Cross Lingual Speaker Adaptation for TTS Applications

Software Project-2021/2022

Rasul, Anna, Claesia, Sharmila

Outline

- Finished Tasks
- Unresolved Issues
- Training Difficulties
- Training Status
- Plans
- More Architectures
- Timeline

Finished Tasks

- Checkpoints
  - Allows us to train models for longer (limit on training time from Grid 5000)
- Start training

Resolved Issues

- Memory restrictions on Grid 5000
  - Applied for additional storage
- VCTK EN dataset had too many speakers
  - Chose top 5 from them

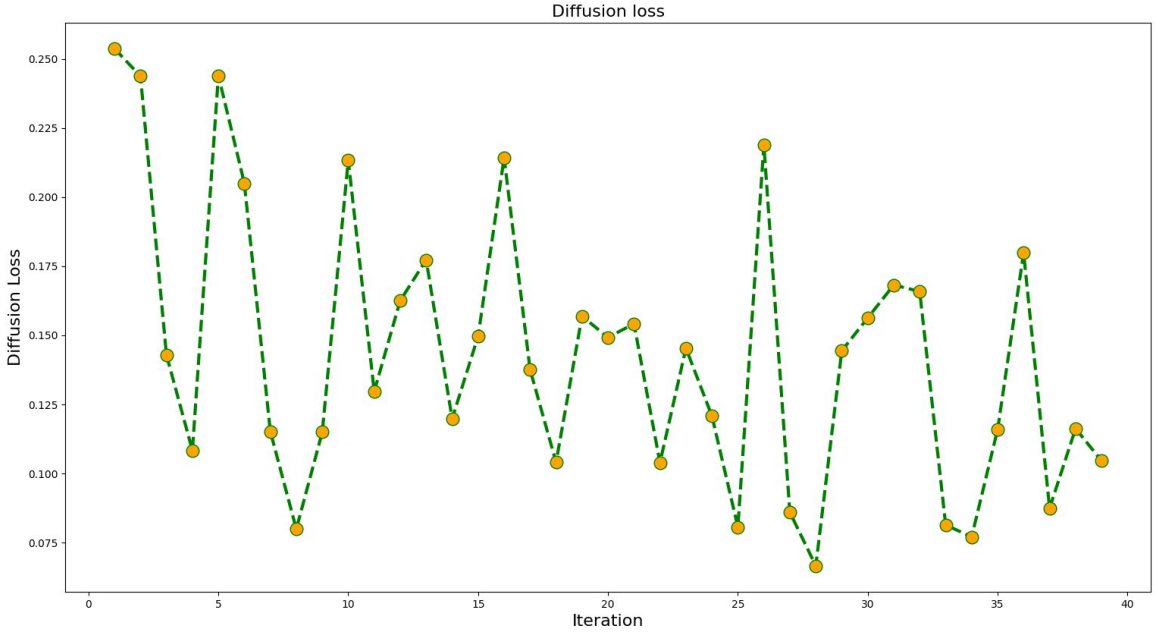
Unresolved Issues

- Multi-gpu support
  - Package conflicts
- Docker
  - Operating system incompatibilities
  - Installing dependencies non-interactively
  - CPU inference in the absence of GPU

Training Difficulties

- High training time, while the walltime that the grid provide might not be sufficient
- Adding and validating French dataset for speaker representation learning.
- Losses decreasing rather slowly in the first 40 epochs

Training Status



Diffusion loss seems to converge very slowly after the beginning epochs, such long training is essential to get a good model. [Referred from GradTTS paper]

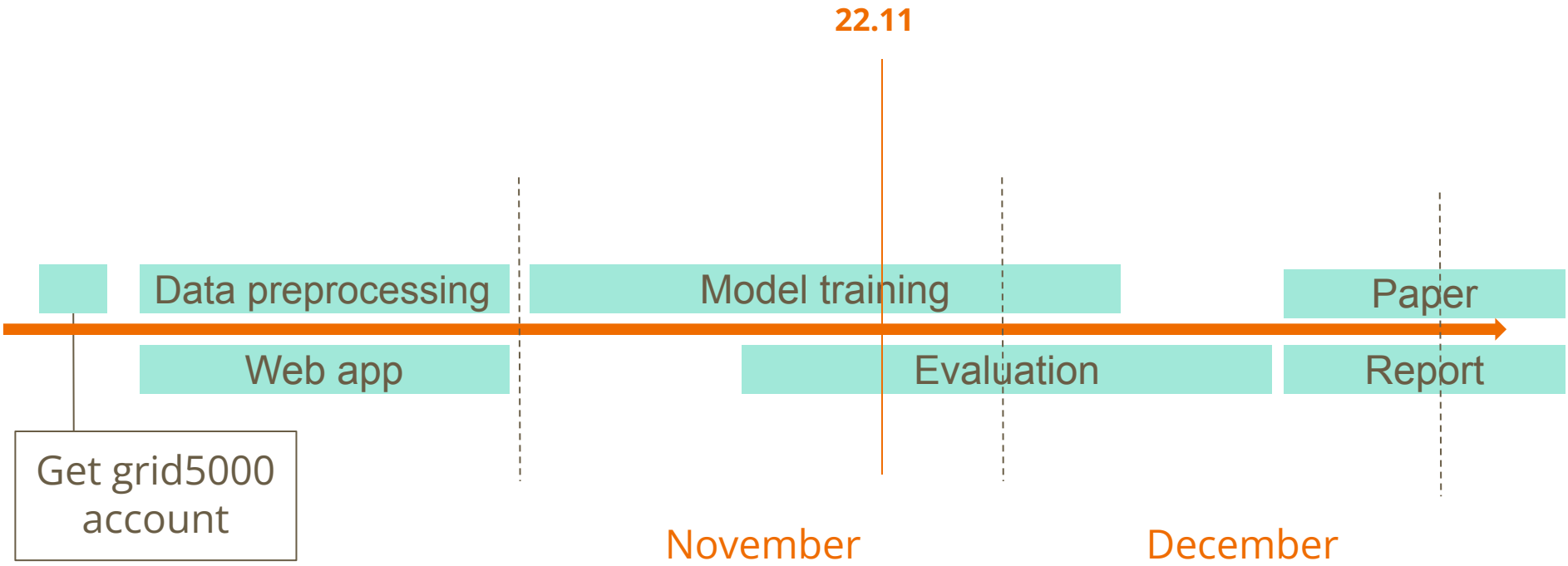
Plans

- Evaluation
- Add other FR datasets
  - Tundra Multilingual Dataset
  - Synpaflex French Speech Dataset
- Try different architectures

More Architectures

	Language embedding	Speaker embedding
Model 1	ID	ID
Model 2	ID	Embedding Network
Model 3	Embedding Network	ID
Model 4	Embedding Network	Embedding Network

Timeline


22.11
Data preprocessing
Model training
Paper
Web app
Evaluation
Report
Get grid5000 account
November
December

Thank you! Any questions?