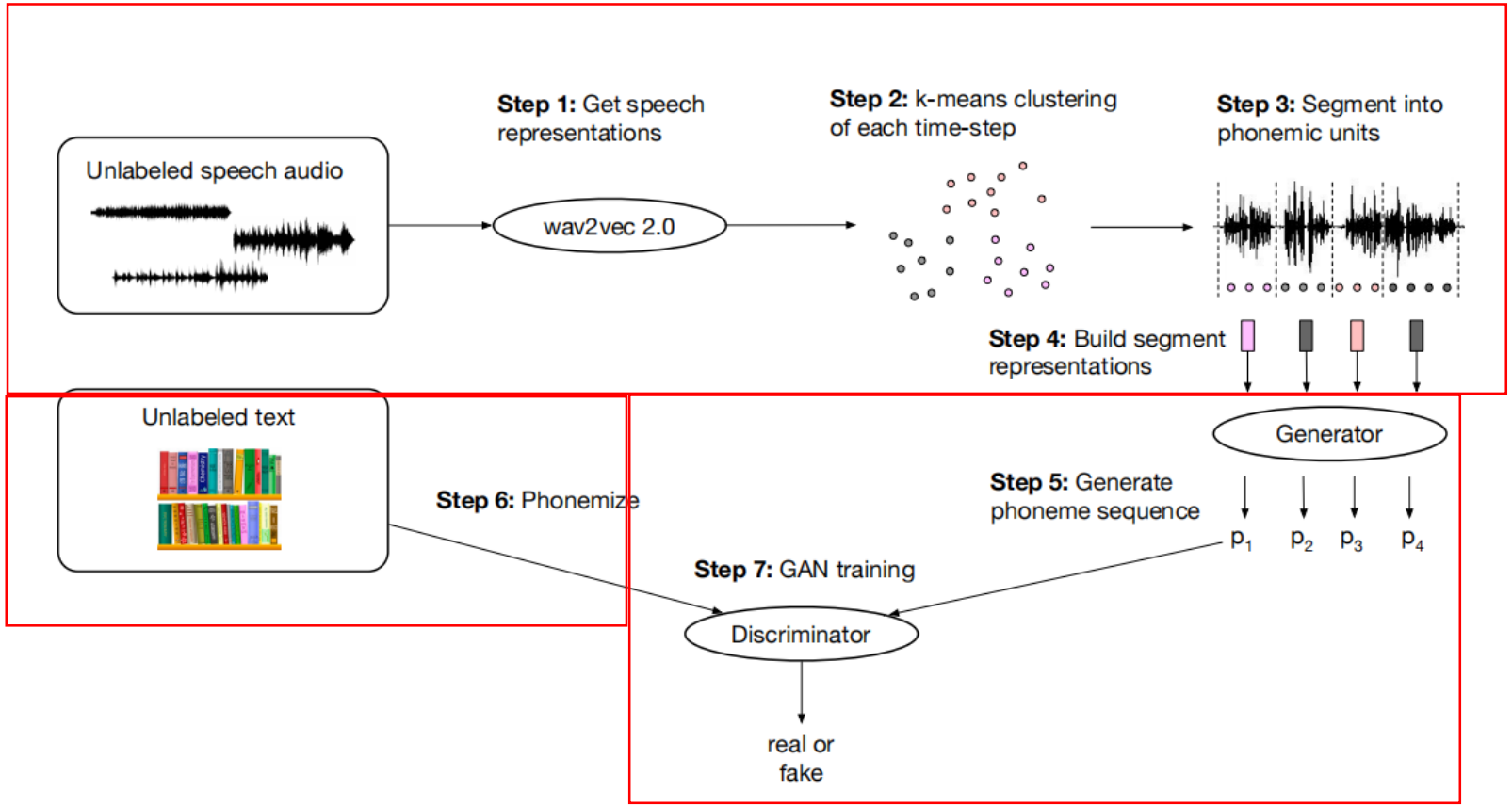
Wav2vec-U 实验报告

报告人：严子曦、李思瑞

Illustration of wav2vec Unsupervised

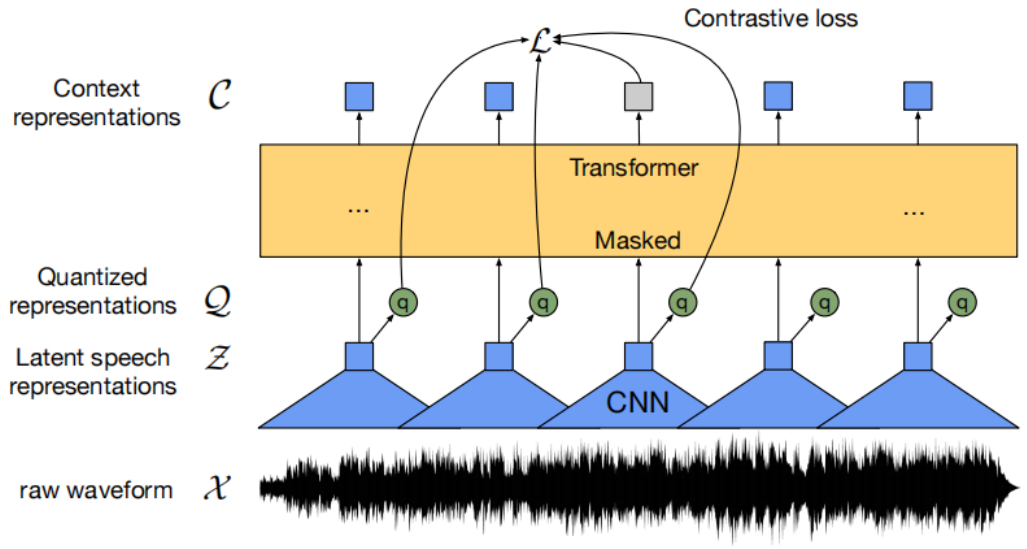
1、 Speech Audio Representations

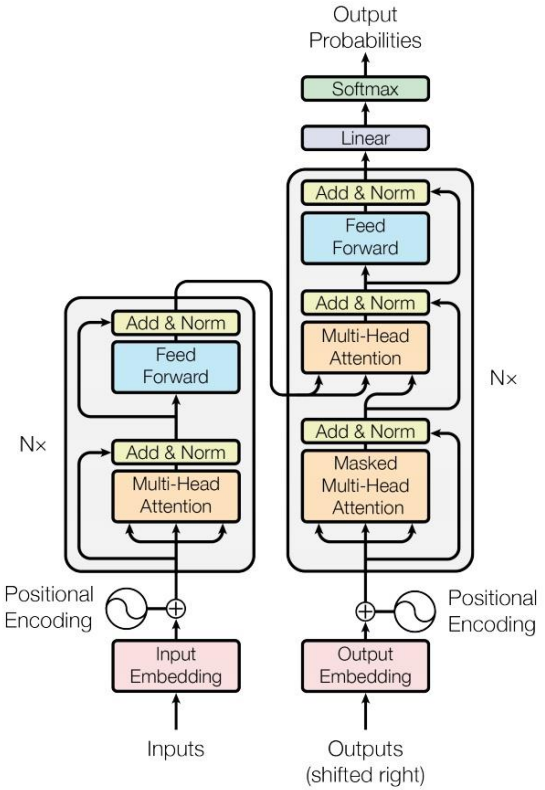


2、 Pre-processing the Text Data

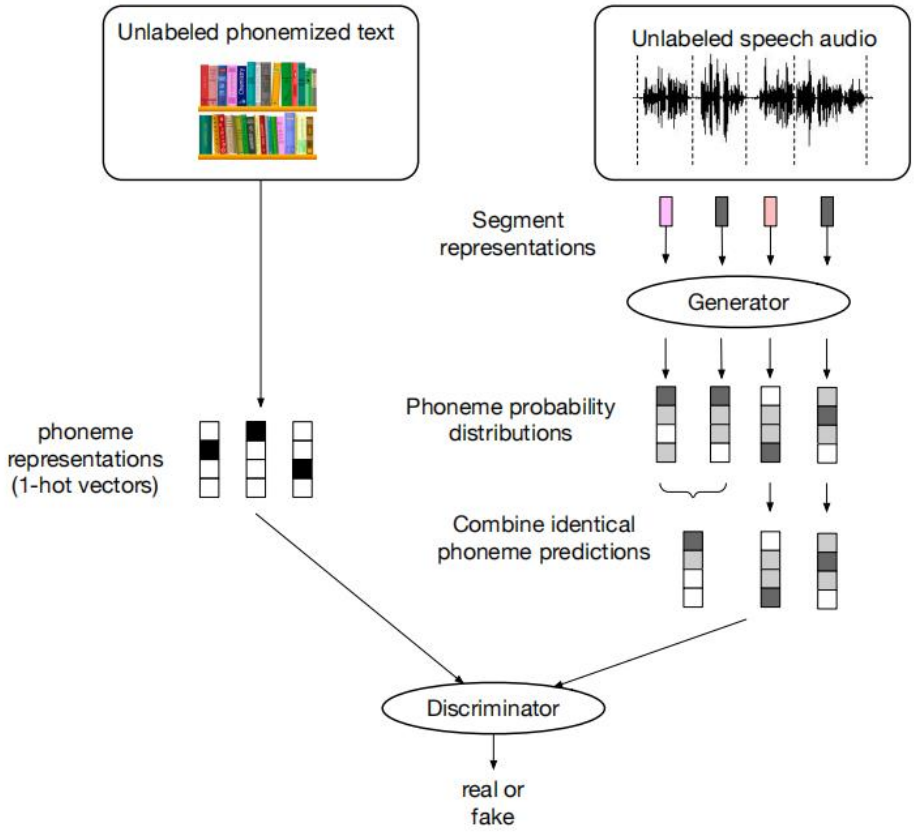
3、 Unsupervised Learning

Wav2vec2.0 and transformer





Unsupervised Learning



Objective

- the original GAN objective with a gradient penalty, a segment smoothness penalty and a phoneme diversity penalty

$$\min_{\mathcal{G}} \max_{\mathcal{C}} \mathop{\mathbb{E}}_{P^r \sim \mathcal{P}^r} [\log \mathcal{C}(P^r)] - \mathop{\mathbb{E}}_{S \sim \mathcal{S}} [\log (1 - \mathcal{C}(\mathcal{G}(S)))] - \lambda \mathcal{L}_{gp} + \gamma \mathcal{L}_{sp} + \eta \mathcal{L}_{pd}$$

$$\mathcal{L}_{gp} = \mathop{\mathbb{E}}_{\tilde{P} \sim \tilde{\mathcal{P}}} \left[\left(\|\nabla \mathcal{C}(\tilde{P})\| - 1 \right)^2 \right]$$

$$\mathcal{L}_{pd} = \frac{1}{|B|} \sum_{S \in B} -H_{\mathcal{G}}(\mathcal{G}(S))$$

$$\mathcal{L}_{sp} = \sum_{(p_t, p_{t+1}) \in \mathcal{G}(S)} \|p_t - p_{t+1}\|^2$$

Unsupervised Cross-Validation Metric

- We use the metric for early stopping and to select training hyper-parameters (λ, γ, η).

- We consider two quantities in our metric: language model entropy and vocabulary usage.

1) language model entropy: $H_{LM}(\mathcal{P})$

2) vocabulary usage: $U(\mathcal{P}) \in [0, 1]$

select training hyper-parameters、 utterance

$$\hat{\mathcal{P}} = \arg\min_{\mathcal{P}} H(\mathcal{P}) - \log U(\mathcal{P})$$

$$H(\mathcal{P}) < H(\hat{\mathcal{P}}) + \log\left(1.2 \times \frac{U(\mathcal{P})}{U(\hat{\mathcal{P}})}\right)$$

$$\arg\min_{\mathcal{P}} H_{LM}(\mathcal{P}) = \frac{1}{N_s}\sum_{j=1}^{N_s}\sum_{t=1}^{M} p_{LM}(p_t^j)\log p_{LM}(p_t^j), M = |P^j|, P^j = [p_1^j, \ldots, p_M^j]$$

TIMIT Phoneme Error Rate

Setup	Model	LM	core-dev	core-test
matched	wav2vec-U	4-gram	17.0	17.8
	wav2vec-U_now	4-gram	20.6	21.7
	wav2vec-U+ST	4-gram	11.3	12.0
	wav2vec-U+ST_now	4-gram	22.6	22.9
unmatched	wav2vec-U	4-gram	21.3	22.3
	wav2vec-U_now	4-gram	27.6	28.2
	wav2vec-U+ST	4-gram	13.8	15.0
	wav2vec-U+ST_now	4-gram	27.5	25.7

Model	LM	core-dev	core-test	all-test
wav2vec-U	4-gram	17.0	17.8	16.6
+ HMM	4-gram	13.7	14.6	13.5
+ HMM + HMM	4-gram	13.3	14.1	13.4
+ HMM resegment + GAN	4-gram	13.6	14.4	13.8
+ fine-tune	4-gram	12.0	12.7	12.1
+ fine-tune	-	12.1	12.8	12.0
+ fine-tune + fine-tune	-	12.0	12.7	12.0
+ HMM + fine-tune	-	11.3	11.9	11.3
+ HMM + fine-tune	4-gram	11.3	12.0	11.3

版本依赖问题

大型项目经常会依赖很多第三方的项目，有些第三方的项目并不会做版本的向下兼容，且如果依赖的第三方项目较多，存在各项目更新进度不一致的问题。在安装时需要注意两点，最好不使用最新版本，各项目的版本依赖最好能在安装前弄清楚，避免安装了一大半，后续的项目与前面项目出现版本依赖的冲突。

python环境版本退化问题

使用conda工具管理python环境在安装一些包时会导致python版本的退化，因此如果需要对conda管理的python环境进行及时的备份，避免python版本退化导致原来安装好的环境崩溃。

https://hub.docker.com/

https://huggingface.co/

git clone https://gitclone.com/github.com/pytorch/fairseq.git

git config --global url."https://gitclone.com/".insteadOf https://

git clone https://github.com/pytorch/fairseq.git

Issue

mail

谢谢大家观看