



ການຮຽນຮູ້ ແລະ ການຈື່ຈຳ ຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນ
ຂໍ້ຄວາມສຽງ ດ້ວຍເຕັກນິກຂອງປັນຍາປະດິດຂັ້ນສູງ

Lao Text To Speech Recognition Using Deep Learning Technology

ນຳສະເໜີໂດຍ: 205N0014.19 ທ້າວ ຄຳປະເສີດ ໄຊຍະວົງ ຫ້ອງ: 4CS2

205N0045.19 ທ້າວ ໃຫຍ່ ແສງວິໄຊ ຫ້ອງ: 4CS2

ອາຈານຜູ້ນຳພາ: ອຈ.ປອ ລັດສະໝີ ຈິດຕະວົງ

ອາຈານຜູ້ຊ່ວຍນຳພາ: ອຈ.ປທ ສົມມິດ ທຸມມາລີ

ເນື້ອໃນລວມການນຳສະເໜີ

01 ຄວາມສຳຄັນຂອງບັນຫາ

02 ຈຸດປະສົງຂອງການຄົ້ນຄວ້າ

03 ຂອບເຂດຂອງການຄົ້ນຄວ້າ

04 ປະໂຫຍດຄາດວ່າຈະໄດ້ຮັບ

05 ວິທີດຳເນີນການຄົ້ນຄວ້າ

06 ຜົນຂອງການປະມວນຜົນ
ແບບຈຳລອງ

07 ສະຫຼຸບ

08 Demo

ໃນປັດຈຸບັນໂລກຂອງພວກເຮົາໄດ້ມີການພັດທະນາຄວາມສາມາດຂອງອຸປະກອນຕ່າງໆ ດ້ານຄອມພິວເຕີ ໃຫ້ມີການຮຽນຮູ້ຕາມແບບຢ່າງຂອງມະນຸດ. ໃນດ້ານການຈີຈໍາສຽງເວົ້າ (Speech Recognition) ກໍມີຄວາມກ້າວໜ້າໄປຫຼາຍ ໄດ້ມີການນໍາໄປປະຍຸກໃຊ້ໃນລະບົບສື່ສານໂທລະຄົມມະນາຄົມ ເຊັ່ນ: ການສັ່ງໂທອອກຂອງໂທລະສັບມືຖືໂດຍໃຊ້ສຽງ. ຈຸດປະສົງຫຼັກຂອງການຈີຈໍາສຽງເວົ້າຄືການເພີ່ມຄວາມສາມາດໃຫ້ອຸປະກອນຕ່າງໆ ສາມາດຮັບຮູ້ ແລະ ມີການໂຕ້ຕອບກັບມະນຸດໄດ້ຫຼາຍຂຶ້ນ ເຊິ່ງການໃຊ້ສຽງເວົ້າໃນການສື່ສານຄວບຄຸມສັ່ງການຖືວ່າເປັນວິທີທີ່ມະນຸດໃຊ້ຢ່າງເປັນທໍາມະຊາດທີ່ສຸດ.



Virtual



Phone



Voice Control

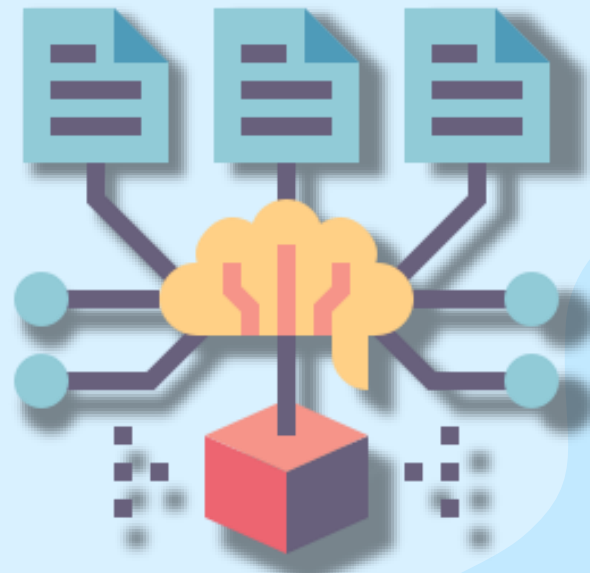


Computer

01

ຄວາມສໍາຄັນຂອງບັນຫາ (ຕໍ່)

ເນື່ອງຈາກວ່າ, ພາສາລາວຍັງບໍ່ມີນັກຄົ້ນຄວ້າທ່ານ ໃດທີ່ສ້າງລະບົບການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນ ຂໍ້ຄວາມສຽງ (Lao Text-To-Speech) ມາເຜີຍແຜ່ ແລະ ໃຫ້ນຳໃຊ້ກັນໃນດ້ານອຸປະກອນຄອມພິວເຕີຕ່າງໆ. ສະນັ້ນ, ພວກຂ້າພະເຈົ້າ ຈຶ່ງມີຄວາມສົນໃຈຢາກສຶກສາ, ປັບປຸງ ແລະ ນຳໃຊ້ແບບຈຳລອງການແປຂໍ້ຄວາມຂຽນໄປເປັນ ຂໍ້ຄວາມສຽງ (Text-To-Speech) ຂອງພາສາລາວໃຫ້ມີ ປະສິດທິພາບ, ສາມາດຮຽນຮູ້ ແລະ ຈຶ່ງຈຳ ໄດ້ຖືກຕ້ອງຊັດ ເຈນຍິ່ງຂຶ້ນ.



Model

02

ຈຸດປະສົງຂອງການຄົ້ນຄວ້າ

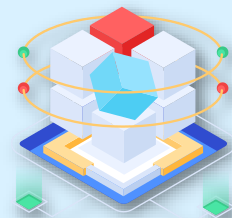
➤ ເພື່ອສຶກສາຂະບວນການເຮັດວຽກຂອງ ລະບົບການຈີຈຳການແປຂໍ້ຄວາມຂຽນໄປເປັນຂໍ້ຄວາມສຽງ (Text-To-Speech) ໃນການຮຽນຮູ້ ແລະ ຈີຈຳ ພາສາລາວ.

➤ ເພື່ອສຶກສາການປະມວນຜົນ ແລະ ວິເຄາະພາສາລາວມີທັງຂໍ້ຄວາມຂຽນ ແລະ ຂໍ້ຄວາມສຽງ ໂດຍການພັດທະນາມາຈາກເຄືອຂ່າຍປະສາດທຽມ DNN ຂອງ Deep Learning ໂດຍນຳໃຊ້ສະຖາປັດຕະຍະກຳ RNN ດ້ວຍເຕັກນິກ YourTTS.

➤ ເພື່ອປະມວນຜົນການຮຽນຮູ້ ແລະ ການຈີຈຳຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ (Lao Text-To-Speech) ດ້ວຍແບບຈຳລອງ YourTTS ທີ່ປັບປຸງແລ້ວ.



Text-To-Speech



YourTTS Model

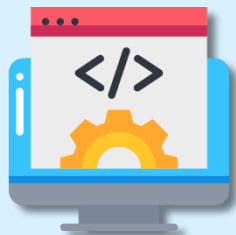


Processing

Speech-To-Text



- ເພື່ອສະແດງຜົນການຮຽນຮູ້ ແລະ ການຈື່ຈຳຂໍ້ຄວາມສຽງພາສາລາວ ໄປເປັນຂໍ້ຄວາມຂຽນ (Lao Speech-To-Text) ໂດຍນຳໃຊ້ແບບຈຳລອງ Google Speech To Text API.



Create Website

- ເພື່ອສ້າງເວັບໄຊການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ ແລະ ການແປຂໍ້ຄວາມສຽງພາສາລາວໄປເປັນຂໍ້ຄວາມຂຽນ ທີ່ມີປະສິດທິພາບ.

03

ຂອບເຂດຂອງການຄົ້ນຄວ້າ

- ກະກຽມຂໍ້ມູນສຽງພາສາລາວສໍານຽງພາກກາງ ເພື່ອນໍາໄປໃຊ້ໃນການຮຽນຮູ້ ແລະ ຈຶ່ງຈໍາ ຂໍ້ຄວາມຂຽນພາສາລາວ ເປັນຈໍານວນ 600 ຄໍາ.



Dataset



YourTTS Model

- ສ້າງຂັ້ນຕອນວິທີແບບ DNN ໂດຍນໍາໃຊ້ວິທີການຮຽນຮູ້ແບບການແປຂໍ້ຄວາມຂຽນໄປເປັນຂໍ້ຄວາມສຽງ (Text-To-Speech) ໂດຍນໍາໃຊ້ສະຖາປັດຕະຍະກຳ RNN ດ້ວຍເຕັກນິກ YourTTS ເພື່ອສ້າງແບບຈຳລອງການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ (Lao Text-To-Speech).

- ປະເມີນປະສິດທິພາບ ແລະ ການທົດສອບແບບຈຳລອງການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ (Lao Text-To-Speech).



Processing



Google Speech-To-Text API

- ສຶກສາ ແລະ ນຳໃຊ້ແບບຈຳລອງຂໍ້ມູນທີ່ຖືກພັດທະນາຂຶ້ນໂດຍ Google Speech To Text API ເພື່ອແປຄຳເວົ້າພາສາລາວໄປເປັນຂໍ້ຄວາມຂຽນ (Lao Speech To Text).



Create Website

- ສ້າງເວັບໄຊການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ ແລະ ການແປຂໍ້ຄວາມສຽງໄປເປັນຂໍ້ຄວາມຂຽນ.

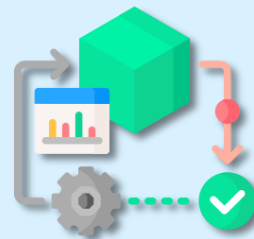
04

ປະໂຫຍດຄາດວ່າຈະໄດ້ຮັບ

- ໄດ້ຮູ້ເຖິງຂະບວນການເຮັດວຽກ ແລະ ບັນຫາທີ່ເກີດຂຶ້ນຂອງແບບຈຳລອງການແປຂໍ້ຄວາມຂຽນໄປເປັນຂໍ້ຄວາມສຽງ (Text-To-Speech).
- ໄດ້ແບບຈຳລອງການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ (Lao Speech-To-Text) ໃໝ່ ທີ່ມີປະສິດທິພາບ.
- ໄດ້ເວັບໄຊການແປຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ ແລະ ການແປຂໍ້ຄວາມສຽງພາສາລາວໄປເປັນຂໍ້ຄວາມຂຽນທີ່ມີປະສິດທິພາບ.



Work Process of TTS



Updated YourTTS



Website

05

ວິທີດໍາເນີນການຄົ້ນຄວ້າ

1)

ວິທີສຶກສາ ແລະ ຄົ້ນຄວ້າ

STEP 1

ການເກັບກຳ
ຂໍ້ມູນສຽງ



STEP 2

ສຶກສາ ແລະ ນຳໃຊ້ແບບ
ຈຳລອງຂອງ YourTTS

STEP 3

ປະເມີນປະສິດທິພາບຂອງ
ແບບຈຳລອງ



STEP 4

ນຳເອົາແບບຈຳລອງມາພັດທະນາ
ເປັນ Website

2) ການວິເຄາະຂໍ້ມູນ

STEP 2

ການປະມວນຜົນຟາຍສຽງຂັ້ນຕົ້ນ

ກ່ອນທີ່ຈະເຂົ້າສູ່ຂະບວນການປະມວນຜົນ ຕ້ອງໄດ້ກະກຽມດັ່ງນີ້:

- ✓ ກຳຈັດສຽງລົບກວນອ້ອມຂ້າງ (Eliminate Noise)
- ✓ ຕັດເອົາສະເພາະສຽງທີ່ເຮົາບັນທຶກ (ຫຼື ຕັດຊ່ອງວ່າງຂອງສຽງທີ່ບັນທຶກ)
- ✓ ການເຮັດໃຫ້ສຽງເປັນປົກກະຕິ



STEP 1

ການຈັດກຽມຂໍ້ມູນສຽງ

ຊຸດຂໍ້ມູນສຽງແມ່ນໄດ້ກຽມຢູ່ 3 ຊຸດຄື:

- ✓ metadata.csv ມີຂໍ້ມູນຢູ່ 100% ຫຼື 900 ຟາຍສຽງ
- ✓ metadata_train.csv ມີຂໍ້ມູນຢູ່ 95% ຫຼື 855 ຟາຍສຽງ
- ✓ metadata_eval.csv ມີຂໍ້ມູນຢູ່ 5% ຫຼື 45 ຟາຍສຽງ



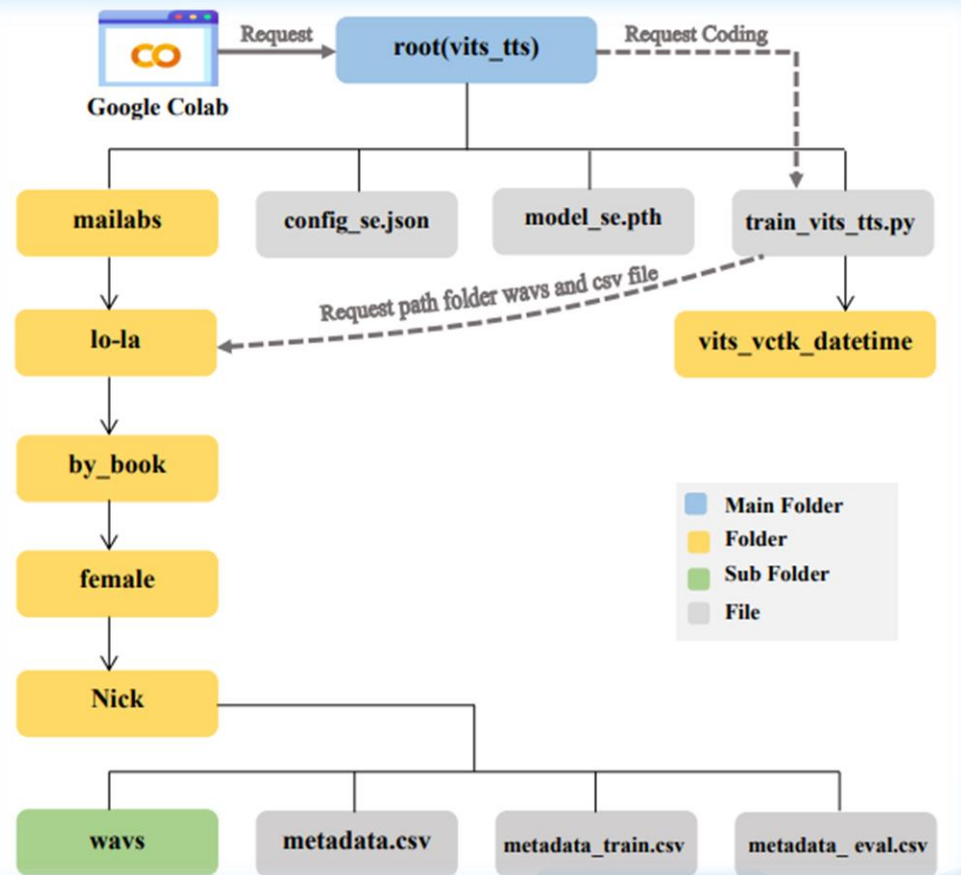
STEP 3

ການນຳໃຊ້ແບບຈຳລອງ YourTTS ແລະ ການປະມວນຜົນ

ການນຳໃຊ້ແບບຈຳລອງ YourTTS ແມ່ນໄດ້ນຳເອົາຊຸດຂໍ້ມູນຟາຍສຽງທີ່ໄດ້ມາມີທັງໝົດ 900 ຟາຍສຽງ ນຳໄປປະມວນຜົນ, ເຝິກສອນ ແລະ ທົດສອບແບບຈຳລອງ ໃນ Google Colab ໂດຍມີຂັ້ນຕອນການປະມວນຜົນດັ່ງຕໍ່ໄປນີ້:

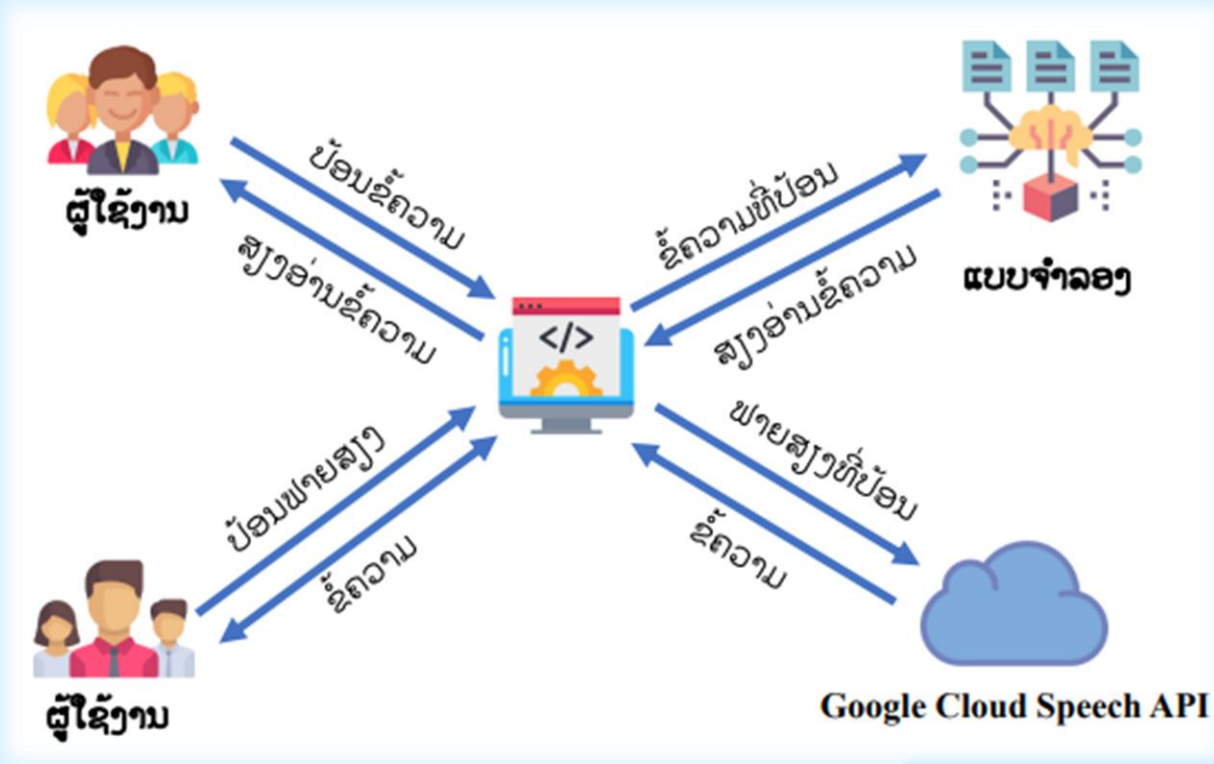
2) ການວິເຄາະຂໍ້ມູນ (ຕໍ່)

STEP 3 (ຕໍ່)



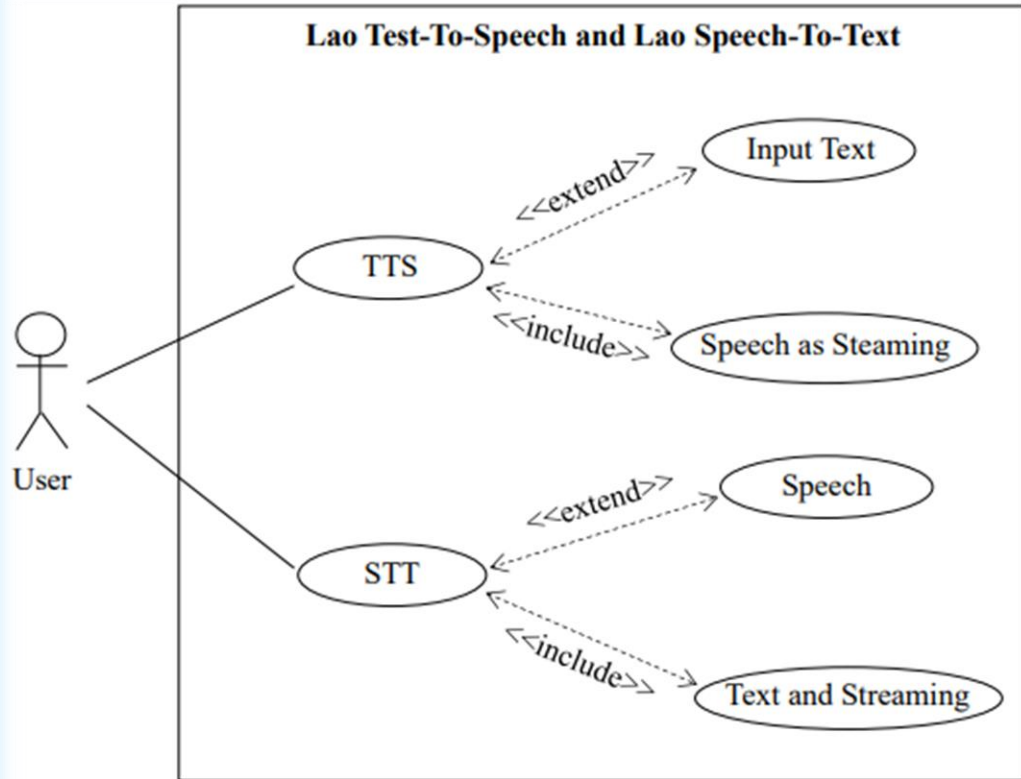
3) ການວິເຄາະລະບົບ

❖ ແຜນວາດລວມຂອງລະບົບ



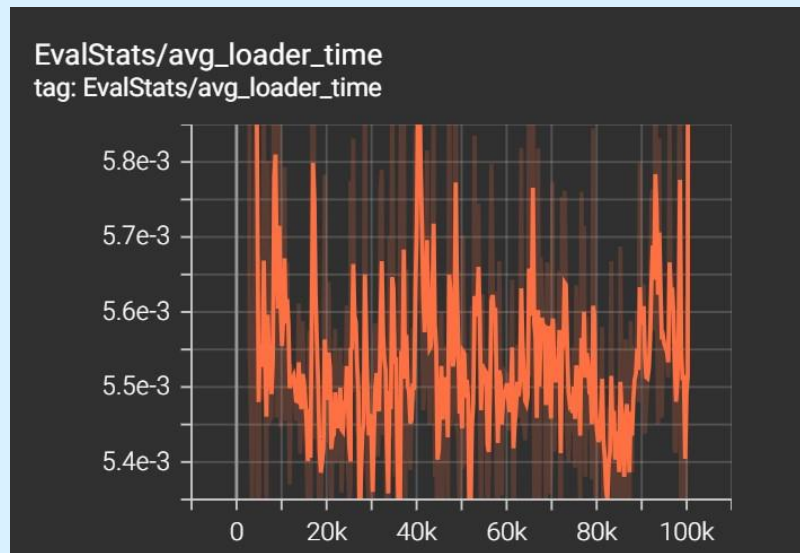
3) ການວິເຄາະລະບົບ (ຕໍ່)

❖ ແຜນວາດ Use Case Diagram

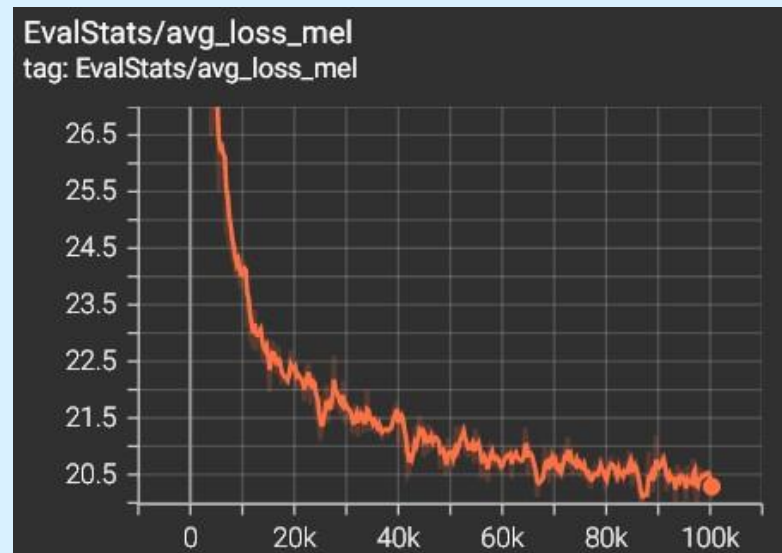


06

ຜົນຂອງການປະມວນຜົນແບບຈຳລອງ



❖ ຮູບພາບສະແດງເວລາສະເລ່ຍໃນການ
ໂຫຼດຂໍ້ມູນໃນລະຫວ່າງການທົດລອງ
(EvalStats/avg_loader_time)

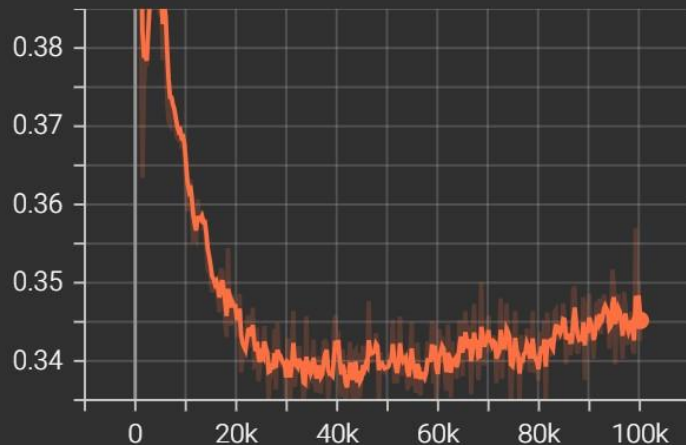


❖ ຮູບພາບສະແດງການສູນ
ເສຍ mel-spectrogram
(EvalStats/avg_loss_mel)

06

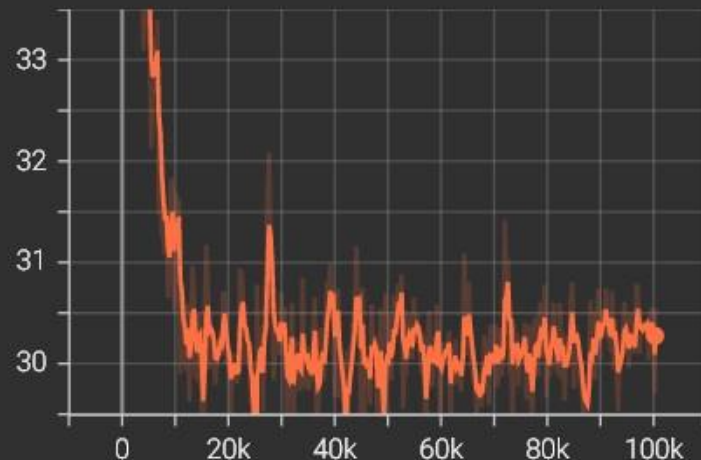
ຜົນຂອງການປະມວນຜົນແບບຈຳລອງ (ຕໍ່)

EvalStats/avg_loss_duration
tag: EvalStats/avg_loss_duration



❖ ຮູບພາບສະແດງໄລຍະເວລາການສູນເສຍ
ການຄາດຄະເນ ແລະ ໄລຍະເວລາຄວາມຈິງ
ພື້ນຖານຂອງ phoneme ໃນການເວົ້າທີ່ສ້າງ
ຂຶ້ນ (EvalStats/ avg_loss_duration)

EvalStats/avg_loss_1
tag: EvalStats/avg_loss_1



❖ ຮູບພາບສະແດງການສູນ
ເສຍ mel-spectrogram ໃນ
ລະຫວ່າງການທົດລອງ
(EvalStats/ avg_loss_1)

ຜົນຂອງການຮຽນຮູ້ຂອງແບບຈຳລອງ YourTTS ເຫັນວ່າມີ ຄ່າການເຝິກສອນ ໄດ້ຄົດເປັນຄ່າຕ່າງໆຄື: ເວລາສະເລ່ຍໃນການໂຫຼດຂໍ້ມູນ (avg_loader_time) ມີຄ່າ 0.0054 ວິນາທີ, ການສູນເສຍ mel-spectrogram (avg_loss_mel) ມີຄ່າ 20.29, ເວລາການສູນເສຍການຄາດຄະເນ ແລະ ໄລຍະເວລາຄວາມຈິງພື້ນຖານຂອງ phoneme ໃນການເວົ້າທີ່ສ້າງຂຶ້ນ (avg_loss_duration) ມີຄ່າ 0.34 ວິນາທີ ແລະ ການສູນເສຍ mel-spectrogram (avg_loss_l) ມີຄ່າ 30.27 ໃນການເຝິກສອນໃນຮອບທີ່ 100,000.

ຈາກຄ່າດັ່ງກ່າວສະແດງໃຫ້ເຫັນວ່າ ການຮຽນຮູ້ ແລະ ການຈື່ຈຳ ຂໍ້ຄວາມຂຽນພາສາລາວໄປເປັນຂໍ້ຄວາມສຽງ ແມ່ນເຮັດໄດ້ຢ່າງວ່ອງໄວ ແລະ ຂ້ອນຂ້າງມີປະສິດທິພາບທີ່ດີ.





❖ ຂໍ້ຈຳກັດຂອງການຄົ້ນຄວ້າ

- ການຮຽນຮູ້ຂອງແບບຈຳລອງ ຍັງບໍ່ສາມາດອ່ານຕົວເລກສາກົນໄດ້.
- ການຮຽນຮູ້ຂອງແບບຈຳລອງ ຍັງບໍ່ສາມາດອ່ານຂໍ້ຄວາມທີ່ພົມໃນຊ່ອງພົມຂໍ້ຄວາມເວັບໄຊ ທີ່ບໍ່ມີສຽງໃນຊຸດຂໍ້ມູນ (Dataset) ທີ່ເຝິກສອນ (Train) ໄດ້.
- ຂໍ້ຈຳກັດທາງດ້ານການອ່ານສັນຍາລັກ ຍັງບໍ່ສາມາດອ່ານສັນຍາລັກຕ່າງໆໄດ້.
- ຄ່າການເຝິກສອນທີ່ຄິດເປັນຄ່າຕ່າງໆຍັງບໍ່ສາມາດຕໍ່າກວ່ານັ້ນໄດ້



I DON'T
KNOW





❖ ຂໍສະເໜີໃນການຄົ້ນຄວ້າຕໍ່ໄປ

- ການແນະນຳແນວທາງການພັດທະນາໄດ້ແກ່ ການພັດທະນາ Noise Removal ໃນສ່ວນຂອງ Pre-processing ເພື່ອໃຫ້ສາມາດລົບຄື້ນສຽງລົບກວນ ໃນກໍລະນີສຽງທີ່ມີຄື້ນສຽງລົບກວນຈຳນວນຫຼາຍອອກໄປ ແລະ ເພື່ອໃຫ້ແບບຈຳລອງສາມາດອ່ານຂໍ້ຄວາມສຽງໄດ້ຢ່າງມີປະສິດທິພາບຫຼາຍຂຶ້ນ.
- ແນະນຳການຊອກຫາເຕັກນິກ ຫຼື Algorithm ກ່ຽວກັບ TTS ພາສາລາວ ທີ່ມີຄວາມຊັດເຈນ ແລະ ສາມາດປັບຄຳຄວາມໄວ ແລະ ຊ້າຂອງສຽງເວລາປະມວນອອກມາ.
- ແນະນຳການພັດທະນາໃນສ່ວນ Post-processing ທີ່ເປັນສ່ວນທີ່ສຳຄັນໃນການປັບພາສາລາວໃຫ້ຖືກຕ້ອງຕາມຫຼັກວັດຈະນານຸກົມພາສາລາວ.
- ສຳລັບການພັດທະນາໃນສ່ວນ STT ແມ່ນຢາກໃຫ້ພັດທະນາເພີ່ມເຕີມໃນສ່ວນທີ່ສາມາດແປງຂໍ້ຄວາມໄປເປັນຟາຍຕ່າງໆເຊັ່ນ: .txt, .docx, .doc, .pdf ແລະ ອື່ນໆ.



08 Demo



Q & A

Thank you