



Создание карточки модели

Карточка модели — это файл, который, возможно, так же важен, как файлы модели и токенизатора в репозитории моделей. Это центральное описание модели, обеспечивающее возможность повторного использования другими членами сообщества и воспроизводимость результатов, а также предоставляющее платформу для других участников.

Документирование процесса обучения и оценки помогает другим понять, чего ожидать от модели, а предоставление достаточной информации об использованных данных, а также о проведенной предварительной и постобработке. По карточке модели ясны ограничения, предубеждения и контексты, в которых модель может быть полезна, а в каких случаях окажется бесполезной.

Поэтому создание карточки модели, которая четко определяет вашу модель, является очень важным шагом. Здесь мы даем несколько советов, которые помогут вам в этом. Карточка модели создается с помощью файла *README.md*, который вы видели ранее, который представляет собой файл Markdown.

Концепция «карточки модели» возникла в результате исследовательского направления Google, впервые представленного в статье [“Model Cards for Model Reporting”] (<https://arxiv.org/abs/1810.03993>) Маргарет Митчелл и др. Большая часть информации, содержащейся здесь, основана на этом документе, и мы рекомендуем вам ознакомиться с ним, чтобы понять, почему карточки с моделями так важны в мире, который ценит воспроизводимость, возможность повторного использования и честность.

Карточка модели обычно начинается с очень краткого общего обзора того, для чего предназначена модель, за которым следуют дополнительные сведения в следующих разделах:

- Описание модели
- Предполагаемое использование и ограничения
- Как использовать
- Ограничения и предубеждения
- Тренировочные данные
- Процедура обучения
- Результаты оценки

Давайте посмотрим, что должен содержать каждый из этих разделов.

Описание модели

Описание содержит основные сведения о модели. Оно включает в себя архитектуру, версию, если модель была представлена в статье - автора, ссылку на оригинальную реализацию (если доступна), автора и общую информацию о модели. Любые авторские права должны быть указаны здесь. В этом разделе также можно упомянуть общую информацию о процедурах обучения, параметрах и важных отказах от ответственности.

Предполагаемое использование и ограничения

Здесь вы описываете варианты использования, для которых предназначена модель, включая языки, области и домены, в которых она может применяться. В этом разделе карты модели также можно документировать те области, которые являются неподходящими для модели.

Как использовать

Этот раздел должен включать несколько примеров того, как использовать модель. Это может быть продемонстрировано с использованием функции `pipeline()`, использованием классов модели и токенизатора, а также любым другим способом, удобным на ваш взгляд.

Обучающие данные

В этой части должно быть указано, на каком наборе данных обучалась модель. Также приветствуется краткое описание набора(ов) данных.

Процедура обучения

В этом разделе вы должны описать все важные аспекты обучения, которые полезны с точки зрения воспроизводимости. Раздел включает в себя любую предварительную и постобработку данных, а также такие детали, как количество эпох, на которых была обучена модель, размер батча, скорость обучения и т. д.

Variable and metrics

Здесь вы должны описать метрики, которые вы используете для оценки, и прочие величины, которые вы измеряете. Напишите, какие метрики использовались, в каком датасете и какое разделение разбиение датасета позволяет легко сравнивать производительность вашей модели с другими моделями.

Результаты валидации

Наконец, укажите, насколько хорошо модель работает с набором данных для оценки. Если в модели используется порог принятия решения (threshold)– укажите его, либо предоставьте подробные сведения об оценке при различных порогах для предполагаемого использования.

Пример

Ознакомьтесь с несколькими примерами хорошо сделанных карточек моделей:

- [bert-base-cased](#)
- [gpt2](#)
- [distilbert](#)

Больше примеров от других организаций и компаний доступны: [здесь](#).

Примечание

Карточки моделей не являются обязательным требованием при публикации моделей, и вам не нужно включать все разделы, описанные выше, при их создании. Однако подробное документирование модели может принести только пользу будущим пользователям, поэтому мы рекомендуем вам заполнить как можно больше разделов в меру своих знаний и способностей.

Метаданные карточки модели

Если вы немного изучили Hugging Face Hub, вы должны были заметить, что некоторые модели относятся к определенным категориям: вы можете фильтровать их по задачам, языкам, библиотекам и т. д. Категории, к которым принадлежит модель, идентифицируются в соответствии с метаданными, которые вы добавляете в заголовок карточки модели.

Например, если вы посмотрите на [карточку модели camembert-base](#), вы должны увидеть следующие строки в заголовке карточки модели:

```
---
language: fr
license: mit
datasets:
- oscar
---
```

Эти метаданные анализируются Hugging Face Hub, который затем идентифицирует эту модель как французскую модель с лицензией MIT, обученную на наборе данных Oscar.

[Полная спецификация карточки модели](#) позволяет указать языки, лицензии, теги, датасеты, метрики, а также результаты валидации модели.