

Transcription de prises de parole

Réalisé par :

- Othmane PREURE
- NECHAB Idir
- Fawzi KETREB
- Yacine LARIBI

Encadrant :

- M. Boubekur Merabet

Table des matières

- Introduction
- Évaluation des performance des différents modèles
- Prompt engineering
- Post-processing
- Conclusion

Introduction

- La SNCF communique avec ses usagers de différentes manières : à travers son application, des annonces vocales diffusées par haut-parleurs, et des affichages sur écrans.
Par ailleurs, la SNCF est tenue de communiquer ses informations dans plusieurs langues : espagnol, anglais et français.

Problematique

- Les outils de transcription disponibles ont leurs limites, notamment lorsqu'ils sont confrontés à des ambiguïtés, à des termes techniques spécifiques au domaine ferroviaire ou à des noms de lieux tels que les gares.

Pré-traitement des données

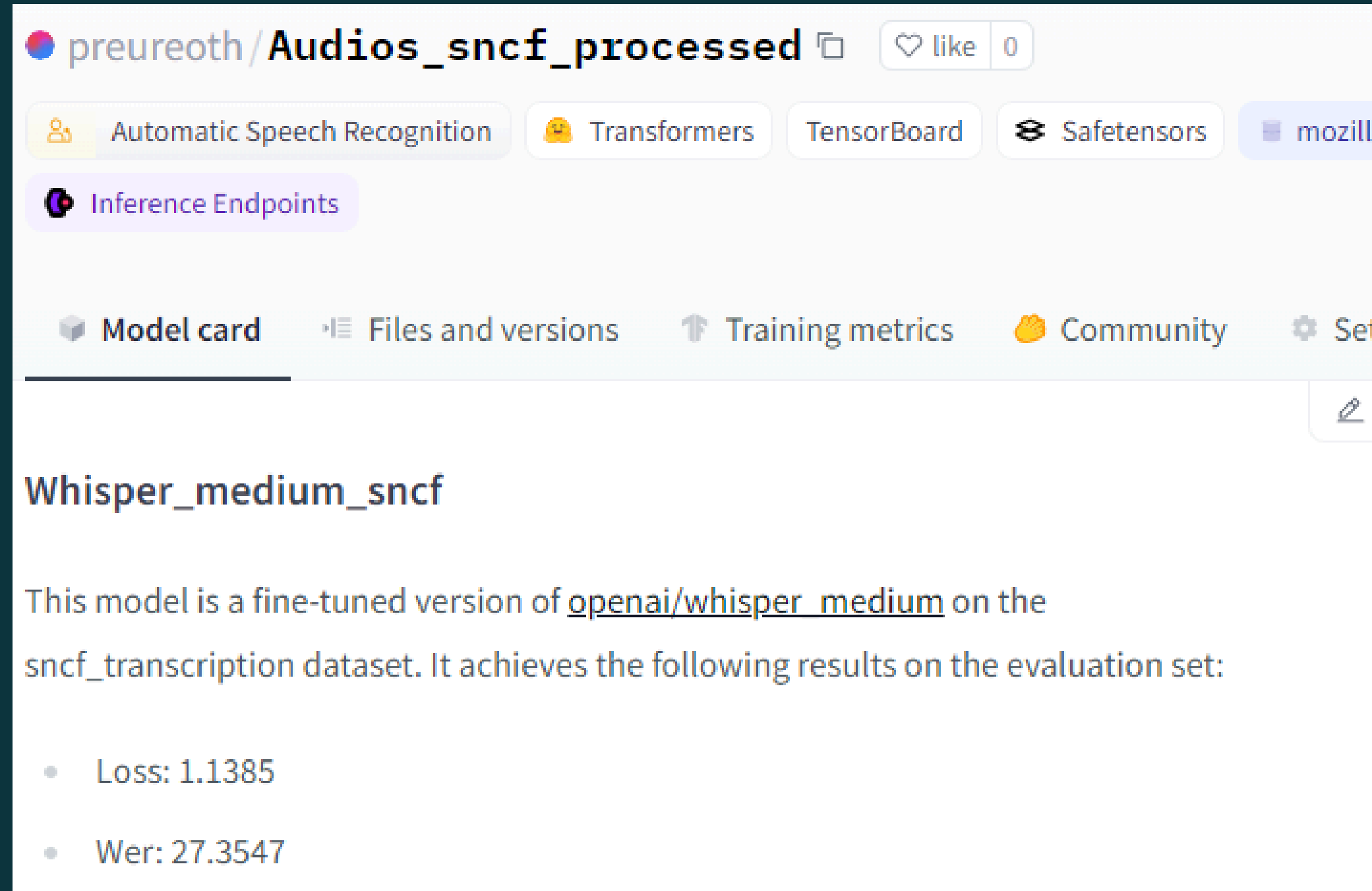
- Environ 200 fichiers audio ont été fournis, totalisant près de deux heures d'enregistrement.
- Des corrections ont été apportées aux transcriptions pour améliorer leur précision.
- Les fichiers audio d'une durée supérieure à 30 secondes ont été divisés en segments plus petits pour être traités par le modèle.
- Les fichiers audio ont été convertis en spectrogrammes, une représentation visuelle du son, après avoir ajusté leur fréquence d'échantillonnage à 16000 Hz.
- Une tentative d'augmentation des données a été effectuée, mais elle a été abandonnée en raison de l'absence de timestamp pour chaque intervalle de temps dans la transcription.

Évaluation des performance des différents modèles

- Nous avons évalué un modèle spécialisé pour la langue française, mais les résultats étaient décevants, avec un taux d'erreur de 39.22 %.
- Nous avons entraîné le modèle Whisper de taille moyenne disponible sur Hugging Face.
- Avant l'entraînement du modèle, le taux d'erreur était de 28.38 %.
- L'application d'un filtre pour améliorer la qualité des fichiers audio n'a pas donné les résultats escomptés, le taux d'erreur augmentant à 37 %.

Évaluation des performance des différents modèles

- Après l'entraînement sur nos données, nous avons atteint un taux d'erreur de 27.35 %, ce qui reste insuffisant.



The screenshot shows a Hugging Face model card for the repository 'preureoth/Audios_sncf_processed'. The card is titled 'Whisper_medium_sncf' and describes it as a fine-tuned version of 'openai/whisper_medium' on the 'sncf_transcription' dataset. It lists the following evaluation results:

- Loss: 1.1385
- Wer: 27.3547

The interface includes a top navigation bar with links to 'Automatic Speech Recognition', 'Transformers', 'TensorBoard', 'Safetensors', and 'mozilla'. Below the title, there are tabs for 'Model card', 'Files and versions', 'Training metrics', 'Community', and 'Settings'. The 'Model card' tab is currently selected.

Prompt engineering

Nous avons exploré une autre approche en utilisant désormais la bibliothèque OpenAI, qui propose un paramètre "prompt" permettant de fournir un contexte au modèle. Cela accorde plus d'importance à certains mots et facilite la transcription des termes tels que les noms de sites de la SNCF qui ne sont pas toujours reconnus correctement.

Le modèle disponible sur OpenAI a une précision de 22% sur un ensemble de 200 fichiers audio.

- Après l'ajout des prompts pour préciser le contexte, nous obtenons un taux d'erreur de 16%.

Conclusion

- Exploré les défis de la transcription automatique des communications multilingues de la SNCF
- Description du dataset utilisé (200 fichiers audio, deux heures d'enregistrement) et pré-traitement
- Entraînement d'un modèle Whisper de OpenAI et réduction du taux d'erreurs à 16%
- Réalisation d'une interface graphique pour enregistrer un audio en français puis le transcrire et le traduire vers l'anglais et l'espagnol
- Nos perspectives pour améliorer le modèle de transcription

Merci pour votre attention
Vos questions sont les bienvenues

Demo

Transcription et Traduction



0:05

En Français

Traduction en anglais

Traduction en espagnol

© 2024 SNCF. Tous droits réservés.



Rechercher



FRA



09:06

26/04/2024