

Khắc phục sự cố

Hướng dẫn này nhằm mục đích cung cấp cho bạn các công cụ và kiến thức cần thiết để điều hướng một số vấn đề. Nếu những gợi ý được liệt kê trong hướng dẫn này không đề cập đến tình huống như vậy của bạn, vui lòng tham khảo phần Yêu cầu trợ giúp học ở đâu tìm trợ giúp về vấn đề cụ thể của bạn.

Sự cố khi tải lên tập dữ liệu bằng `push_to_hub`

Vấn đề xác thực

Nếu bạn gặp sự cố xác thực khi chia sẻ tập dữ liệu trên Hub bằng cách sử dụng `Dataset.push_to_hub()` and a Hugging Face mã thông báo truy cập:

- Đảm bảo rằng mã thông báo Ôm Mặt bạn đang sử dụng để xác thực chính mình là mã thông báo với sự cho phép viết.
- Trên OSX, nó có thể giúp xóa tất cả mật khẩu ôm mặt.co trên móc khóa của bạn truy cập, cũng như cấu hình lại git config `--global credential.helper osxkeychain` , trước khi sử dụng đăng nhập ôm mặt-cli.

Ngoài ra, bạn có thể sử dụng khóa SSH để tự xác thực - đọc thêm trong Hub tài liệu.

Mất kết nối khi tải lên tập dữ liệu lớn

Khi tải các tập dữ liệu lớn lên Hub, nếu số lượng phân đoạn tập dữ liệu lớn, nó có thể tạo ra quá nhiều cam kết cho Hub trong một thời gian ngắn. Điều này sẽ dẫn đến lỗi kết nối.

Lỗi kết nối cũng có thể do lỗi HTTP 500 được nhóm AWS S3 trả về mà Hub sử dụng nội bộ.

In either situation, you can re-run `Dataset.push_to_hub()` to proceed with the dataset upload.

Hub sẽ kiểm tra SHA

của các phân đoạn đã được tải lên để tránh tải chúng lên lại.

Chúng tôi đang nỗ lực làm cho quá trình tải lên trở nên mạnh mẽ hơn trước các lỗi nhất thời, vì vậy việc cập nhật phiên bản thư viện mới nhất là luôn luôn là một ý tưởng tốt.

Quá nhiều yêu cầu

Uploading large datasets via `push_to_hub()` can result in an error:

`HfHubHTTPError: 429 Lỗi máy khách: Quá nhiều yêu cầu đối với url: ...`

Bạn đã vượt quá hạn ngạch hành động hàng giờ của chúng tôi: cam kết. Chúng tôi mời bạn thử lại sau.

If you encounter this issue, you need to upgrade the `datasets` library to the latest version (or at least `2.15.0`).

Các vấn đề khi tạo tập dữ liệu từ dữ liệu tùy chỉnh

Đang tải hình ảnh và âm thanh từ một thư mục

Khi tạo tập dữ liệu từ một thư mục, một trong những vấn đề phổ biến nhất là cấu trúc tệp không tuân theo định dạng mong muốn hoặc có vấn đề với tệp siêu dữ liệu.

Tìm hiểu thêm về cấu trúc thư mục bắt buộc trong các trang tài liệu tương ứng:

- Thư mục âm thanh
- Thư mục Hình ảnh

Vấn đề ngâm chua

Sự cố tấy khi sử dụng `Dataset.from_generator`

When creating a dataset, `IterableDataset.from_generator()` and `Dataset.from_generator()` mong đợi một chức năng tạo "có thể chọn được".

Điều này là bắt buộc để đảm bảo cách sử dụng dựa trên hàm để có thể lưu tập dữ liệu vào bộ đệm trên đĩa.

Mặc dù các chức năng của trình tạo nói chung là "có thể chọn được", hãy lưu ý rằng các đối tượng của trình tạo bạn đang sử dụng một đối tượng máy phát điện,

bạn sẽ gặp `TypeError` như thế này:

`TypeError: không thể chọn đối tượng 'máy phát điện'`

Lỗi này cũng có thể xảy ra khi sử dụng hàm tạo sử dụng đối tượng chung không phải là

"có thể ngâm được", chẳng hạn như

Kết nối DB chẳng hạn. Nếu đúng như vậy, bạn có thể khởi tạo đối tượng đó trực tiếp bên trong chức năng của máy phát điện tránh lỗi này.

Sự cố xảy ra với `Dataset.map`

Pickling errors can also happen in the multiprocessing `Dataset.map()` - objects are pickled to be được truyền cho các tiến trình con.

Nếu các đối tượng được sử dụng trong quá trình chuyển đổi không thể chọn được thì không thể lưu kết quả `map`, dẫn đến lỗi được nêu ra.

Dưới đây là một số cách để giải quyết vấn đề này:

- A universal solution to pickle issues is to make sure the objects (or generator classes) are có thể chọn thủ công bằng cách triển khai `__getstate__` / `__setstate__` / `__reduce__`.
- Bạn cũng có thể cung cấp hàm băm duy nhất của riêng mình trên bản đồ với đối số `new_fingerprint`.
- You can also disable caching by calling `datasets.disable_caching()`, however, this is không mong muốn - đọc thêm về tầm quan trọng của bộ đệm

Yêu cầu giúp đỡ

Nếu lời khuyên khắc phục sự cố ở trên không giúp bạn giải quyết được sự cố của mình, hãy liên hệ để được t cộng đồng và đội ngũ.

Diễn đàn

Yêu cầu trợ giúp trên diễn đàn Ôm mặt - đăng câu hỏi của bạn trong danh mục Bộ dữ liệu

Đảm bảo viết một bài đăng mô tả có ngữ cảnh phù hợp về thiết lập của bạn và khả năng tái tạo mã để tối đa hóa khả năng vấn đề của bạn được giải quyết!

Bất hòa

Đăng câu hỏi trên Discord và để nhóm cũng như cộng đồng giúp đỡ bạn.

Thảo luận cộng đồng trên Hub

Nếu bạn đang gặp vấn đề khi tạo tập dữ liệu tùy chỉnh trên Hub, bạn có thể hỏi Khuôn mặt ôm nhóm để được trợ giúp bằng cách mở một cuộc thảo luận trong tab Cộng đồng của tập dữ liệu của bạn với t

Dataset review request for <Dataset name>

Sự miêu tả

<brief description of the dataset>

Tập cần xem lại

- tập tin1

- tập tin2

None

cc @lhoestq @albertvillanova

Sự cố GitHub

Cuối cùng, nếu bạn nghi ngờ đã tìm thấy lỗi liên quan đến chính thư viện, hãy tạo Sự cố trên Bộ dữ liệu

Kho lưu trữ GitHub. Bao gồm bối cảnh liên quan đến lỗi: đoạn mã để sao chép, chi tiết về môi trường và dữ liệu của bạn, v.v. để giúp chúng tôi tìm ra điều gì sai và cách chúng tôi có thể sửa nó