Rap Machine

CMPS242

De Huo
Ke Wang
Nursultan Kabylkas
Ramesh K. Jayaraman
Yuan Yang
Introduction

Proposal: train image captioning and poem generation models to generate rap lyrics based on the image.
Quick overview

Image Captioning was implemented by Inception V3 architecture.

Lyrics generation was implemented by seq2seq encoder-decoder RNN model.
Work split

Image Captioning:
- De Huo
- Ke Wang
- Yuan Yang

Lyrics generation
- Nursultan Kabylkas
- Ramesh Jayaraman

CNN model → “Man in black shirt playing a guitar.” → RNN model

Come on come on...
I see no changes wake up in the morning and I ask myself
Is life worth living should I blast myself?
Image captioning

test image

sample
<END> token
=> finish.
Image Feature Extraction Model

- Inception V3
- VGG-16
- Resnet-50
Inception V3
Dataset

• Flickr-8k (1.6GB)
• Flickr-30K (4.1GB)
• Microsoft COCO datasets (14GB)
• Imagenet datasets (56GB)
Inception V3 → Image Feature vector → Dense → repeat Image Feature matrix → Embedding matrix

One-Hot Encoder → LSTM → Dense → Caption Feature matrix

Concatenated Feature Matrix

<begin> man → Bidirectional LSTM → Dense → Activation → in
Inception V3
Image Feature vector
Dense
repeat
Image Feature matrix

Embedding matrix
One-Hot Encoder
LSTM
Dense
Caption Feature matrix

<begin> man in

Bidirectional LSTM
Dense
Activation

black

Concatenated Feature Matrix
Embedding matrix → One-Hot Encoder → LSTM → Dense → Caption Feature matrix → Concatenated Feature Matrix

Inception V3 → Image Feature vector → Dense → repeat

(begin) man in black shirt is playing

Bidirectional LSTM → Dense → Activation → Guitar
Inception V3 Image Feature vector Dense Image Feature matrix repeat

Embedding matrix One-Hot Encoder LSTM Dense Caption Feature matrix

<begin> man in black shirt is playing guitar

Bidirectional LSTM Dense Activation <end>

Concatenated Feature Matrix
Hyperparameters

- Optimizer: Adam
- Batch Size: 256
- Epochs: 100 (6000 images per epoch, 5 captions per image)
- Embedding size: 256
- Adam hyperparameters are from original paper
Results

• Loss: 2.016
• Training Accuracy: 0.501
Out[305]:

Normal Max search: A dog is jumping up at the camera.
Beam Search, k=3: A brown dog is excitedly greeting another dog.
Beam Search, k=5: A brown dog is excitedly greeting another dog.
Beam Search, k=7: A brown dog is excitedly greeting another dog.
Normal Max search: A woman with a necklace and a necklace poses for the camera.
Beam Search, k=3: A woman with a necklace poses for the camera.
Beam Search, k=5: A woman with a necklace poses for the camera.
Beam Search, k=7: A woman with a necklace poses for the camera.
Out[309]:

In [310]:

print ('Normal Max search:', predict_captions(try_image))
print ('Beam Search, k=3:', beam_search_predictions(try_image, beam_index=3))
print ('Beam Search, k=5:', beam_search_predictions(try_image, beam_index=5))
print ('Beam Search, k=7:', beam_search_predictions(try_image, beam_index=7))

Normal Max search: A man in a red vest and white vest is water-skiing.
Beam Search, k=3: A man is wakeboarding through the water.
Beam Search, k=5: A man is wakeboarding through the water.
Beam Search, k=7: A man is wakeboarding through the water.
Normal Max search: A man with a hat standing on a rock.
Beam Search, k=3: A man wearing a hat is standing on a rock wall.
Beam Search, k=5: A man wearing a hat is standing on a rock wall.
Beam Search, k=7: A man wearing a hat is standing next to a rock wall.
In [322]:

print ('Normal Max search:', predict_captions(try_image))
print ('Beam Search, k=3:', beam_search_predictions(try_image, beam_index=3))
print ('Beam Search, k=5:', beam_search_predictions(try_image, beam_index=5))
print ('Beam Search, k=7:', beam_search_predictions(try_image, beam_index=7))

Normal Max search: A child climbing on a rope ropes.
Beam Search, k=3: A little girl in a red shirt and green shirt climbs on a playground.
Beam Search, k=5: A child in a red shirt and red shirt climbs up ropes.
Beam Search, k=7: A child in a red shirt and green climbs on a playground.
Lyrics generation

- we implemented encoder-decoder model
- typically used for machine translation
Encoder-Decoder model
Encoder-Decoder model for lyrics gen.
Encoder-Decoder model for lyrics gen.

Example:

Line 1: Guess who’s back?
Line 2: Back again!
Can Machine Generate Traditional Chinese Poetry? A Feigenbaum Test

Qixin Wang¹,⁴*, Tianyi Luo¹,³*, Dong Wang¹,²†
¹CSLT, RIIT, Tsinghua University, China
²Tsinghua National Lab for Information Science and Technology, Beijing, China
³Huilan Limited, Beijing, China
⁴CIST, Beijing University of Posts and Telecommunications, China
{wqx, lty}@cs1t.riit.tsinghua.edu.cn
wangdong99@mails.tsinghua.edu.cn

Abstract

Recent progress in neural learning demonstrated that machines can do well in regularized tasks, e.g., the game of Go. However, artistic activities such as poem generation are still widely regarded as human

tonal patterns. The structural pattern regulates how many lines and how many characters per line; the rhythmical pattern requires that the last characters of certain lines hold the same or similar vowels; and the tonal pattern requires characters in particular positions hold particular tones, i.e., ‘Ping’ (level tone), or ‘Ze’ (downward tone). A
Architecture

Loss layer

Projection layer

Hidden layer

Embedding layer

Loss:
sparse_softmax_cross_entropy_with_logits
Optimizer: AdamOptimizer
learning rate = 0.001

GRU cell
state_size = 256

cocab_size = 2k - 7k
Data set

- Kaggle
- AZLyrics.com scraping
- Millionsong dataset

```
1 Tommy Still drives that ol' pickup truck
2 He had back in high school
3 He says he won't trade it in
4 'Cause they don't make 'em like they used to
5 His wife Donna's got a picture of Elvis
6 By the clock in the kitchen
7 Her hair and clothes come back in style
8 Ever now and then
9 They made a promise on their wedding day
10 Nothin's gonna take this feeling away
11 Let the world turn
12 Let the years drift by
13 The fire will burn
14 In their hearts just as bright
15 Time may take it's toll
16 But their love won't grow old
17 And if they live to be a hundred and one
18 They're still gonna die young
19 Tommy spent years fixin' wheels and brakes
20 at the texaco on the highway
21 Donna kept the kids and old Permanent waves
22 At the beauty shop on fridays
23 It's been forty years but who'd ever know
24 By the way they carry on
25 Just the same two kids as much in love
26 as the day they married on
27 They made a promise on there wedding day
28 Nothings gonna take this feeling away
29 Let the world turn
30 Let the years drift by
```
Generated lyrics

- Trained on 1 album of Eminem
- 1000 lines of lyrics
- 2500 vocab size

Seed: “I love you, dear”

You’re, i want you to focus
If you face me crazy,
Every time i think of you, puke
I was gonna take the time to sit down and write you to go,
But i thought thought a song would probably be a little better,
Instead of just a letter
Trick to improve performance

The results were worse

Seed: “I love you, dear”

you’re like an argument over a parking spot their front porch, their front porch but i ain’t tryna have none of my people hurt and murdered im supposed to be the soldier who never blows his composure
Trick to improve performance

- Take output of image captioning model and find the closely matched line in training set based on cosine difference.
## Trick to improve performance

<table>
<thead>
<tr>
<th>Seed</th>
<th>Closest line in training set</th>
</tr>
</thead>
<tbody>
<tr>
<td>I love you, dear</td>
<td>cause you love me, and i love you more</td>
</tr>
<tr>
<td>A woman with the necklace posing for the camera</td>
<td>got our name from a woman and our game from a woman</td>
</tr>
<tr>
<td>A man in red vest wakeboarding</td>
<td>already in the red</td>
</tr>
<tr>
<td>A child in a red shirt and green climbs on a playground</td>
<td>professor x vanglorious exists in a state of red, black, and green</td>
</tr>
</tbody>
</table>
Results

Seed:
“A woman with the necklace posing for the camera”
Closest seed from training set:
“got our name from a woman and our game from a woman”

Generated lyrics:

she wrapped the baby up and threw him in a trash heap
she didn’t know what to throw away and what to keep
she sees sex as a way of leavin’ hell
she wrapped the baby up and threw him in a trash heap
guess she thought she’d get away, wouldn’t hear the cries
she tried to sell crack but end up gettin’ robbed
now brenda really never knew her moms
Results

Seed:
“A little child climbs...”
Closest seed from training set:
“now hush little baby, don’t you cry”

Generated lyrics:
everythings everything’s gonna be alright
that upper lip up, little lady, i told ya
daddy’s here to say ya through the night
i know mommys not here, daddys here
but i promise mama’s being alright
...
Conclusion

● Bigger dataset will definitely increase performance
● Implement attention models
● Tune