменениям условий эксплуатации и улучшать точность локализации с течением времени.

Процесс локализации железнодорожных колёс с помощью самообучающейся нейронной сети включает в себя несколько этапов:

1. Подготовка данных: необходимо собрать достаточное количество видеозаписей с производства железнодорожных колёс. Видеозаписи должны содержать чёткие изображения колёс и быть разнообразными по условиям съёмки (освещение, фон, угол обзора и т. д.).

2. Выбор архитектуры нейронной сети: для решения задачи локализации колёс можно использовать различные архитектуры нейронных сетей, такие как свёрточные нейронные сети (CNN) или рекуррентные нейронные сети (RNN). Выбор архитектуры зависит от требуемой точности, скорости работы и доступных вычислительных ресурсов.

3. Обучение нейронной сети: на этом этапе нейронная сеть обучается на подготовленных данных. В процессе обучения сеть постепенно улучшает свои прогнозы, адаптируясь к изменениям в данных.

4. Тестирование нейронной сети: после обучения нейронная сеть тестируется на новых данных, чтобы оценить её точность и эффективность.

5. Применение нейронной сети: после успешного тестирования нейронная сеть может быть применена для локализации колёс в реальном времени.

В данной работе рассматривается пример использования инкрементного обучения нейросети YOLO[2]. В настоящее время в архитектуру нейросети YOLO были внедрены следующие усовершенствования для инкрементного обучения:

– Динамическая настройка архитектуры: новые модели YOLO включают динамическую настройку архитектуры, которая позволяет им адаптировать свою структуру в зависимости от поступающих потоков данных. Такая гибкость крайне важна для непрерывного обучения, поскольку позволяет модели расширять свои возможности без переобучения с нуля.

– Методы управления памятью: в архитектуры YOLO были интегрированы передовые методы управления памятью, такие как воспроизведение опыта и выборочное

сохранение данных в памяти. Эти методы помогают модели сохранять важные характеристики предыдущих задач во время обучения новым, эффективно предотвращая катастрофическое забывание.

– Стратегии регуляризации: в последних версиях YOLO используются сложные стратегии регуляризации для поддержания производительности при выполнении различных задач. Такие методы, как эластичная консолидация весов (EWC) и прогрессивные нейронные сети, изучаются для того, чтобы веса модели корректировались таким образом, чтобы сохранялась ранее изученная информация.

Самообучающиеся нейронные сети могут значительно повысить эффективность и безопасность железнодорожного транспорта. Они способны адаптироваться к изменениям в данных и улучшать качество своих прогнозов с течением времени, что делает их перспективным инструментом для решения различных задач в этой области.

Однако стоит отметить, что применение нейронных сетей в железнодорожном транспорте требует тщательного тестирования и оптимизации. Необходимо учитывать различные факторы, такие как условия съёмки, освещение, фон и т. д., чтобы обеспечить высокую точность и эффективность локализации колёс.

Литература

1. ГОСТ 10791–2011. Колеса железнодорожные. Общие технические условия. — Москва: Стандартинформ.
2. YOLO Documentation. You Only Look Once: Unified, Real-Time Object Detection. URL: <https://docs.ultralytics.com>. Дата обращения: 2024.

Платонова А.С., Рыжкова М.Н.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264 г. Муром, Владимирской обл., ул. Орловская, д. 23
allaplatoнова@inbox.ru*

Разработка онлайн-курса по теории информации

Рынок онлайн-образования очень разнообразен – это массовые открытые онлайн-курсы (такие как Coursera, EdX, Udacity, Udemy), образовательные платформы широкого профиля, b2b-решения (в первую очередь, различные LMS), обучающие мобильные приложения и компьютерные игры, инструменты для видеоконференций, виртуальные уроки с преподавателями и многое другое. Например, «Яндекс. Практикум», «Stepik» «Codecademy», «Data Camp», «Coursera», «Проект тренажер», «Skillbox», «Khan academy». В большинстве из них содержится теоретический материал для изучения общеобразовательных предметов и учебных дисциплин, есть возможность проверить знания с помощью тестов, присутствует быстрая обработка результатов, предлагаются образцы решения, есть модули статистики.

Использование для проверки умений решать учебные задачи тестовых заданий с вводом ответа или выбором правильного варианта может быть недостаточным для качественного усвоения учебного материала. Примеры и задачи, решаемые в рамках учебных дисциплин, могут быть достаточно сложными и трудоёмкими. Сложность задач может увеличиваться по мере изучения разделов, например, как в учебной дисциплине «Теория информации», начиная с заданий по вычислению количества полученной информации и заканчивая кодированием. Поэтапное решение задач – это последовательное выполнение шагов для нахождения решения, которое позволит обучающемуся понять структуру решения, увидеть прогресс, сосредоточиться на выполнении конкретных действий, что в результате может повысить учебную мотивацию, производительность и результативность. Анализ онлайн-курсов с точки зрения поэтапного решения учебных задач показал недостаточное наличие такой возможности.

Разработка онлайн-курса по дисциплине «Теория информации» была начата с проектирования структуры курса. Для этого были проанализированы соответствующие учебные и учебно-методические пособия: «Основы теории информации» автор О.В. Мотовилова, «Теория информации и кодирования» автор Д.Е. Чикрин, «Теория информации и кодирования» автор Е.В. Гошин, «Теория информации» автор В.В. Лидовский, «Теория информации» авторы И.В. Блинова, В.В. Котенко. Основные разделы, из которых состоит курс дисциплины «Теория информации», следующие: предмет теории информации, основные понятия и определения; энтропия системы, энтропия сложной системы; количественные меры Хартли и Шеннона; скорость передачи информации и пропускная способность дискретного канала без помех и с помехами; эффективное кодирование информации Шеннона-Фано и Хаффмана, арифметическое кодирование информации, помехоустойчивое кодирование, коды Хемминга, циклические коды. В указанных выше учебных пособиях по каждому из разделов были подобраны типовые задачи, которые затем представлялись последовательностью более мелких шагов [1].

Литература

1. Платонова А.С. Разработка онлайн-тренажера по теории информации // Радиотехнические и телекоммуникационные системы. 2023. № 2. С. 46-55.

Панкратов Д.А., Андрианов Д.Е.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
denis_pankratov2000@mail.ru*

Применение топологической декомпозиции для нахождения признаков при классификации изображений

Задачи классификации являются одними из основных в области машинного обучения и анализа данных. Они находят применение в различных сферах, таких как медицина, финансы, производство, безопасность и многих других. Однако при решении данных задач возникают проблемы подбора значимых признаков изображений для из дальнейшей классификации.

Для нахождения подобных признаков предлагается использование концепции яркостного разложения изображения по топологическим особенностям, а именно метода разложения изображения на матрицы специального типа (компоненты) [1-2].

Основная суть метода лежит в просмотре пикселей изображения по мере убывания яркости, при нахождении данного значения яркости на изображении создаётся компонента, повторяющая размеры изображения и первоначально заполненная нулями, у которой возрастает значение яркости в рассматриваемой точке. Нумерация компонент происходит в порядке их появления. Изображения рассматриваются слева направо и сверху вниз.

Если рассматриваемая точка на изображении находится слева/справа/сверху/снизу от уже созданной компонентой, то эта точка будет относиться к данной компоненте, у которой в рассматриваемых на изображении координатах значение увеличится на единицу.

При соприкосновении нескольких компонент происходит поглощение одной компоненты другой, приоритет отдаётся компоненте с наибольшим количеством пикселей. При этом значения поглощаемой компоненты в ненулевых координатах становятся равными единице.

При рассмотрении следующей яркости изображения, ненулевые значения компонент увеличиваются на единицу.

Также в ходе исследования было обнаружено, что практической значимостью помимо поглощённых компонент также имеют поглощающие.

После нахождения компонент для определения характеристик изображения они разбиваются на кластеры при помощи метода k-средних, где для расчёта евклидова расстояния используются начало и конец существования компонент. Кластеризация выполняется отдельно для поглощённых и поглощаемых компонент.

В ходе тестирования различных признаков, выделенных при помощи топологической декомпозиции были выделены такие характеристики, как количество классов компонент, количество классов с поглощёнными и поглощаемыми компонентами, а также данные количества, где компонент больше одной, а также количество поглощённых компонент с разным количеством пикселей, и количество поглощаемых компонент с различным временем существования.

В качестве примера применения топологической декомпозиции для нахождения признаков была выбрана задача классификации изображений криогелей поливинилового спирта (КГПВ). Криогели образуются в результате криогенной обработки (замораживание - выдерживание в замороженном состоянии- оттаивание) концентрированных растворов этого полимера в воде или диметилсульфоксиде (ДМСО) и имеют макропористую гетерофазную морфологию.

Большое воздействие на структуру криогеля оказывает температура, так выявлены немонотонные изменения общего характера текстуры и размеров её элементов в зависимости от режимов замораживания/вымораживания. На характеристики вещества помимо режима и способа криогенной обработки также оказывает влияние исходная концентрация поливинилового спирта. Решение данной задачи осложнено наличием схожей текстурой изображения при различных параметрах получения образцов [3-4]. В качестве алгоритма классификации использовался метод опорных векторов. Обучение проводилось на 20 классах. В ходе экспериментов удалось добиться точности распознавания классов криогелей поливинилового спирта в 77 процентов.

При попытке классификации изображений криогелей без выделения признаков с использованием архитектуры нейронной сети ResNet удалось добиться точности в 60 процентов.

Литература

1. Еремеев, С.В. Алгоритм совмещения пространственных объектов разномасштабных карт на основе топологического анализа данных / С.В. Еремеев, Д.Е. Андрианов, В.С. Титов // Компьютерная оптика. – 2019. – Т. 43, № 6. – С. 1021-1029. – DOI: 10.18287/2412-6179-2019-43-6-1021-1029.
2. Еремеев, С.В. Метод разложения изображения по топологическим признакам / С.В. Еремеев, А.В. Абакумов, Д.Е. Андрианов, Д.В. Титов // Компьютерная оптика. – 2022. – Т. 46, № 6. – С. 939-947. – DOI: 10.18287/2412-6179-СО-1080.
3. Лозинский В. И., Дамшкалн Л. Г., Шаскольский Б. Л. [и др.] Изучение криоструктурирования полимерных систем. 27. физико-химические свойства криогелей поливинилового спирта и особенности их макропористой морфологии // Коллоидный журнал. – 2007. – Т. 69. – № 6. – С. 798-816. – EDN IBGVJJ4.
4. Лозинский В. И., Дамшкалн Л. Г., Курочкин И. Н., Курочкин И. И. Изучение криоструктурирования полимерных систем. 28. физико-химические свойства и морфология криогелей поливинилового спирта, сформированных многократным замораживанием-оттаиванием // Коллоидный журнал. – 2008. – Т. 70. – № 2. – С. 212-222. – EDN IJKFWF.

Стрыгин Д.А., Андриевская Е.А.

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

г. Томск, пр. Ленина, 36

d.strygin@mail.ru; ja.konovalova@gmail.com

Финансово устойчивая модель работы искусственного интеллекта: анализ использования больших языковых моделей в образовательных системах

В условиях растущих финансовых затрат на использование искусственного интеллекта (ИИ) в образовательных системах становится важным подход к оптимизации операционных расходов. Одним из решений является внедрение смешанного подхода, сочетающего как платные, так и бесплатные большие языковые модели (LLM), что позволяет снизить общие расходы на ИИ-услуги, сохраняя при этом качество и точность предоставляемых услуг [1]. В данной работе рассматриваются возможности реализации такого подхода в образовательных платформах.

Образовательные системы, использующие ИИ, сталкиваются с несколькими важными вызовами, связанными с высокой стоимостью использования платных языковых моделей [2]. Платные модели, такие как GPT-4, имеют высокую вычислительную нагрузку и требуют значительных финансовых затрат на каждом запросе, что может быть нецелесообразным при массовом использовании. В то же время бесплатные модели, например, версии GPT-2 или другие открытые LLM [3], обладают ограниченными возможностями, но могут быть эффективно использованы для решения менее сложных задач.

Основным преимуществом смешанного подхода является возможность динамически выбирать модель в зависимости от сложности запроса. Например, для базовых и стандартных запросов можно использовать бесплатные модели, что существенно снижает затраты на вычисления и инфраструктуру. При этом для более сложных задач, таких как анализ текста с высоким уровнем сложности или генерация специализированного контента, подключаются платные модели, что обеспечивает необходимое качество и точность.

Кроме того, такой подход позволяет оптимизировать использование ресурсов, распределяя нагрузку между моделями в зависимости от текущей ситуации. Использование бесплатных моделей в период низкой нагрузки на систему или для предварительных этапов обработки запросов, например, для фильтрации или предварительной классификации данных, дает возможность уменьшить количество обращений к более дорогим платным моделям.

Финансовая составляющая также играет важную роль в обеспечении устойчивости образовательных платформ. Платежи за использование платных языковых моделей часто основаны на объеме обработанных данных или числе запросов, что может привести к значительным затратам при массовом использовании. Смешанный подход позволяет более эффективно планировать и контролировать расходы, выделяя ресурсы только на те задачи, которые действительно требуют применения более мощных моделей.

В заключении следует отметить, что использование смешанного подхода в образовательных системах позволяет существенно снизить операционные расходы при сохранении высокого качества обслуживания пользователей. Такой подход является эффективным решением для образовательных платформ, стремящихся сбалансировать экономическую целесообразность и высокое качество образовательных сервисов, используя как бесплатные, так и платные модели в зависимости от сложности задачи.

Литература

1. Smith J. Optimization Strategies for Large Language Models // AI and Education. 2023. № 4. С. 45-53.
2. Brown T., Mann B. Scalable AI Systems in Education // Advances in NLP. 2022. №3. С. 12-18.
3. OpenAI Research Team. Cost-efficient AI: A Guide for Educators // White Paper. 2024.

Холкина Н.Е.

*Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
E-mail: holnataly@mail.ru*

Об опыте организации промежуточного контроля знаний в процессе изучения дисциплин линии «Программирование»

На разных ИТ – специальностях эти дисциплины называются по-разному, но по сути речь идет о систематизации знаний языка программирования, получении опыта решения стандартных задач, изучении особенностей организации базовых структур данных, и др. В сумме эти дисциплины изучаются на первом и втором курсе и завершают базовую подготовку студентов по языкам и алгоритмам.

Ровных групп не существует, ребята приходят с очень разной подготовкой, кто-то в олимпиадах участвовал, а несколько человек язык программирования впервые видят. Рассматривая статистику поступления можно отметить что число таких ребят не превышает 5-10%, но не учитывать такую подготовку первокурсников нельзя.

Все дисциплины цикла «Программирование» имеют ярко выраженную практическую направленность, поэтому за первый семестр в среднем на специальности студент решает 50-60 задач, во втором семестре по мере усложнения их число сокращается примерно в двое и на этом уровне остается и во втором и третьем и в четвертом семестрах. Контроль решения задач это один из видов промежуточного контроля. Здесь помощником может выступать тестовая система с хорошим набором тестовых данных.

Несмотря на практическую направленность дисциплин проверять знания теории надо. Проведение письменного опроса плохая идея – если задания одинаковые сложно предотвратить списывание, подготовить, напечатать и проверить большое количество различных вариантов очень трудозатратно. Выходов из ситуации несколько. Первый подход подготовить задания с коротким ответом, и давать задания на проверку знаний в высоком темпе. Этот подход годится на этапах обучения, когда требуется запомнить много новых терминов и имеет смысл если эта новая терминология будет использована на текущем занятии далее. Время отводимое на такую проверку знаний не должно превышать 10-15 минут. Особенно на первых занятиях не должно быть сложных заданий. За экватором курса можно добавлять сложные или неожиданные задания, но не более одного за опрос. Появление таких заданий снижает темп опроса и оставляет простор списываниям.

Второй подход: использование тестового опроса. Но этот подход не имеет особых достоинств перед описанным ранее, кроме одного: темп прохождения теста для студента комфортный. Но на этом преимущества и заканчиваются. Задания на написание небольших фрагментов код сложно проверить в автоматизированном режиме, но читать гораздо удобнее чем написанные от руки. Так же как и в предыдущем случае нет оценки сразу, что снижает ценность машинного теста.

Один из важных моментов любых коротких опросов – анализ результатов. Анализировать ответы на предложенные вопросы на следующем занятии не имеет смысла: за одну-две недели студент успеет забыть задания, и затраты времени на разбор ошибок будут не меньше чем на сам опрос, да и ценность такого разбора низкая, если отрабатывать эту теорию планируется на

текущем занятии, а значит проводить анализ возможных вариантов решения поставленных задач нужно сразу, после того как работы будут сданы: задания и их решение студенты ещё помнят.

Что же дают студенту и преподавателю такие контрольные опросы? Студент, зная о проводимом контроле готовится к занятию. Он сам может оценить уровень своих знаний, зная максимальный балл теста, свой балл и средний балл группы. Преподаватель видит, кто из студентов готовится к занятиям. Видит качество подготовки в динамике. Знает темы, материал по которым усвоен хорошо, и темы, которые необходимо рассмотреть повторно.