

Разработка инструмента оценки качества работы алгоритмов разметки медицинских изображений

ГБУЗ «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» создан в 1996 году и является ведущей организацией в России в области развития и повышения эффективности службы лучевой и инструментальной диагностики. Осуществляет научную, технологическую, учебную, организационную и методологическую поддержку рентгенологов, радиологов и врачей функциональной диагностики. Центр ведёт научно-исследовательские и практические работы в области искусственного интеллекта в здравоохранении и телемедицины на национальном и международном уровнях.

АКТУАЛЬНОСТЬ ЗАДАЧИ

На сегодняшний день здравоохранение является одной из перспективных сфер для разработки решений и технологий, работающих на основе искусственного интеллекта (ИИ). Уже за последние годы в клиническую практику были внедрены многие инновационные продукты.

- Системы распознавания речи помогают врачам заполнять медицинскую документацию и взаимодействовать с интерфейсом медицинских информационных систем.
- Чат-боты позволяют регулярно взаимодействовать с пациентами, например напоминать о приёме лекарств и самостоятельном проведении диагностических манипуляций (измерение артериального давления, уровня сахара крови и т.п.), уточнять состояние здоровья пациента, записывать и сохранять результаты измерений.
- Автоматический анализ Big data позволяет определять на основе лабораторных анализов предикторы и риски возникновения заболеваний.

- Автоматический анализ медицинских изображений, например маммограмм и томограмм, позволяет выявлять патологии, проводить приоритизацию пациентов по тяжести состояния и в целом облегчает врачу-рентгенологу интерпретацию исследований.

В конце 2019 года в Москве стартовал эксперимент по использованию технологий компьютерного зрения, работающий на основе ИИ, для анализа медицинских изображений. Предполагалось, что разработанные решения смогут автоматически анализировать и выявлять признаки патологии на диагностических изображениях. И уже сейчас внедрены ИИ-сервисы, помогающие врачам в диагностике COVID-19, рака лёгкого, рака молочной железы, туберкулёза. В ближайшем будущем запланировано расширение эксперимента и внедрение сервисов для диагностики заболеваний сердечно-сосудистой системы, желудочно-кишечного тракта, центральной нервной и опорно-двигательной систем.

ОПРЕДЕЛЕНИЕ ПРОБЛЕМЫ

Перед тем как внедрить сервис, работающий на основе ИИ, в рутинную клиническую практику, его необходимо протестировать на техническую готовность, а также проверить, насколько он соответствует заявленным характеристикам диагностической точности. Это называется аналитической валидацией алгоритма. Прошедшие данную проверку ИИ-сервисы допускаются к интеграции с медицинскими информационными системами, в том числе городскими системами здравоохранения.

Сейчас аналитическая валидация проводится вручную. Это достаточно трудоёмкий процесс. К тому же неавтоматизированная валидация допускает случайные или намеренные отклонения от утверждённого алгоритма тестирования, позволяет манипулировать наборами данных, а также может потенциально ставить в неравные условия участников тестирования.

ЗАДАЧА

Мы предполагаем, что возможно разработать единые подходы для автоматической аналитической валидации алгоритма и набора метрик для последующего сравнения с эталонным набором данных (результат экспертной разметки).

В рамках хакатона командам предлагается разработать инструмент, который будет оценивать качество разметки медицинских изображений ИИ-сервисами. Он должен автоматически определять набор метрик, а результаты его работы должны наиболее точно совпадать с мнением врача-эксперта.

ЧТО ПРЕДОСТАВИМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ

Командам будут предоставлены рентгенограммы органов грудной клетки (подвыборка из датасета [NIH Chest X-rays](#)), содержащие как патологию, так и «норму» (Origin), и наборы разметок. В наборах будет представлена разметка врача-эксперта (Expert) и трёх интеллектуальных алгоритмов (sample_1, sample_2, sample_3).

Также врач-эксперт оценил корректность работы интеллектуальных алгоритмов по 5-ти бальной шкале (OpenPart.csv), где 5 – полностью согласен с разметкой интеллектуального алгоритма, 1 – полностью не согласен. Для усложнения задачи часть оценок была скрыта.

Структура данных:

- Корневая папка – Dataset
 - DX_TEST_RESULT_FULL.csv – описание находок
 - Origin – входные изображения
 - Expert – разметка врача-эксперта (Ground Truth)
 - sample_1 – первая разметка для сравнения
 - sample_2 – вторая разметка для сравнения
 - sample_3 – третья разметка для сравнения
 - OpenPart.csv – оценка корректности работы интеллектуальных алгоритмов

Разметка предоставляется в двух форматах для простоты использования:

- Описание находок в виде набора прямоугольников и эллипсов. Описание для прямоугольников и эллипсов (DX_TEST_RESULT_FULL.csv):
 - Тип объекта Shape, бывает Rectangle – прямоугольник и Circle – эллипс.
 - Эллипс имеет ориентацию параллельно осям.

- Координаты центра x , y : $xcenter$, $ycenter$.
- Половина ширины и половина высоты области $w/2$, $h/2$: $rhorizontal$, $rvertical$.
- Сегментационная карта находок. Отрисованные значения файла `DX_TEST_RESULT_FULL.csv`. Файлы в формате `png`. При этом если значение пикселя 0 — область не аномальная, если значение 255 — аномальная.

КРИТЕРИИ ОЦЕНКИ РАЗРАБОТАННЫХ РЕШЕНИЙ

ОБЩИЕ ПРАВИЛА

- Не допускается использование исходников с закрытой лицензией.
- Не разрешается использовать закрытые данные от прочих участников (возможность предоставления своих данных другим участникам следует обсуждать с жюри).
- Если вы используете при разработке свои наработки, то они должны быть опубликованы и описаны.
- Запрещается ручная разметка датасета (для обучения, в качестве ответа и т.д.).

Оценивать предложенное решение мы будем по трем основным критериям:

1. Точность предсказания.
2. Сопроводительная документация.
3. Программный код.

ТОЧНОСТЬ ПРЕДСКАЗАНИЯ

Точность предсказания — это основной критерий, по которому будет определен победитель. Ведь в дальнейшем именно точность предсказания поможет оценить качество работы ИИ-сервиса и принять решение о его внедрении в рутинную практику.

СОПРОВОДИТЕЛЬНАЯ ДОКУМЕНТАЦИЯ

В сопроводительной документации необходимо максимально подробно описать гипотезу и алгоритм работы вашего решения. Что ещё важно отразить в документации:

- Должны быть описаны все применяемые методы. Если использовались готовые реализации — вам необходимо указать ссылку на этот метод.

- В случае проведения процесса обучения — он должен быть строго описан.
- Если вы будете использовать модифицированные классические методы оценки качества (mAP, IoU, Dice и т.п.) — все изменения должны быть описаны.

Главная задача сопроводительной документации — предоставить возможность воспроизведения вашего решения специалистами нашего Центра!

ПРОГРАММНЫЙ КОД

К программному коду мы предъявляем три обязательных требования:

1. Исходные коды должны соответствовать сопроводительной документации.
2. Должен быть пример запуска приведённого кода, например с помощью консоли. Формат входа должен быть вида «папка с входными исследованиями» → «отчёт в описанном формате».
3. По запросу жюри необходимо продемонстрировать запуск и исполнение кода.

! Плюсами при оценке решений будут:

- Собранный докер-контейнер.
- Google Colab или аналог, позволяющий напрямую проверить работоспособность вашего алгоритма.
- Комментарии в коде.

! Мы хотим обратить внимание, что в случае несоблюдения обязательных требований, участник не будет допущен в следующий этап конкурса.