

Test du Stage ASR : Optimisation fine basée sur des adaptateurs (Adapter-Based Fine-Tuning) pour les langues à faibles ressources

Table des matières

1. Aperçu du défi
 2. Objectif principal
 3. Commencement
 4. Votre tâche
 5. Règles et contraintes
 6. Liste de vérification de soumission
 7. Critères d'évaluation
 8. Documents pertinents
-

Aperçu du défi

Bienvenue ! Ce test fait partie du processus de sélection pour le stage en reconnaissance vocale automatique (ASR). Nous recherchons des personnes talentueuses et passionnées par la mise à disposition de la technologie vocale pour les langues à faibles ressources afin de la rendre accessible à tous.

La technique présentée dans ce test s'inspire de recherches de pointe. Lorsqu'un modèle pré-entraîné est parfaitement ajusté sur un jeu de données spécifique à un domaine, il existe un risque d'« oubli catastrophique », c'est-à-dire que le modèle oublie les informations précédemment apprises lors de l'introduction de nouvelles données. Les adaptateurs résolvent ce problème en insérant de petits modules efficaces dans le modèle de base figé, et en y dirigeant le nouvel apprentissage. Cela permet au modèle de maîtriser de nouveaux domaines tout en préservant les connaissances préexistantes. Votre objectif est de démontrer l'efficacité de cette méthode.

Objectif principal

L'objectif est de réduire le taux d'erreur sur les mots (WER) sur l'ensemble de données du domaine fourni. Pour ce faire, vous entraînerez un module d'adaptation tout en conservant les poids du modèle de reconnaissance automatique de la parole (ASR) pré-entraîné. L'évaluation portera sur la précision de l'adaptateur et son amélioration par rapport au modèle de base.

Commencement

Choisissez un modèle ASR et affinez-le à l'aide d'adaptateurs sur l'ensemble de données ASR_Fellowship_Challenge_Dataset.

Jeu de données :

```
depuis hugingface_hub import snapshot_download
```

```
snapshot_download(repo_id="DigitalUmuganda/  
ASR_Fellowship_Challenge_Dataset",repo_type='dataset',local_dir='<destination_dir>')
```

Exemple d'instance de données :

```
{'creator_id': 'bcSXMbErjM6pJyAwwLs7NAxA9v2',  
  
'raw_text': 'Ingagi ihagaze yonyine. Ingagi ni nziza cyane, kuko zikurura ba  
mukerarugendo bakazana amadovize mu Gihugu cyacu.',  
  
'duration': 15.06,  
  
'LUFS': -25.5,  
  
'text': 'ingagi ihagaze yonyine ingagi ni nziza cyane kuko zikurura ba mukerarugendo  
bakazana amadovize mu gihugu cyacu',  
  
'audio_filepath': 'audio_1751479904-bcSXMbErjM6pJyAwwLs7NAxA9v2.webm',  
  
'image_filepath': 'restyf.jpg',  
  
'age_group': '50+',  
  
'gender': 'Male',  
  
'location': 'Musanze',  
  
'shard_id': 0,  
  
'image_shard_id': 0}
```

Votre tâche

Améliorez les performances du système de base sur l'ensemble de données Afrivoice_Kinyarwanda health en utilisant des adaptateurs comme stratégie de réglage fin.

Règles et contraintes

1. Modèle de base : Les poids du modèle ASR pré-entraîné doivent rester inchangés(frozen). **Toute soumission les modifiant sera disqualifiée.**
2. Aucune donnée externe : utilisez uniquement l'ensemble d'entraînement fourni.
3. Langages et frameworks : implémentez votre solution en Python en utilisant PyTorch (ou un framework construit sur celui-ci).

Liste de vérification de soumission

Créez un dépôt GitHub contenant :

1. **base_transcriptions.txt** – transcriptions de l'ensemble de données de test utilisant votre modèle de base.
2. **finetuned_transcriptions.txt** – transcriptions de l'ensemble de données de test utilisant votre modèle affiné.
3. Un répertoire contenant tout le code source (commenté et reproductible).
4. Poids du modèle de base.
5. Poids adaptateurs entraînés.
6. **rapport.pdf** – inclure :
 - Nom complet et coordonnées.
 - WER final pour les modèles de base et les modèles affinés.
 - Nombre de paramètres entraînables.
 - Description de l'architecture de votre adaptateur et de votre stratégie de formation.
 - Instructions étape par étape pour reproduire votre configuration et votre évaluation.

Soumettez votre dépôt via le formulaire Google fourni :

<https://docs.google.com/forms/d/e/1FAIpQLSdS9jc6PFPqBYFU1CFhE2DIjaFZztNmN1FN3AOTVHBCVbgSg/viewform?usp=dialog>

Critères d'évaluation

| Critère | Poids | Description |
|---------|-------|-------------|
|---------|-------|-------------|

| | | |
|----------------------------|-----|--|
| WER du jeu de test | 40% | Différence de WER entre les transcriptions de base et les transcriptions affinées. Plus la valeur est élevée, mieux c'est. |
| Innovation et méthodologie | 30% | Clarté de votre approche dans le rapport. |
| Reproductibilité | 30% | Organisation du code, documentation et facilité d'exécution. |

Ce défi est l'occasion de démontrer vos compétences sur un problème concret de reconnaissance automatique de la parole.

Bonne chance !

Documents pertinents

- B. Thomas, S. Kessler, and S. Karout, “Efficient Adapter Transfer of Self-Supervised Speech Models for Automatic Speech Recognition,” arXiv:2202.03218 (2022).
- W. Hou et al., “Exploiting Adapters for Cross-lingual Low-resource Speech Recognition,” arXiv:2105.11905 (2021).
- N. Houlsby et al., “Parameter-Efficient Transfer Learning for NLP,” arXiv:1902.00751 (2019).