

RE: DLCP2023. Proceedings
Subject:
RE: DLCP2023. Proceedings
From:
С уважением, Михаил Ледовских <misha_ledov@list.ru>
Date:
29.09.2023, 23:42
To:
Alexander Kryukov <kryukov@theory.sinp.msu.ru>

Андрей Павлович, добрый день!

Благодарим за рецензию.

Мы ознакомились с замечаниями и внесли правки в статью. Исправленная версия статьи в приложении к письму. Тем не менее некоторые замечания требуют скорее ответа, чем внесения исправлений в текст. Поэтому излагаем ответы в письме.

1) Замечание: Постановка задачи упомянута лишь опосредованно в Разделе

6, где упоминается, что выбираются модели MO для решения задачи

классификации дефектов кожи.

Ответ: Благодарим за комментарий. Действительно структура введения была недостаточно проработана. Добавлено описание задачи в Введение.

2) Замечание: В эпоху искусственных нейронных сетей, демонстрирующих решающее преимущество в большинстве задач обработки изображений перед классическими моделями MO, и при условии постановки задачи как классификации изображений, выбор способа решения в пользу экспертного дизайнера признаков и применения классических моделей MO выглядит необоснованным. Подход выглядит тем более неоправданным при условии, что начиная с 2017 года уже опубликован целый ряд статей (например, [1,2]), демонстрирующих весьма высокое качество в задачах (бинарной) классификации меланомы и в целом поражений кожи с использованием сверточных нейросетей. Тем более неясно, почему эти исследования, довольно широко цитируемые, не упомянуты в обзорной части во Введении. В связи с этим современное состояние проблемы, решаемой в исследовании, нельзя считать представленным даже в начальном объеме.

Ответ: Мы ознакомились с предложенными статьями. В одной из этих работ рассматривается подход основанный на нейронных сетях и результатами более слабыми, чем представленными в работах, на которые мы ссылаемся. Во второй представлено неоконченное исследование без содержательных результатов и декларированием дальнейшего исследования, которое нигде не было опубликовано. Возможно, это замечание возникло по причине того, что современное состояние области описывалось не во введении, а в главе Related Works. Для улучшения читаемости работы, обзор был перенесен во введение. Работа [Romero-Lopez] была добавлена в обзор, как одна из самых первых по теме.

3) Замечание: В случае, если выбор методов решения и подходов формирования признакового пространства был продиктован какими-то специальными соображениями, это следует отдельно подробно описать во Введении или в отдельном разделе после Введения, посвященном формулировке задачи и выбору подхода ее решения.

Выбор моделей MO, примененных в задаче, никак не обоснован. Особенно странно выглядит включение модели решающих деревьев (DT), которая известна склонностью к переобучению и чрезвычайно высокой чувствительностью обобщающей способности к выбору гиперпараметров. В то же время, в исследовании не использованы ансамблевые модели, демонстрирующие наиболее высокое качество на современных задачах с табличными данными, а именно градиентный бустинг над решающими деревьями или полносвязные искусственные нейронные сети. Реализации таких моделей доступны и в пакете scikit-learn, и отдельно (например, CatBoost). Выбор списка моделей может быть продиктован разными соображениями, и можно предположить, что в свете каких-то резонансов приведенный крайне неудачный набор тестируемых моделей может оказаться обоснованным. Эти соображения должны быть приведены.

Ответ: Благодарим за комментарий. В работе не был явно описан используемый в работе подход близкий по идее к бустингу и алгоритмам максимального использования. Этот подход использует предварительный статистический анализ, он был выбран по двум причинам: понижение размерности задачи и увеличение корреляции вектора признаков с распознаваемыми классами. Этот подход имеет преимущества как по сравнению с градиентным бустингом, имеющим слишком слабую сходимость, так и перед сверточными подходами, теряющими значительную часть полезной информации. Также повышение корреляции вектора признаков позволяет использовать рассматриваемые в работе простые модели классификации, а также снижает их склонность к переобучению. Это позволяет получить очень вычислительно легковесное решение. Добавлено описание в работу.

4) Замечание: Приведенная в статье расшифровка "K-means method" - модель кластеризации, не релевантная исследованию.

Ответ: Спасибо за замечание. Опечатка. Поправили.

5) Замечание: Выбор мер качества, использованных в исследовании, никак не обоснован. Более того, применяется либо мера Accuracy (доля верных ответов), известная ненадежностью в случае сильного дисбаланса классов (что наблюдается в исследовании), либо Recall (полнота, чувствительность), односторонняя мера, не характеризующая специфичность или точность. В существующих исследованиях такого же рода наряду с полнотой применяются другие меры, такие как специфичность, индекс Жаккара, F1-мера, etc. Подбор меры качества в задаче классификации дефектов кожи для неидеального классификатора при условии сильно несбалансированной выборки - сложный многогранный вопрос, в котором затрагиваются этические, социальные, статистические аспекты. При условии несогласованности выбранных мер качества с существующими высоко цитируемыми исследованиями обязательно следует обосновать выбор именно Accuracy и Recall. Альтернативно следует воспользоваться сложившейся практикой и опираться на меры, используемые в существующих исследованиях.

Ответ: В статье использованы меры Accuracy и Recall, как наиболее принятые в сложившейся практике. Во всех публикациях по задаче распознавания поражений кожи используются именно они (включая публикации указанные рецензентом). Было решено в работе приводить именно их, для более наглядного сравнения с существующими работами. При этом, при исследовании использовались и другие меры, например F1-мера. Результаты этого анализа доступны по ссылке указанной в статье. В работу добавлено упоминание о этом.

6) Замечание: Отдельно следует отметить, что в результате чистки наборов данных существенное количество экземпляров были исключены из рассмотрения. При этом исследование позиционируется как шаг в направлении к автоматизированной системе классификации дефектов кожи, которая должна обрабатывать все снимки, поступающие на анализ. Таким образом, вопрос ограничений применимости разработанных моделей никак не освещен. В тексте заключения должны быть упомянуты такие ограничения.

Ответ: При исследовании были исключены снимки, содержащие повторы, а также снимки содержащие лишь частичное изображение поражения или имеющие дополнительные подписи и разметку. Эти снимки были исключены не по причине невозможности обработки, а из-за несоответствия методике сбора данных. В Заключение добавлено обсуждение применимости подхода

7) Замечание: Описание процесса подбора признаков приведено чересчур подробно. В деталях описывается трудный многофазный путь авторов, который в конечном итоге приводит к одному единственному способу формирования признакового описания. В случае, если была поставлена цель продемонстрировать исследование релевантности признаков, это должно быть описано отдельно; должны быть приведена методика исследования, приведены результаты этого исследования и проведены соответствующие выводы. Однако ни в аннотации, ни в заголовке статьи, ни во Введении не обозначена цель исследования релевантности признаков. Считаю, что описание подробного пути, приводящего к правильным признакам, следует существенно сократить и переформулировать.

Ответ: Спасибо за замечание. Во Введение и аннотацию добавлено в качестве одной из целей исследования, описание исследования признаков. На наш взгляд такой анализ является одной из важнейших составляющих описываемого подхода, во многом определяющий довольно высокие результаты классификации, и потому не допускающий сокращения описания. Без такого метода понижения размерности результаты классификации на простых моделях классификации были бы крайне сомнительными. Тогда как полученные результаты не только сравнимы с подходами рассмотренными в работах других авторов, использующими нейронные сети глубокого обучения, но и могут быть скомбинированы с такими подходами, что планируется сделать в последующих работах.

8) Замечание: Саму задачу, решаемую в исследовании, следует сформулировать во Введении, а не в Разделе 6. Читателю с самого начала статьи должно быть ясно, какая решается задача. Это не до конца ясно даже из аннотации, в которой значится буквально следующее: "In this work, a program was developed for segmentation of affected and unaffected skin areas. The resulting masks were used to train machine learning models." Для решения какой именно задачи обучаются модели MO - неясно. Для чего разрабатывается программа сегментации чистых и поврежденных участков кожи - тоже неясно. Эти детали должны быть обоснованы и прояснены во Введении или отдельном разделе, посвященном формулировке задачи.

Ответ: Спасибо за замечания. Аннотацию поправили. Постановку задачи привели во Введении.

9) Замечание: Общая рекомендация по статье состоит в том, чтобы существенно

редуцировать статью до объема краткой (5-9 стр.) в силу отсутствующей или не представленной новизны. Альтернативно можно рекомендовать сместить фокус статьи на исследование экспертно разрабатываемых признаков или подчеркнуть новизну. Как дополнительная альтернатива, статья может быть опубликована в форме некоторого обучающего материала (т.н. tutorial), однако для этого следует существенно переформатировать ее под такой формат.

Ответ: Спасибо за замечание. В доработанной версии статьи было уделено большее внимание работе с признаками, которые являются неотъемлемой частью всего подхода. Соответствующие описания добавлены во введении и другие части работы. Было уделено большее внимание описанию новизны работы, заключающейся во впервые разработанной автоматизации получения признаков по изображению поражений, предложенном новом подходе основанном на комбинированном применении статистического анализа и применении моделей машинного обучения, а также получение результатов позволяющие усилить подходы рассматриваемые в работах других авторов. Сокращение статьи или перевод в другой формат требует полной переработки текста, что занимает больше чем 2 дня между получением рецензии и поставленным дедлайном.

С уважением
Михаил, Вячеслав

От: Alexander Kryukov
Отправлено: 26 сентября 2023 г. в 23:31
Кому: Михаил Ледовских
Тема: Fwd: DLCP2023. Proceedings

Добрый день!

На вашу работу была получена рецензия с рядом существенных замечаний. С учетом этих замечаний Вам предлагается сократить объем статьи до краткой (5-9 стр.).

Просьба дать ответ рецензенту и внести необходимые изменения в содержание статьи до 29 сентября.

С уважением, А.Крюков

===== Рецензия =====

В работе представлен подход, основанный на методах машинного обучения (МО), для классификации дерматоскопических фотографий, содержащих повреждения кожи, в отношении типа повреждений; преимущественный фокус исследования направлен на классификацию новообразований. В работе

представлен подход, основанный на классических моделях МО, который резонно потребовал реализации этапа дизайна признакового описания фотографий, который расписан достаточно подробно для воспроизведения исследования. Постановка задачи упомянута лишь опосредованно в Разделе 6, где упоминается, что выбираются модели МО для решения задачи классификации дефектов кожи. Дизайн исследования описан достаточно ясно; четко описаны меры качества, которыми оцениваются решения. Результаты представлены достаточно четко и ясно. Выводы подкреплены результатами.

К исследованию есть целый ряд существенных замечаний.

В эпоху искусственных нейронных сетей, демонстрирующих решающее преимущество в большинстве задач обработки изображений перед классическими моделями МО, и при условии постановки задачи как классификации изображений, выбор способа решения в пользу экспертного дизайна признаков и применения классических моделей МО выглядит необоснованным. Подход выглядит тем более неоправданным при условии, что начиная с 2017 года уже опубликован целый ряд статей (например, [1,2]), демонстрирующих весьма высокое качество в задачах (бинарной) классификации меланомы и в целом поражений кожи с использованием свёрточных нейросетей. Тем более неясно, почему эти исследования, довольно широко цитируемые, не упомянуты в обзорной части во Введении. В связи с этим современное состояние проблемы, решаемой в исследовании, нельзя считать представленным даже в начальном объеме. Скорее, следует считать представленный обзор современных методов для решения похожих конечных задач сильно устаревшим. В связи с этим нельзя считать, что в исследовании сформулирована новизна по сравнению с существующими исследованиями. В случае, если выбор методов решения и подходов формирования признакового пространства был продиктован какими-то специальными соображениями, это следует отдельно подробно описать во Введении или в отдельном разделе после Введения, посвященном формулировке задачи и выбору подхода ее решения.

Выбор моделей МО, примененных в задаче, никак не обоснован. Особенно странно выглядит включение модели решающих деревьев (DT), которая известна склонностью к переобучению и чрезвычайно высокой

чувствительностью обобщающей способности к выбору гиперпараметров. В то же время, в исследовании не использованы ансамблевые модели, демонстрирующие наиболее высокое качество на современных задачах с табличными данными, а именно градиентный бустинг над решающими деревьями или полносвязные искусственные нейронные сети. Реализации таких моделей доступны и в пакете `scikit-learn`, и отдельно (например, `CatBoost`). Выбор списка моделей может быть продиктован разными соображениями, и можно предположить, что в свете каких-то резонансов приведенный крайне неудачный набор тестируемых моделей может оказаться обоснованным. Эти соображения должны быть приведены.

Модель KNN - "K nearest neighborhood" (модель классификации или регрессии в зависимости от сценария применения). Приведенная в статье расшифровка "K-means method" - модель кластеризации, не релевантная исследованию. Выбор мер качества, использованных в исследовании, никак не обоснован. Более того, применяется либо мера Accuracy (доля верных ответов), известная ненадежностью в случае сильного дисбаланса классов (что наблюдается в исследовании), либо Recall (полнота, чувствительность), односторонняя мера, не характеризующая специфичность или точность. В существующих исследованиях такого же рода наряду с полнотой применяются другие меры, такие как специфичность, индекс Жаккара, F1-мера, etc. Подбор меры качества в задаче классификации дефектов кожи для неидеального классификатора при условии сильно несбалансированной выборки - сложный многогранный вопрос, в котором затрагиваются этические, социальные, статистические аспекты. При условии несогласованности выбранных мер качества с существующими высоко цитируемыми исследованиями обязательно следует обосновать выбор именно Accuracy и Recall. Альтернативно следует воспользоваться сложившейся практикой и опираться на меры, используемые в существующих исследованиях.

Отдельно следует отметить, что в результате чистки наборов данных существенное количество экземпляров были исключены из рассмотрения. При этом исследование позиционируется как шаг в направлении к автоматизированной системе классификации дефектов кожи, которая должна обрабатывать все снимки, поступающие на анализ. Таким образом, вопрос ограничений применимости разработанных моделей никак не освещен. В

тексте заключения должны быть упомянуты такие ограничения.

Далее приводятся замечания по форме изложения исследования.

Описание процесса подбора признаков приведено чересчур подробно. В деталях описывается трудный многофазный путь авторов, который в конечном итоге приводит к одному единственному способу формирования признакового описания. В случае, если была поставлена цель продемонстрировать исследование релевантности признаков, это должно быть описано отдельно; должны быть приведена методика исследования, приведены результаты этого исследования и проведены соответствующие выводы. Однако ни в аннотации, ни в заголовке статьи, ни во Введении не обозначена цель исследования релевантности признаков. Считаю, что описание подробного пути, приводящего к правильным признакам, следует существенно сократить и переформулировать. Саму задачу, решаемую в исследовании, следует сформулировать во Введении, а не в Разделе 6. Читателю с самого начала статьи должно быть ясно, какая решается задача. Это не до конца ясно даже из аннотации, в которой значится буквально следующее: "In this work, a program was developed for segmentation of affected and unaffected skin areas. The resulting masks were used to train machine learning models." Для решения какой именно задачи обучаются модели MO - неясно. Для чего разрабатывается программа сегментации чистых и поврежденных участков кожи - тоже неясно. Эти детали должны быть обоснованы и прояснены во Введении или отдельном разделе, посвященном формулировке задачи.

Общая рекомендация по статье состоит в том, чтобы существенно редуцировать статью до объема краткой (5-9 стр.) в силу отсутствующей или не представленной новизны. Альтернативно можно рекомендовать сместить фокус статьи на исследование экспертно разрабатываемых признаков или подчеркнуть новизну. Как дополнительная альтернатива, статья может быть опубликована в форме некоторого обучающего материала (т.н. tutorial), однако для этого следует существенно переформатировать ее под такой формат.

СПИСОК ЛИТЕРАТУРЫ

[1] Yu, Lequan, et al. "Automated melanoma recognition in dermoscopy images via very deep residual networks." IEEE transactions on medical imaging 36.4 (2016): 994-1004

[2] Lopez, Adria Romero, et al. "Skin lesion classification from dermoscopic images using deep learning techniques." 2017 13th IASTED international conference on biomedical engineering (BioMed). IEEE, 2017.

--

A.Kryukov, PhD

Head of laboratory, SINP MSU

Phone: +7 495 939-3173

Attachments:

d\cp2023-Ledovskikh-3.rar 4,2 MB