Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS COC

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

Discussio

Question

References

References

Generating Natural Questions About an Image

[5]

Ryan Callihan & Sarah Taylor

Seminar für Sprachwissenschaft Universität Tübingen

Janurary 12, 2018

Sarah Taylor

Generating Questions fro Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

Bing

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

it iteatest iteigi

ъ.

Question

References

References

Outline

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Datasets

MS COCO

Bing

Flicker

Generative Models

Maximum Entropy Language Model

Long Short-Term Memory

Gated Recurrent Neural Network

Retrieval Model

K-Nearest Neighbor

Evaluation

Discussion

Questions

References

Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Elista

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

Lvaluatioi

. .

References

Poforoncos

Getting Started

Make groups!

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flicke

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neigh

Evaluation

Lvaiuatioi

. ..

References

References

Challenge One - Setup

Please go to: and enter in the code:

Ryan Callihan & Sarah Taylor

Generating Questions f

The Authors' Objective

Image Recognition

Datasets

MS CO

Flicke

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neighbo

Evaluation

Discussion

Deference

Reference

References

Challenge One



Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognitionusing a CNN

Dataset

Bing

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neigh

Evaluation

Ouestions

References

References

Challenge One - Corpus Results



- ► Is this a religious ceremony?
- ► That looks very interesting, don't you think?
- ► What are they all gathered for?
- ▶ What are these people gathered for?
- ▶ Is this a satanic ritual?

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flick

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

Lvaiuatioi

. ..

References

References

Challenge Two - Setup

Please go to: and enter in the code:

Ryan Callihan & Sarah Taylor

Generating Questions fro

The Authors' Objective

mage Recognition

Dataset

MS CO

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neighbo

Evaluation

Outsi

Deference

Deference

Challenge Two



Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognitionusing a CNN

Dataset

Bing Elister

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model K-Nearest Neighbor

Evaluation

Questions

References

References

Challenge Two - Corpus Results



- ▶ What are the people demonstrating about?
- ▶ What rally are they attending?
- ► What are these people protesting?
- ▶ What are they protesting?
- ▶ Who is in the gray jacket?

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognition

Dataset

MS CO

5...8

Generative Models

Maximum Entropy

Long Short-Term Memory

Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

D1.......

Questio

Reference

5.6

References 10

Generating Natural

What is Image Recognition?

Ryan Callihan &

Evaluation

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

.

using a CNN

Dataset

MS COC

Elicko

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbor

Evaluation

Discussion

Question

References

References

Image recognition and neural networks

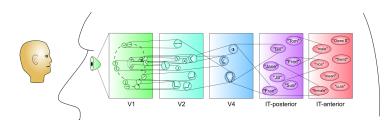


Figure: Visual network representation. Image from grey.colorado.edu [6]

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

Image Recognition

Dataset

MS COC

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

K-Nearest Neigh

Evaluation

Discussion

Question

References

References

Convolutional neural network



Figure: CNN representation. Image from blog.floydhub.com/building-your-first-convnet [8]

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

lmage Recognitio using a CNN

Dataset

MS COC

Flicke

Generative Mode

Maximum Entropy Language Model Long Short-Term Memor Gated Recurrent Neural Network

Retrieval Mode

Evaluation

a. .

Question

References

References

CNN visualized

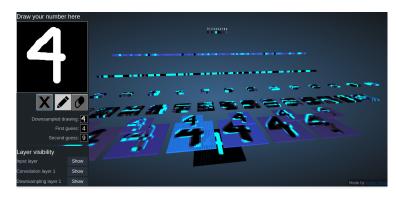


Figure: CNN visualized. Image from scs.ryerson.ca/ [4]

Ryan Callihan & Sarah Taylor

Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO Bing

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Question

References

References

Generative Models



Caption Bot [2]

captionbot.ai was used throughout this paper to automatically generate captions. It is a Microsoft project based on the Computer Vision API, Emotion API, and Bing Image Search API.

Generating Natural

Ryan Callihan &

References

Caption Bot example



I think it's a person smiling for the camera and she seems 🖨



Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures
The Authors' Objects

Image Recognition

Dataset

Bing

Flicker

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode K-Nearest Neighbor

Evaluatio

Discussion

Questions

References

References

Computer Vision API example



FEATURE NAME:	VALUE
Description	{ "tags"; { "person", "man", "indoor", "table", "sitting", "holding", "food", "woman", "glasses", "people", "posing", "drinking", "wine", "restaurnt", "plate", "smillag", "pizza", "phone", "young", "standing", "store", "group", "white" }, "captions" [{ "text": "a man sitting at a table in a restaurant", "confidence" 0.9105153 }] }
Tags	[{"name": "person", "confidence": 0.999498367), {"name": "man", "confidence": 0.926230047), {"name": "indoor", "confidence": 0.8648256), {"name": "restaurant", "confidence": 0.193121776 }]
lmage format	"Jpeg"

- ► Description "tags": ["person", "man", "indoor", "table",
 "sitting", "holding", "food", "woman", "glasses", "people",
 "posing", "drinking", "wine", "restaurant", "plate", "smiling",
 "pizza", "phone", "young", "standing", "store", "group", "white"
], "captions": ["text": "a man sitting at a table in a
 restaurant", "confidence": 0.9105153]
- ► Faces ["age": 25, "gender": "Male", "faceRectangle": "top": 94, "left": 149, "width": 79, "height": 79, "age": 33, "gender": "Male", "faceRectangle": "top": 97, "left": 343, "width": 79, "height": 79]

Ryan Callihan & Sarah Taylor

Generating
Questions from

The Authors' Objective

Image Recognitio

Datacete

MS CO

Ding

Generative Model

M. F.

Language Model

Gated Recurrent Neural

Retrieval Model

K-Nearest Neighl

Evaluation

D:-----

. .

Referenc

References 18/37

Ryan Callihan & Sarah Taylor

Generating

The Authors' Objective

Image Recognition

Dataset

MS CO

D....8

Generative Models

Maximum Entrony

Long Short-Term Memory

Cated Recurrent Neural

Retrieval Mode

K-Nearest Neight

Evaluation

Discussio

Questio

Reference

References 19/37

Ryan Callihan & Sarah Taylor

Generating Questions from

The Authors' Objective

Image Recognitio

_ . . .

.

Bing

Generative Models

....

Language Model

Long Short-Term Memory

Retrieval Model

K-Nearest Neighb

Evaluation

D.

Questio

Reference

recreme

References 20/37

Ryan Callihan & Sarah Taylor

Generating Questions fron

The Authors' Objective

Image Recogniti

_

NC CO

...5 00

Eliek

Generative Models

Maximum Entropy

Long Short-Term Memory

Gated Recurrent Neural

Retrieval Model

K-Nearest Neighbo

Evaluation

D.

Quartic

Reference

receive

References

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition

ъ. .

... ...

Bing

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

K-Nearest Neighb

Evaluation

Discussio

Questions

References

References

Generative models

Three generative models were used in this study

- Maximum entropy language model (MELM)
- ► RNN with long short-term memory cells (LSTM)
- ► RNN with gated recurrent units (GRU)

Ryan Callihan & Sarah Taylor

Questions from Pictures

The Authors' Objectiv

mage Recognition

Dataset

Bing

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Elizabeth de

Evaluation

Question

References

References

Model input



The last fully connected layer of the image recognition CNN was used as input

It was a 4096-dimentional vector

Ryan Callihan &

Sarah Taylor

Generating Questions from Pictures

The Authors' Objecti

Image Recognition using a CNN

Б. .

MS COC

Citation.

Generative Model

Maximum Entrop

Long Short-Term Memor Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. ..

References

References

Maximum Entropy Language Model

► Generated word probabilities using a CNN

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS COC

Dillig

Generative Models

Maximum Entrop

Long Short-Term Memor Gated Recurrent Neural

Retrieval Mode

K Noorget Noighbor

Evaluation

Question

References

References

Maximum Entropy Language Model

- ► Generated word probabilities using a CNN
- ► MELM using candidate word probabilities

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objecti

Image Recognitio using a CNN

Dataset

MS COO

Dillig

Flicke

Generative Model

Maximum Entrop

Long Short-Term Memory
Gated Recurrent Neural

Retrieval Model

Elizabeth de

Evaluation

Discussion

. .

References

References

Maximum Entropy Language Model

- ► Generated word probabilities using a CNN
- ► MELM using candidate word probabilities
- ► Relevant questions but some grammatical errors.

 Warrents futher research

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognitio

Datasets

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigl

Evaluation

Lvaiuatioi

. ..

References

References

RNN with Long Short-Term Memory

► Given a caption, model generates a question

25/37

Ryan Callihan &

Generating Questions from Pictures

The Authors' Objective

Image Recognitio

Datasets

Bing

Flicker

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbor

Evaluation

Questio

References

References

RNN with Long Short-Term Memory

- ► Given a caption, model generates a question
- ► Used gold standard caption-question dataset for training

Questions About an Image

Ryan Callihan &

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Lvaluatioi

Discussio

References

References

RNN with Long Short-Term Memory

- ► Given a caption, model generates a question
- Used gold standard caption-question dataset for training
 - ► Consists of 2 RNNs
 - ► An encoder which processes the caption
 - ► A decoder which generates the question

Retrieval Model

Evaluation

References

References

RNN with Long Short-Term Memory

- ▶ Given a caption, model generates a question
- ► Used gold standard caption-question dataset for training
 - ► Consists of 2 RNNs
 - ► An encoder which processes the caption
 - ► A decoder which generates the question
 - Output was incoherent

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

mage Recognition

using a CNN

Dataset

Bing

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Discussio

Question

References

References

Gated Recurrent Neural Network

 Uses final fully connected layer from the image recognition CNN (Captionbot I assume) into a 500-dimentional vector

an Image

Rvan Callihan &

Retrieval Model

References

References

Gated Recurrent Neural Network

- ▶ Uses final fully connected layer from the image recognition CNN (Captionbot I assume) into a 500-dimentional vector
- Generates a single token at a time until the end of sentence token (aka "?")

Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures
The Authors' Objective

Image Recognitio using a CNN

Dataset

MS COO Bing Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model K-Nearest Neighbor

Evaluation

Question

References

References

Gated Recurrent Neural Network

- Uses final fully connected layer from the image recognition CNN (Captionbot I assume) into a 500-dimentional vector
- ► Generates a single token at a time until the end of sentence token (aka "?")
- ➤ "Where is this?" Is the top generated question for 29.3% of the images. More interesting and meaningful quesions generally ranked lower

Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures
The Authors' Objective

Image Recognitio using a CNN

Dataset

MS COO Bing Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model K-Nearest Neighbor

Evaluation

Question

References

References

Gated Recurrent Neural Network

- Uses final fully connected layer from the image recognition CNN (Captionbot I assume) into a 500-dimentional vector
- ► Generates a single token at a time until the end of sentence token (aka "?")
- ➤ "Where is this?" Is the top generated question for 29.3% of the images. More interesting and meaningful quesions generally ranked lower

Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures
The Authors' Objective

Image Recognition

Dataset

MS COC

Bing Flicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model K-Nearest Neighbor

Evaluation

Ouestion

References

References

Gated Recurrent Neural Network

- Uses final fully connected layer from the image recognition CNN (Captionbot I assume) into a 500-dimentional vector
- ► Generates a single token at a time until the end of sentence token (aka "?")
- ► "Where is this?" Is the top generated question for 29.3% of the images. More interesting and meaningful quesions generally ranked lower
- ► They added a constraint that which rejects questions with less that 6 tokens

Retrieval Model

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

using a CNN

Dataset

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memor Gated Recurrent Neural Network

Retrieval Model

K-Nearest Neigl

Evaluation

Evaluatio

D .500a.

Б.

Reference

Reference

27/37

K-Nearest Neighbor

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K Magraet Nais

Evaluation

Evaluation

D.500.

- -

Reference

28/37

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition using a CNN

Datasas

MS COO

Elicker

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mod

IZ Manage Natio

Evaluation

Lvaluation

. .

References

References

Evaluation

Two methods of evaluation were used

► Human evaluation using AMT on a 3 point scale

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS COC

Flicker

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neighbo

Evaluation

D:-----

Questio

References

References

Evaluation

Two methods of evaluation were used

- ► Human evaluation using AMT on a 3 point scale
 - ► Automatic evaluation
 - ► BLEU [7]
 - ▶ ΔBLEU [3]
 - ► METEOR [1]

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objecti

Image Recognition using a CNN

Datacete

Bing

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

Lvaluatio

. . .

References

References

Human Evaluation

Using AMT, 3 people evaluated each question on a 3 point scale from 1 (worst) - 3 (best)

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Datasets

MS CO Bing

Flicker

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbor

Evaluation

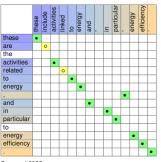
Discus

References

_ .

BLEU, $\triangle BLEU$, & METEOR

Similar concept to Levenshtein or Needleman-Wunsch Distance



Segment 2022

P: 0.897 R: 0.907 Frag: 0.514 Score: 0.440

Figure: Taken from cs.cmu.edu

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition using a CNN

. . . .

MS COO

Bing

Generative Mode

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K-Nearest Neighbor

Evaluation

Discus

2.2

References

Reference

Correlation between human and automatic metrics

ρ 0.628 (1.5e-08) 0.67 (7.0e-10) 0.702 (5.0e-11		METEOR	BLEU	$\Delta BLEU$
, , , , , , , , , , , , , , , , , , , ,	r	0.916 (4.8e-27)	0.915 (4.6e-27)	0.915 (5.8e-27)
	ρ	0.628 (1.5e-08)	0.67 (7.0e-10)	0.702 (5.0e-11)
au 0.476 (1.6e-08) 0.51 (7.9e-10) 0.557 (3.5e-11)	au	0.476 (1.6e-08)	0.51 (7.9e-10)	0.557 (3.5e-11)

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flicker

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mode

K Noorget Noighbor

Evaluation

Lvaiuatioi

. . . .

References

References

Evaluation discussion

With your group, take a look at the evaluation results table and:

▶ Decide which model and data set worked the best and which worked the worst?

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objectiv

Image Recognition using a CNN

Dataset

MS CO

Flicker

Generative Models

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Mod

K-Nearest Neighbor

Evaluation

Question

References

References

Evaluation discussion

With your group, take a look at the evaluation results table and:

- ► Decide which model and data set worked the best and which worked the worst?
- ▶ Why do these differences exist?

Discussion - Authors' Thoughts

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objective

using a CNN

Dataset

MS CO

Eliebe

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural

Retrieval Mode

K-Nearest Neigh

Evaluation

. .

References

Reference

34/37

Questions

Ryan Callihan & Sarah Taylor

Generating
Questions from
Pictures

The Authors' Objective

Image Recognition using a CNN

Dataset

MS CO

Flicke

Generative Model

Maximum Entropy Language Model Long Short-Term Memory

Retrieval Mode

K-Nearest Neigh

Evaluation

Evaluation

Disci

Reference

References

35/37

an Image

Rvan Callihan &

References

References I

- [1] Denkowski, M. and Lavie, A. (2014). Meteor universal: Language specific translation evaluation for any target language. In Proceedings of the EACL 2014 Workshop on Statistical Machine Translation.
- [2] Fang, H., Gupta, S., Iandola, F., Srivastava, R. K., Deng, L., Dollár, P., Gao, J., He, X., Mitchell, M., Platt, J. C., et al. (2015). From captions to visual concepts and back. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1473-1482.
- [3] Galley, M., Brockett, C., Sordoni, A., Ji, Y., Auli, M., Quirk, C., Mitchell, M., Gao, J., and Dolan, B. (2015). deltableu: A discriminative metric for generation tasks with intrinsically diverse targets. CoRR, abs/1506.06863.
- [4] Harley, A. W. (2015). An interactive node-link visualization of convolutional neural networks. In ISVC, pages 867-877.
- [5] Mostafazadeh, N., Misra, I., Devlin, J., Mitchell, M., He, X., and Vanderwende, L. (2016). Generating natural questions about an image. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1802–1813, Berlin, Germany. Association for Computational Linguistics.

Ryan Callihan & Sarah Taylor

Generating Questions from Pictures

The Authors' Objecti

Image Recognitio using a CNN

Dittion

MS COC

Bing

Generative Model

Maximum Entropy Language Model Long Short-Term Memory Gated Recurrent Neural Network

Retrieval Model

Evaluation

Discussion

.

References

References II

- [6] O'Reilly, R. C., Munakata, Y., Frank, M., Hazy, T., et al. (2012). Computational cognitive neuroscience. PediaPress.
- [7] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics, pages 311–318. Association for Computational Linguistics.
- [8] Wadhwa, S. (2017). Building your first convnet.

Reference