The background features a complex network of thin grey lines connecting various points, forming a web-like structure. Scattered throughout are several triangles of different sizes and orientations, some with solid grey outlines and others with hollow outlines. The overall aesthetic is clean, modern, and technical, suggesting a focus on data or technology.

Machine Learning assisted Cantopop Lyric Composition

LYU2101

Tang Ka Lok, 1155125745

01

Nature of Cantonese and Cantopop





Cantonese is a tone language

- A specific tone representing a unique pitch associated to a Chinese Character
 - The pitch is used to distinguish between different words and meanings
-



Tonal System of Cantonese

MARS

MERCURY

VENUS

JUPITER

fan1	fan2	fan3	fan4	fan5	fan6
昏	粉	訓	焚	奮	份

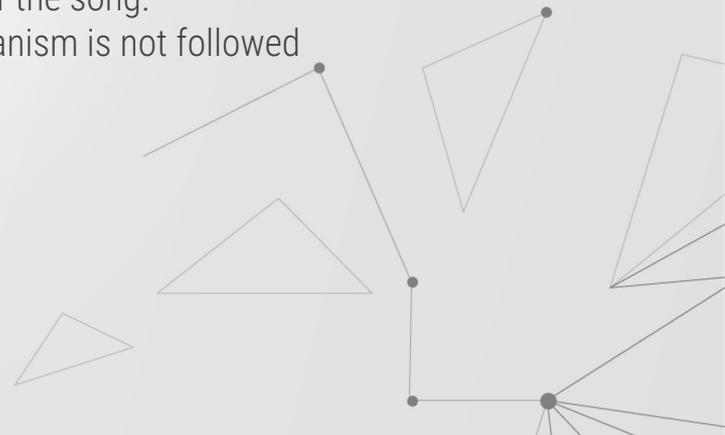
Figure 1.1: 6 characters having same syllable with 6 different tones

- Each Chinese Character is combining the syllable 'fan' and tone from 1 to 6 (Jyutping representation).
- 昏: dizzy
- 粉: pink
- 訓: train
- 焚: burn
- 奮: exert
- 份: portion



Tone-melody Matching Mechanism in Cantopop

- Melody contains the sequence of music notes with different pitch height.
 - Matching the tone of lyrics to the melody of the song.
 - The song is basically not understandable if the mechanism is not followed
-



Tone-melody Matching Mechanism

Original Lyrics

ngo5 mun4 si6 faai3 lok6 dik1 hou2 ji4 tung4 ngo5 mun4 tin1 tin1 jat1 hei2 go1 coeng3
我 們 是 快 樂 的 好 兒 童 我 們 天 天 一 起 歌 唱
(We are happy good children. We are singing together every day)

Music Notation



Lyrics that match the melody

ngo2 mun5 si6 faai3 lok3 dik6 hou2 ji5 tung3 ngo2 mun6 tin1 tin1 jat1 hei3 go3 coeng1
鵝 滿 是 快 烙 滴 好 耳 痛 鵝 悶 天 天 一 戲 個 窗
(not translatable)

Figure 1.2 Example of fitting lyrics with unmatched tone into the melody

Tone-melody Matching Mechanism

Numbered musical notation											
1		7		6		3		2		2	
1		7		7		-		-		6	
5		6		2		1		2		3	
3		6		1		2		5		2	
1		5		1		1		5			
那		夜		誰		將		酒		喝	
掉		因		此		我		講		得	
多		了									
(Who drank the alcohol that night, that's why I talk too much)											

Figure 1.3 Example of the steps to write Cantopop lyrics

Steps to write Cantopop Lyrics

1. Convert the melody to the tones of Cantonese based on the pitch of the music notes.
2. Fill in the lyrics that match the tones.



02

Model Description

Transformer: Self-attention Mechanism

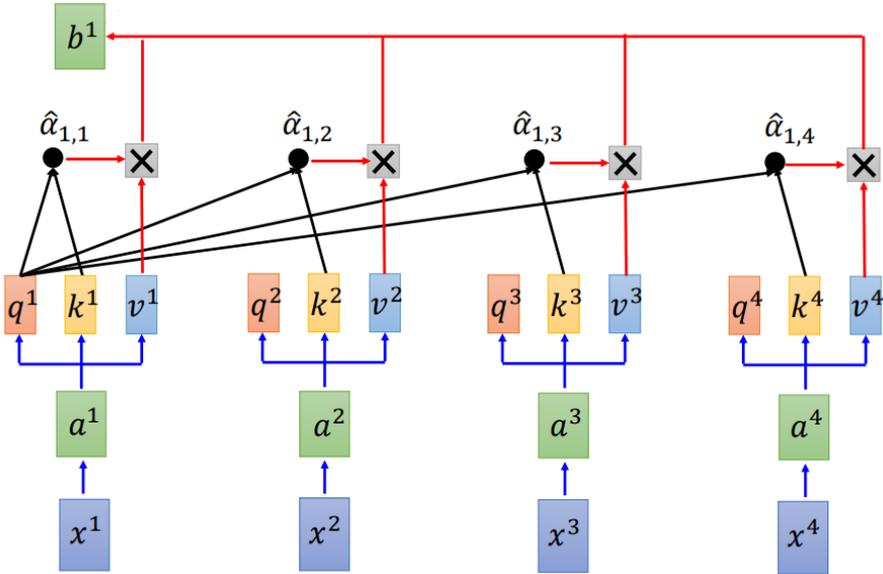


Figure 2.1 Self-Attention mechanism[1]

- With an input sequence passed into self-attention layer, the attention score between each input is calculated.
- With the attention calculated, the contextual information between inputs are captured and learned by the model.

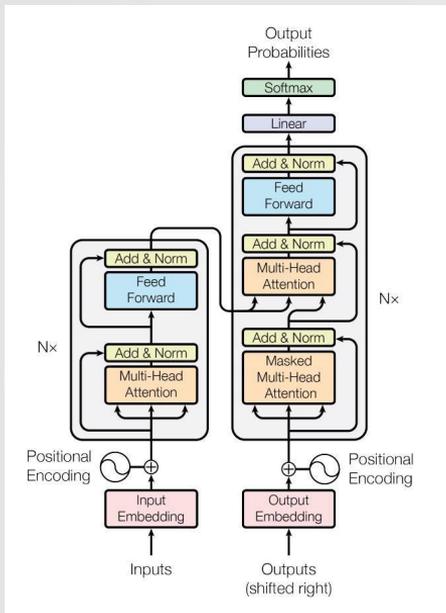


Figure 2.2 Transformer Architecture [2]

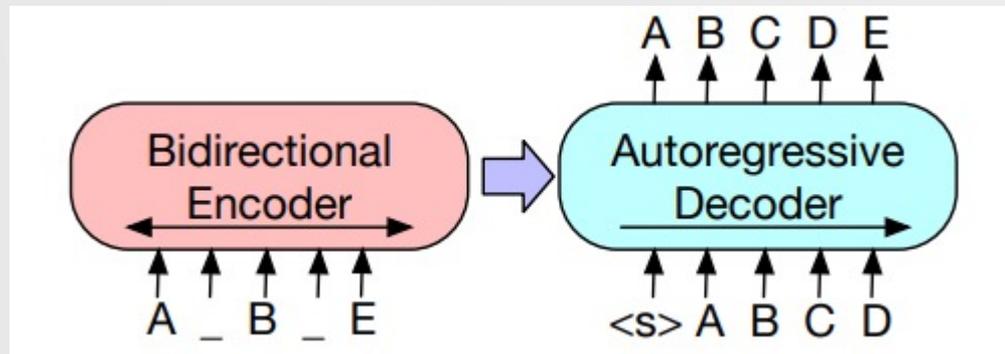


Figure 2.4 Bart [3]

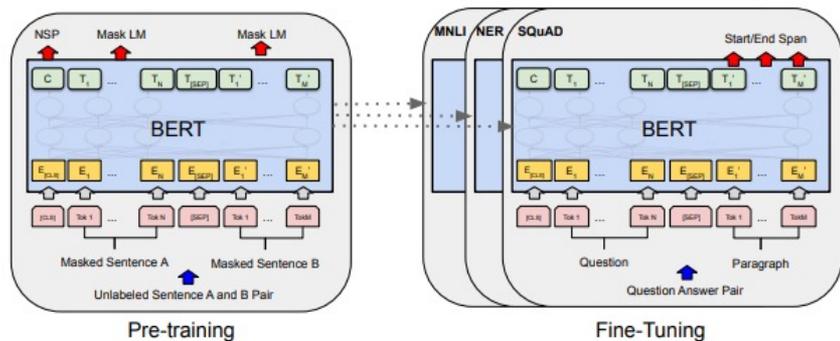


Figure 2.3 Bert [4]

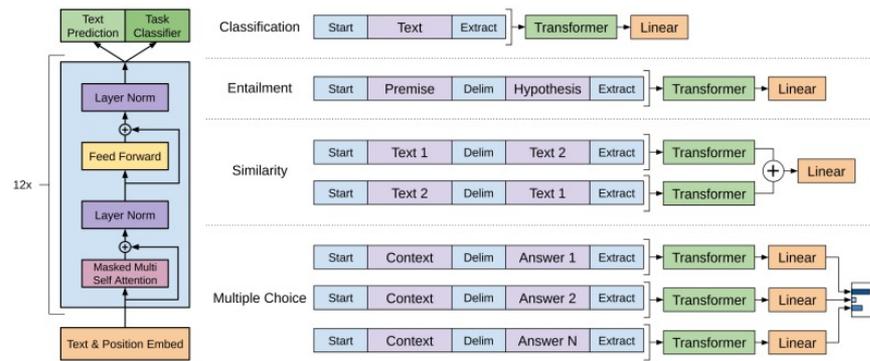


Figure 2.4 GPT*[5]

03

Methodology





Related Works

Given the previous words input, do the next word prediction to generate lyrics

Word-prediction-based Lyrics Generation

Takes the melody as input and generate corresponding lyrics as output. It captures the relationship between melody and lyrics.

Melody-based Lyrics Generation



Recap: Cantopop Lyrics Composition

Numbered musical notation																		
1 7			6 3 2 2			1 7			7 - - 6 5			4 1 7 7			6 5			5
↓																		
Tone																		
5 6			2 1 2			3 6			1 2			5 2 1			1 5			
↓																		
Lyrics																		
那夜			誰將酒			喝掉			因此			我講得			多了			
(Who drank the alcohol that night, that's why I talk too much)																		

Figure 3.1 Example of the steps to write Cantopop lyrics

Steps to write Cantopop Lyrics

1. Convert the melody to the tones of Cantonese based on the pitch of the music notes. (Easy)
2. Fill in the lyrics that match the tones. (Extremely Hard)

Advantages

Easier Data Preparation

- The tone of word can be extracted from the lyrics.
- Difficulties on getting data are reduced while maintaining the nature of Cantopop lyrics composition.

Expand Size of Dataset

- Dataset can be expanded from Cantopop lyrics data to all Chinese pop lyrics data.
- Melody and lyrics are indirectly related to each other
- Greatly increase the diversity of the model.

Training Approaches

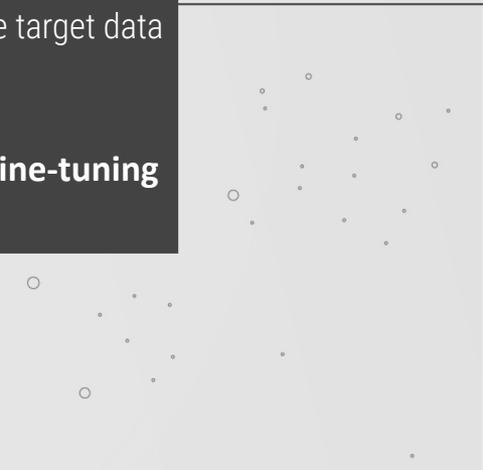


Directly train the model using the target data

Training from Scratch

First pre-train a model using in-domain data and then we fine-tune it to our specific task using the target data

Pre-training and Fine-tuning



04

Workflows



Data Preparation

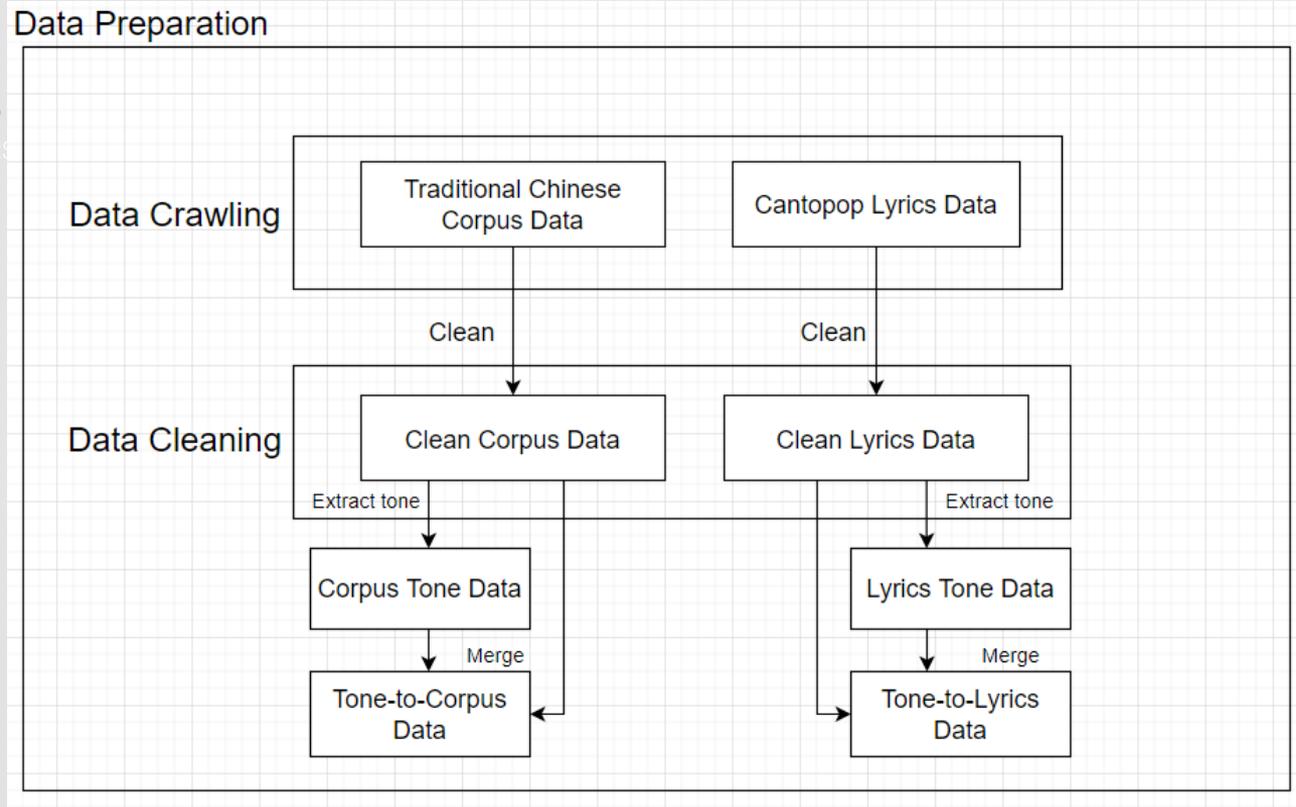


Figure 4.1 Flow chart of data preparation phase

Dataset Overview

10 GB

Traditional Chinese Corpus Data

200 MB

Tone-to-Corpus Data

15000

Lyrics / Tone-to-Lyrics Data



Example of Data Cleaning

Before cleaning:

MARS MERCURY VENUS

Figure 4.2 Lyrics before cleaning

陳奕迅
十年
作詞：林夕
作曲：陳小霞
編曲：陳輝陽

如果那兩個字沒有顫抖
我不會發現 我難受
怎麼說出口 也不過是分手

如果對於明天沒有要求
牽牽手就像旅遊
成千上萬個門口
總有一個人要先走

* 懷抱既然不能逗留
何不在離開的時候
一邊享受 一邊漂流

十年之前 我不認識你
更多更狂盡歌詞 在 ※ Mojim.com 魔鏡歌詞網
你不屬於我 我們還是一樣
陪在一個陌生人左右
走過漸漸熟悉的街頭

十年之後 我們是朋友
還可以問候 只是那種溫柔
再也找不到擁抱的理由
情人最後難免淪為朋友

REPEAT * #

直到和你做了多年朋友 才明白我的眼淚
不是為你而流 也為別人而流

[00:02.18]陳奕迅 - 十年
[00:14.79]如果那兩個字沒有顫抖
[00:19.02]我不會發現 我難受
[00:22.28]怎麼說出口 也不過是分手
[00:29.97]如果對於明天沒有要求
[00:34.87]牽牽手就像旅遊
[00:37.85]成千上萬個門口
[00:41.50]總有一個人要先走
[01:57.00][00:47.56]懷抱既然不能逗留
[02:00.51][00:50.83]何不在離開的時候
[02:03.54][00:53.75]一邊享受 一邊漂流
[02:10.79][01:00.70]十年之前 我不認識你
[02:14.53][01:04.73]你不屬於我 我們還是一樣
[02:19.03][01:09.26]陪在一個陌生人左右
[02:22.66][01:12.92]走過漸漸熟悉的街頭
[02:26.26][02:11.02]十年之後 我們是朋友
[02:29.89][02:10.34]還可以問候 只是那種溫柔
[02:34.44][02:14.64]再也找不到擁抱的理由
[02:38.30][02:18.66]情人最後難免淪為朋友
[02:48.70]直到和你做了多年朋友 才明白我的眼淚
[02:55.37]不是為你而流 也為別人而流

After cleaning:

如果那兩個字沒有顫抖
我不會發現 我難受
怎麼說出口 也不過是分手

如果對於明天沒有要求
牽牽手就像旅遊
成千上萬個門口
總有一個人要先走

懷抱既然不能逗留
何不在離開的時候
一邊享受 一邊漂流

十年之前 我不認識你

你不屬於我 我們還是一樣
陪在一個陌生人左右
走過漸漸熟悉的街頭

十年之後 我們是朋友
還可以問候 只是那種溫柔
再也找不到擁抱的理由
情人最後難免淪為朋友

直到和你做了多年朋友 才明白我的眼淚
不是為你而流 也為別人而流

Figure 4.3 Lyrics After cleaning

Example of Tone-to-Text Data

Original Lyrics:

```
如果爭執 傷口交給我修好\n如果翻風 關起窗給你擁抱\n如果不安 誠實請給我  
知道\n做不好再做 無奈並未代表所需要的好\n\n如果紛擾 耳朵只聽你傾訴\n如果傷風 病菌分給我都好\n如果焦急和暴躁 如果消失能令你息怒\n我可以迴避到避世島\n\n旁人 眼裡 為人 再好\n與愛情無關沒有用途\n投懷 送抱 然而 跌倒\n好的傻子忘記了愛是殘酷\n\n原來奉獻多多少少依然徒勞  
兩個世界也分開\n\n任我花精神期待 清清楚楚辛辛苦苦不會帶來憐愛\n原來是我多此一舉  
傷痕累累要我尷尬也應該\n對你的好仍存在 好先生與愛情競賽 好先生太過難被愛\n始終相信你存在故我在\n\n原來奉獻多多少少依然徒勞  
兩個世界也分開\n\n最固執感情無奈 一起相處有權任性  
不會有權被愛\n原來是我多此一舉  
傷痕累累要我尷尬也應該\n你要的好難被替代 我有的好最後化做 塵埃
```

Figure 4.4 Original Lyrics Data

Extracted Tones:

```
4 2 1 1 1 2 1 1 5 1 2 \n 4 2 1 1 1 2 1 1 5 2 5 \n 4 2 1 1 4 6 2 1 5 1 6 \n 6 1 2 3 6 5 6 6 6 6 2 2 1 3 2 \n 4 2 1 2 5 2 2 1 5 1 3 \n 4 2 1 1 6 2 1 1 5 1 2 \n 4 2 1 1 4 6 3 4 2 1 1 4 6 5 1 6 \n 5 2 5 4 6 2 6 3 2 \n 4 4 5 5 6 4 3 2 \n 5 3 4 5 1 6 5 6 4 \n 4 4 3 5 4 4 3 2 \n 2 1 4 4 4 3 5 3 6 4 6 \n 4 4 6 3 1 1 2 2 1 4 4 4 5 3 3 5 1 1 \n 6 5 1 1 4 4 6 1 1 2 2 1 2 2 1 5 3 4 4 3 \n 4 4 6 5 1 2 1 2 1 4 4 4 3 5 3 5 1 1 \n 3 5 1 2 4 4 6 2 2 3 1 2 1 3 2 3 \n 2 3 \n 1 1 1 2 3 2 1 1 3 2 3 1 1 1 1 2 1 2 2 5 1 2 \n 4 6 1 1 4 6 5 4 2 1 1 4 6 3 1 6 \n 5 2 3 4 6 5 6 5 2 \n 6 4 2 1 4 6 2 5 \n 3 4 4 2 6 5 3 1 \n 4 4 1 2 5 4 1 6 \n 5 5 6 4 2 1 2 2 5 3 2 \n 1 4 4 6 3 1 1 2 2 1 4 4 5 3 3 5 1 1 \n 6 5 1 1 4 4 6 1 1 2 2 1 1 2 2 1 5 3 4 4 3 \n 4 4 6 5 1 2 1 2 1 4 4 4 3 5 3 3 5 1 1 \n 3 5 1 2 4 4 6 \n 2 1 1 5 3 4 3 3 2 1 1 3 3 4 5 3 \n 2 1 2 3 5 4 6 3 5 6 \n 4 4 6 3 1 1 2 2 1 4 4 4 5 3 3 5 1 1 \n 3 3 1 2 4 5 6 1 2 2 3 5 4 6 3 1 5 5 4 5 3 \n 4 4 4 6 5 1 2 1 2 1 4 4 3 5 3 3 5 1 1 \n 5 3 1 2 4 5 3 6 5 1 2 3 6 3 6 4 1
```

Figure 4.5 Extracted Tone Data

Tone-to-Text Data:

```
{  
  "tone": "4 2 1 1 1 2 1 1 5 1 2 \n 4 2 1 1 1 2 1 1 5 2 5 \n 4 2 1 1 4 6 2 1 5 1 6 \n 6 1 2 3 6 5 6 6 6 6 2 2 1 3 2 \n 4 2 1 2 5 2 2 1 5 1 3 \n 4 2 1 1 6 2 1 1 5 1 2 \n 4 2 1 1 4 6 3 4 2 1 1 4 6 5 1 6 \n 5 2 5 4 6 2 6 3 2 \n 4 4 5 5 6 4 3 2 \n 5 3 4 5 1 6 5 6 4 \n 4 4 3 5 4 4 3 2 \n 2 1 4 4 4 3 5 3 6 4 6 \n 4 4 6 3 1 1 2 2 1 4 4 4 5 3 3 5 1 1 \n 6 5 1 1 4 4 6 1 1 2 2 1 2 2 1 5 3 4 4 3 \n 4 4 6 5 1 2 1 2 1 4 4 4 3 5 3 3 5 1 1 \n 3 5 1 2 4 4 6 2 2 2 3 1 2 1 3 2 3 \n 4 \n 1 1 1 1 2 3 2 1 1 3 2 3 1 1 1 1 2 2 5 1 2 \n 4 6 1 1 4 6 5 4 2 1 1 4 6 3 1 6 \n 5 2 3 4 6 5 6 5 2 \n 6 4 2 1 4 6 2 5 \n 4 4 2 1 4 6 2 5 3 1 \n 4 4 1 2 5 4 1 6 \n 5 5 6 4 2 1 2 2 5 3 2 \n 4 4 6 3 1 1 2 2 1 4 4 4 5 3 3 5 1 1 \n 6 5 1 1 4 4 6 1 1 2 2 1 1 2 2 1 5 3 4 4 3 \n 4 4 6 5 1 2 1 2 1 4 4 3 5 3 3 5 1 1 \n 3 5 1 2 4 4 6 \n 2 1 1 5 3 4 3 3 2 1 1 3 3 4 5 3 \n 2 1 2 3 5 4 6 3 5 6 \n 4 4 6 3 1 1 2 2 1 4 4 4 5 3 3 5 1 1 \n 3 3 1 2 4 5 6 1 2 2 3 5 4 6 3 1 5 5 4 5 3 \n 4 4 6 5 1 2 1 2 1 4 4 4 3 5 3 3 5 1 1 \n 5 3 1 2 4 5 3 6 5 5 1 2 3 6 3 6 4 1",  
  "lyrics": "如果爭執 傷口交給我修好\n如果翻風 關起窗給你擁抱\n如果不安 誠實請給我  
知道\n做不好再做 無奈並未代表所需要的好\n\n如果紛擾 耳朵只聽你傾訴\n如果傷風 病菌分給我都好\n如果焦急和暴躁 如果消失能令你息怒\n我可以迴避到避世島\n\n旁人 眼裡 為人 再好\n與愛情無關沒有用途\n投懷 送抱 然而 跌倒\n好的傻子忘記了愛是殘酷\n\n原來奉獻多多少少依然徒勞  
兩個世界也分開\n\n任我花精神期待 清清楚楚辛辛苦苦不會帶來憐愛\n原來是我多此一舉  
傷痕累累要我尷尬也應該\n對你的好仍存在 好先生與愛情競賽 好先生太過難被愛\n始終相信你存在故我在\n\n原來奉獻多多少少依然徒勞  
兩個世界也分開\n\n最固執感情無奈 一起相處有權任性  
不會有權被愛\n原來是我多此一舉  
傷痕累累要我尷尬也應該\n你要的好難被替代 我有的好最後化做 塵埃"
```

Figure 4.6 Tone-to-Text Data

First Stage: GPT-2 (Language Model)

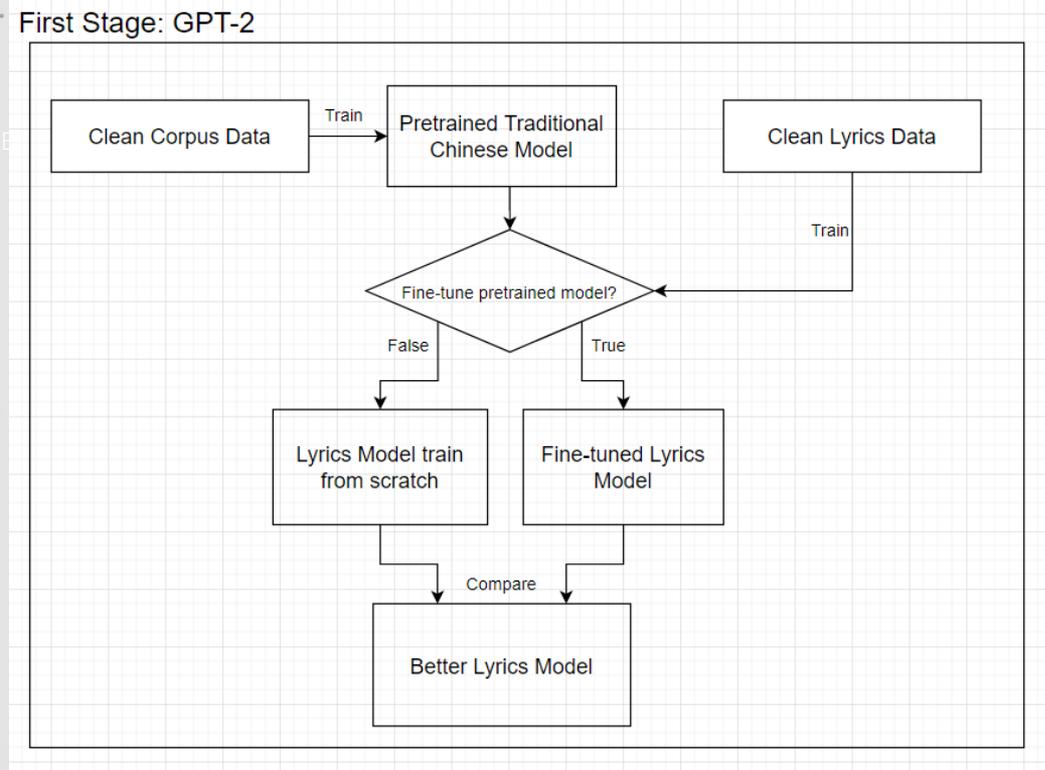


Figure 4.7 Flow chart of GPT-2 model phase

GPT-2 Evaluation

GPT2 trained from scratch

Fine-tuned GPT2

BLEU

0.2

4.8

Perplexity

343.5803

235.6772



Second Stage: Bart (Seq2Seq Model)

Second Stage: Transformer-based Seq2Seq model

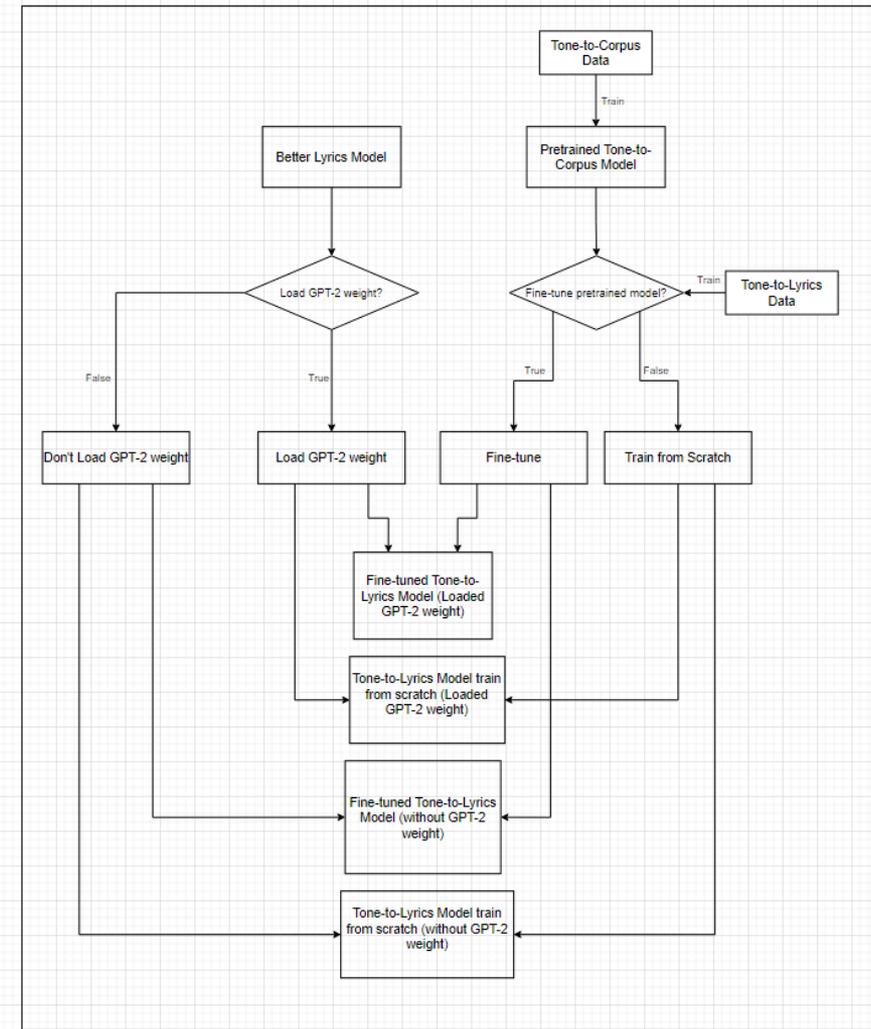


Figure 4.8 Flow chart of Bart model phase

Second Stage: Bart

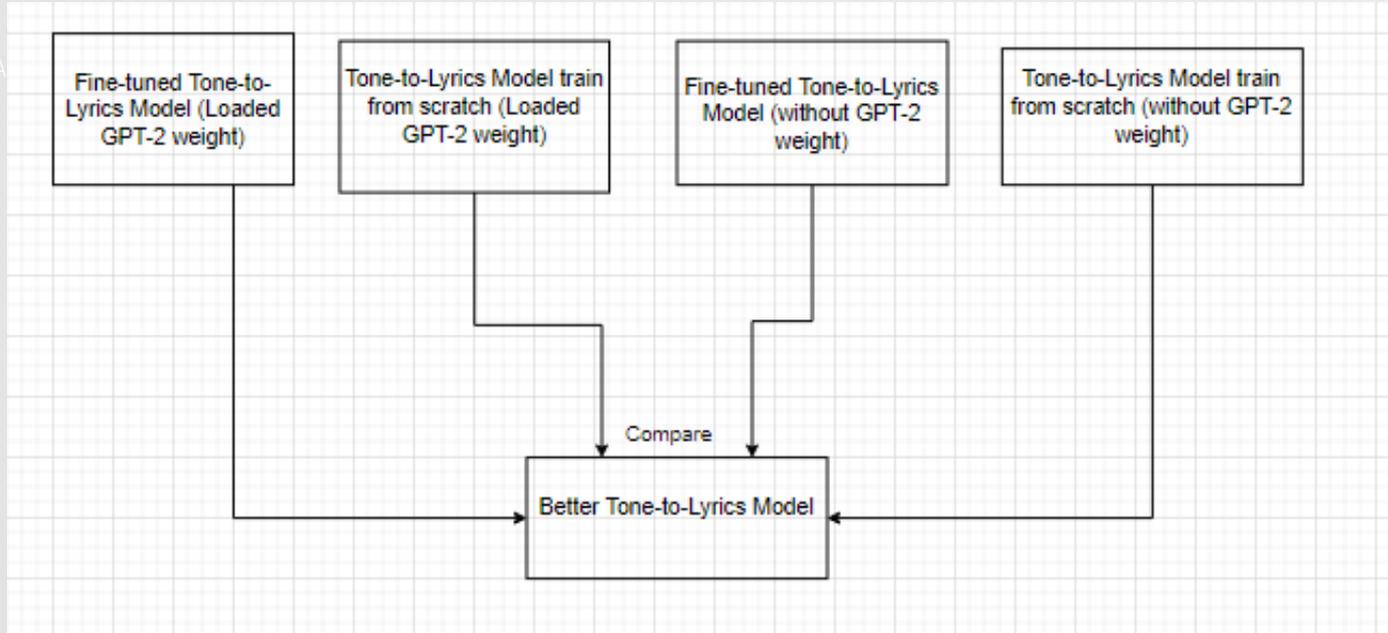


Figure 4.9 Flow chart of Bart models comparison

Second Stage: Bart



Figure 4.10 Training Loss

Second Stage: Bart

MARS

train/final_loss

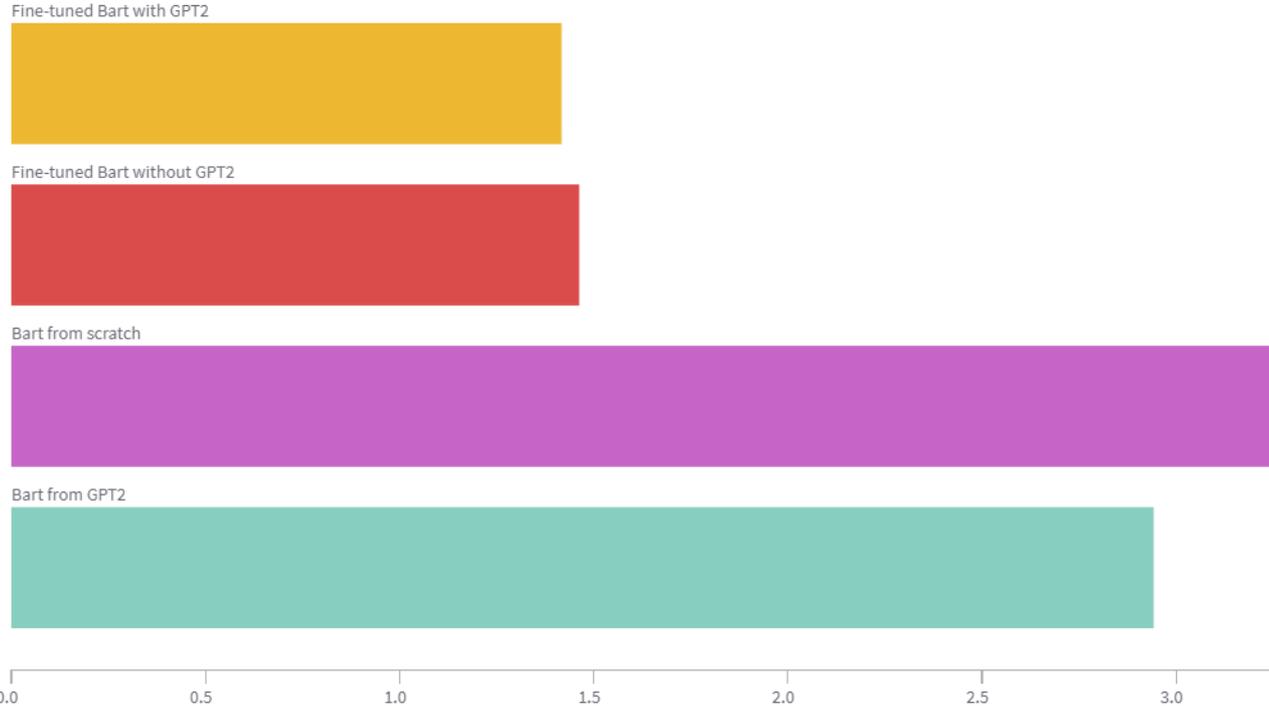


Figure 4.11 Training Final Loss

Second Stage: Bart

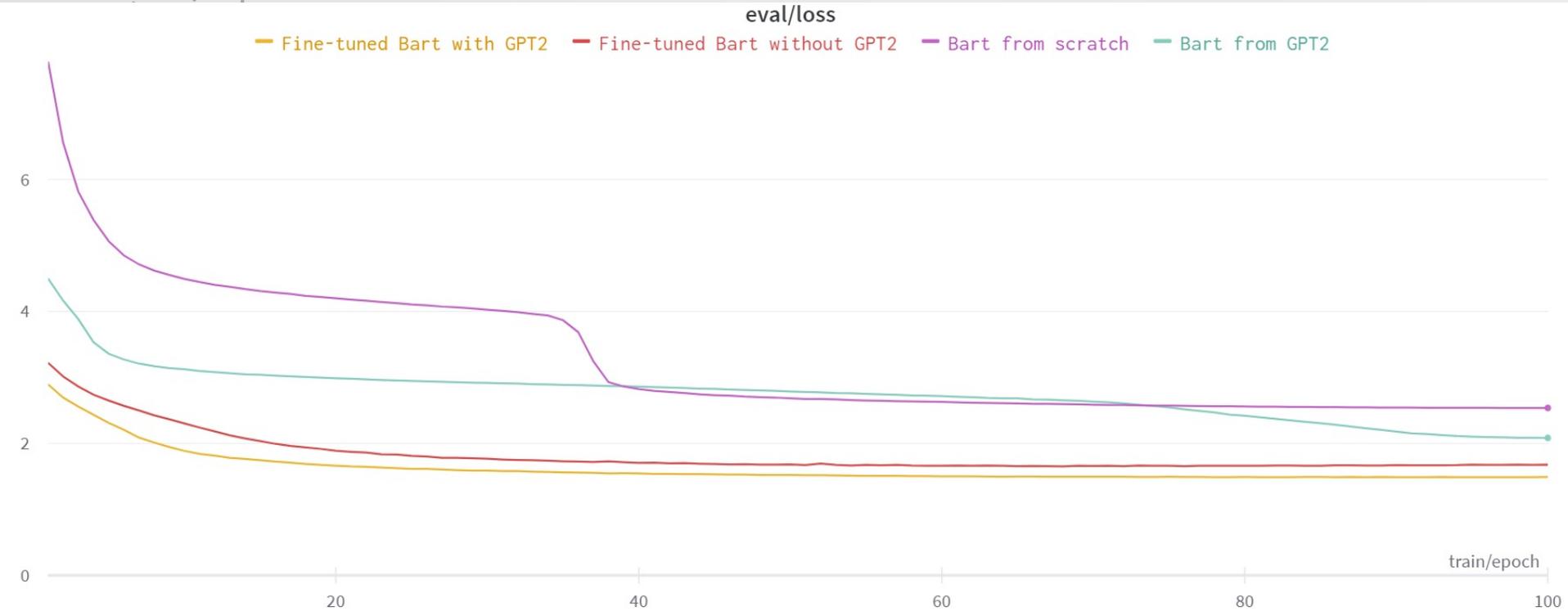


Figure 4.12 Validation Loss

Second Stage: Bart

eval/final_loss

Fine-tuned Bart with GPT2



Fine-tuned Bart without GPT2



Bart from scratch

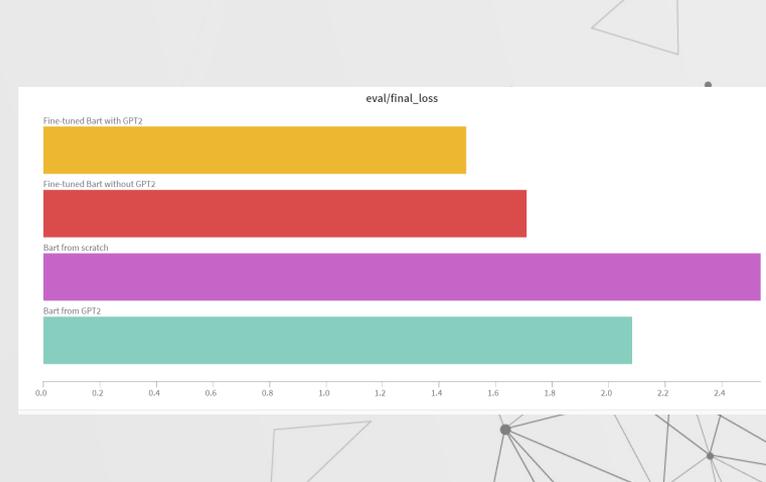
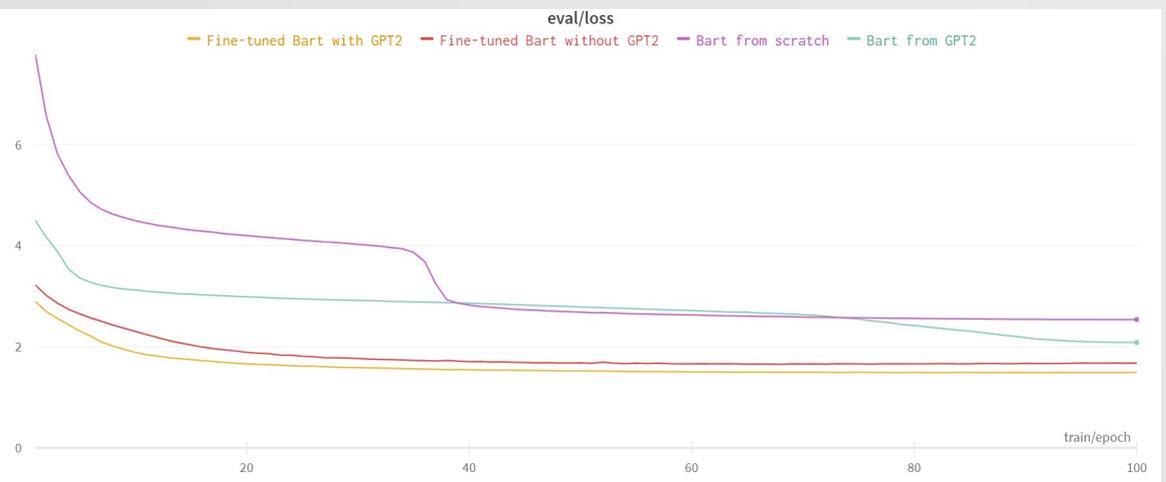
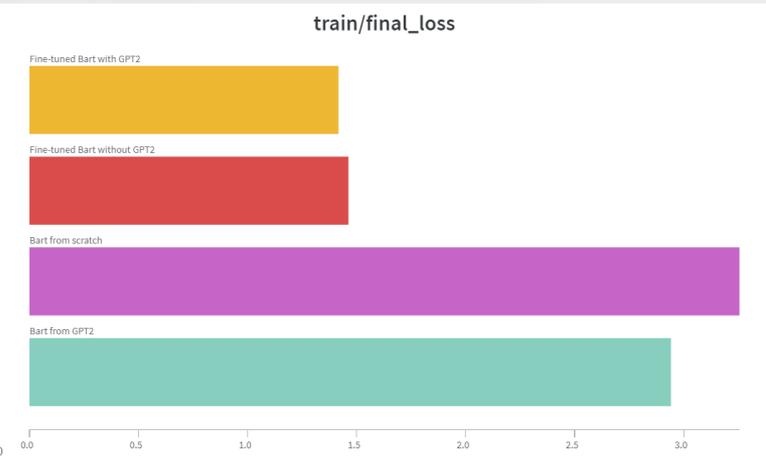


Bart from GPT2



0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4

Figure 4.13 Validation Final Loss



Bart Evaluation

	Bart trained from scratch	Bart trained from a model loaded GPT-2 weight	Bart fine-tuned from a pretrained model	Bart fine-tuned from a pretrained model loaded GPT-2 weight
BLEU	0.3	0.3	0.7	0.8
Perplexity	674.8038	635.6714	344.2669	337.1880
Tone Accuracy	0.98975	0.9877	0.99655	0.9938



Samples

Original Lyrics

我不如過街老鼠

人人亦是個判官審判我壞處

令你羞恥有餘

如果抱住誰都叫住

Input

5 1 4 3 1 5 2

4 4 6 6 3 3 1 2 3 5 6 3

6 5 1 2 5 4

4 2 5 6 2 1 3 6



Sample Output

.Bart trained from scratch without GPT-2 weight

Sample 1

我身從變的吻現
唯求命運氣轉一顆愛你沒變
願你一起你能
而戀你亦始終接近

Sample 2

無聲其實不會相
其情是別要怪不可算無事過
恨與深相距離
情可會做可不變動

Bart trained from a model loaded GPT-2 weight

Sample 1

我的情節分秒到
情形就像要跳得到跳上就快
為了擔險了迷
眉苦了就想不要亂

Sample 2

我的頭看不下這
情人沒問過去多久卻也沒怨
像我不想已從
還好你沒怎麼過沒

Bart fine-tuned from a pretrained model without GPT-2 weight

Sample 1

冷風流過的雨水
原來就是帶著一起向我蕩去
夜裡奔走旅程
回到你在這一個字

Sample 2

我的微笑不渺小
原來命運要看多少次你重看
沒有知己哪尋
而這已是誰的寄望

Bart fine-tuned from a pretrained model loaded GPT-2 weight

Sample 1

曙光融化心裡透
遙遙望外雪卻不懂放下內疚
在你的口裡留
留到你在此刻照舊

Sample 2

無辜來世將永久
流離樹絕世欠生死去了又見
沒有花果也甜
回首也沒感恩怨恨

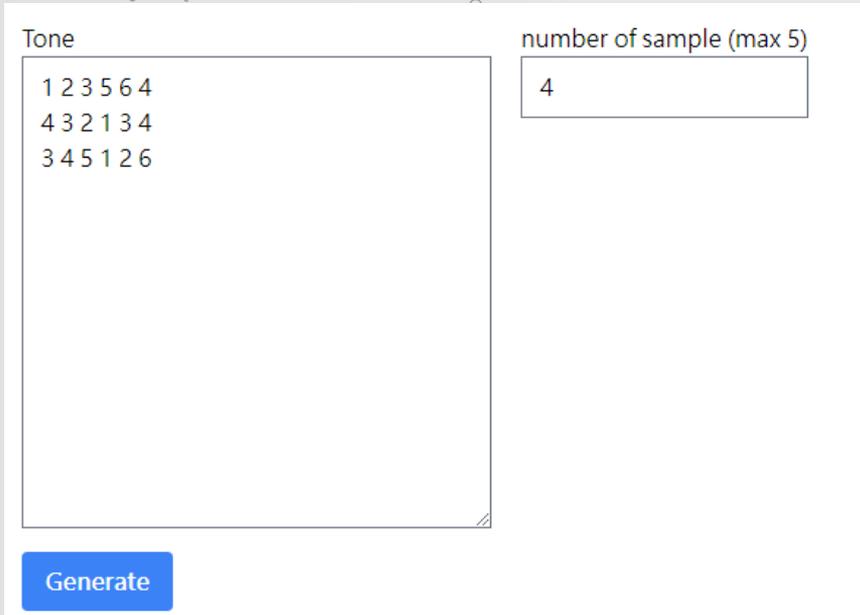


05

Tone2Cantopop

web tool for public to generate lyrics

Lyrics Generation



The input interface consists of a white rectangular area with a thin black border. On the left side, under the label 'Tone', there is a text area containing three lines of numbers: '1 2 3 5 6 4', '4 3 2 1 3 4', and '3 4 5 1 2 6'. On the right side, under the label 'number of sample (max 5)', there is a text input field containing the number '4'. Below the text area is a blue button with the word 'Generate' in white text.

Tone

1 2 3 5 6 4
4 3 2 1 3 4
3 4 5 1 2 6

number of sample (max 5)

4

Generate

Figure 5.1 Input interface for lyrics generation

-
- Input the tones converted from the melody
 - Input the number of sample to be generated

Lyrics Generation

1 2 3 4

speak

不想對你入懷
還要怎麼說明
再和你一起睡

1 2 3 4

speak

傷透過我未來
唇舌竟裝作零
叫停下的好夢

Figure 5.2 Tabs to toggle between generated samples

Text-to-speech

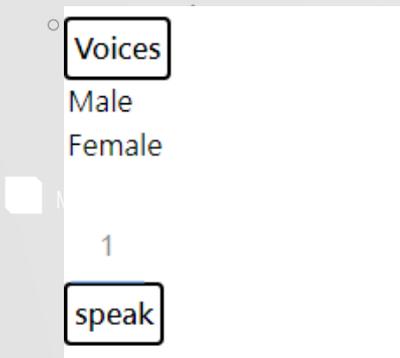


Figure 5.3 Text-to-speech interface

JUPITER

- Two voices are provided
- Spoken word highlighting feature is provided

我不能夠跟你走
從來沒話說對不起對你做錯
沒有一種理由
如果你是否不接受
我不能已不記得

Figure 5.4 Spoken word highlight feature

Tone Comparison

Input Tone	Output Tone	Tone Accuracy
1 2 3 5 6 4	1 2 3 5 6 4	88.89%
4 3 2 1 3 4	4 3 2 1 3 4	
3 4 5 1 2 6	3 4 5 2 1 6	

Figure 5.5 Tone comparison feature

- Wrong output tone is highlighted in red
- Overall Tone Accuracy is calculated

The background features abstract geometric patterns. In the top-left corner, there is a cluster of small circles and dots of varying sizes. In the bottom-right corner, there is a network of interconnected lines forming a complex web, with several triangles of different sizes and orientations scattered throughout. The overall aesthetic is clean and modern, using a light gray color palette.

Live Demonstration

06

Limitation



Size of dataset

10 GB

Traditional Chinese Corpus Data

200 MB

Tone-to-Corpus Data

15000

Lyrics / Tone-to-Lyrics Data



Multiple possible tones of a Chinese character

音節 (香港語言學學會)		粵音	根據	同音字	相關音節	
zung1		黃(p.47) 同(p.1) 李(p.44) 何(p.342)	鐘, 樓, 樓 [46..]	--選擇--	中土, 中心, 中天[32..]	
zung3		黃(p.47) 同(p.1) 李(p.44) 何(p.343)	鍾, 鍾, 癩 [11..]	--選擇--	中邪, 中肯, 中風[7..]	

Figure 6.1 Example of a Chinese character with multiple tones [6]

- Example: 中 (middle)
 - 中心 (center) and 中風 (stroke)
- Character-based Tokenization is used
- Word level / Sentence level Tokenization may solve the problem



07

Possible Future Development



Possible Future Development

- Theme/Title based lyrics generation
- Lyrics generation with partly finished lyrics
- Model Configuration Adjustment

Model improvement

- Account system
- Rating system



App Improvement



Conclusion

- Tone-based lyrics generation
 - A tailor made approach for Cantopop lyrics generation.
 - Sequence-to-Sequence Models (Bart)
 - Trained with Training from scratch, Pre-training and Fine-tuning
 - Tone2Cantopop
 - A Web tool open to public
- 

The background features a complex network of thin grey lines connecting various-sized grey dots. The dots are scattered across the slide, with some appearing as larger nodes and others as smaller points. The lines form a web-like structure that is denser on the left and right sides and sparser in the center. The overall aesthetic is clean, modern, and technical.

THANKS

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.



Reference

[1] Hung-yi Lee, *Transformer*, (Jun. 01, 2019). Accessed: Nov. 23, 2021. [Online Video]. Available: <https://www.youtube.com/watch?v=ugWDIIOHtPA>

[2] A. Vaswani *et al.*, ‘Attention Is All You Need’, *arXiv:1706.03762 [cs]*, Dec. 2017, Accessed: Nov. 15, 2021. [Online]. Available: <http://arxiv.org/abs/1706.03762>

[3] M. Lewis *et al.*, ‘BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension’, *arXiv:1910.13461 [cs, stat]*, Oct. 2019, Accessed: Nov. 15, 2021. [Online]. Available: <http://arxiv.org/abs/1910.13461>

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding’, *arXiv:1810.04805 [cs]*, May 2019, Accessed: Nov. 15, 2021. [Online]. Available: <http://arxiv.org/abs/1810.04805>

[5] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, ‘Improving Language Understanding by Generative Pre-Training’, p. 12.

[6] ‘粵語審音配詞字庫’ . <https://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/> (accessed Dec. 01, 2021).

