Multimodal Image inspired hashtag generator

Varun Gupta	2018201003
Vatsal Soni	2018201005
Darshan Kansagara	2018201033
Dhawal Jain	2018201065

Introduction

- The problem statement is to generate the hashtags for a given multi-modal instagram post that could be a text or an image or set of images. The generated hashtags should be relevant to the existing hashtags that may present along with the post.

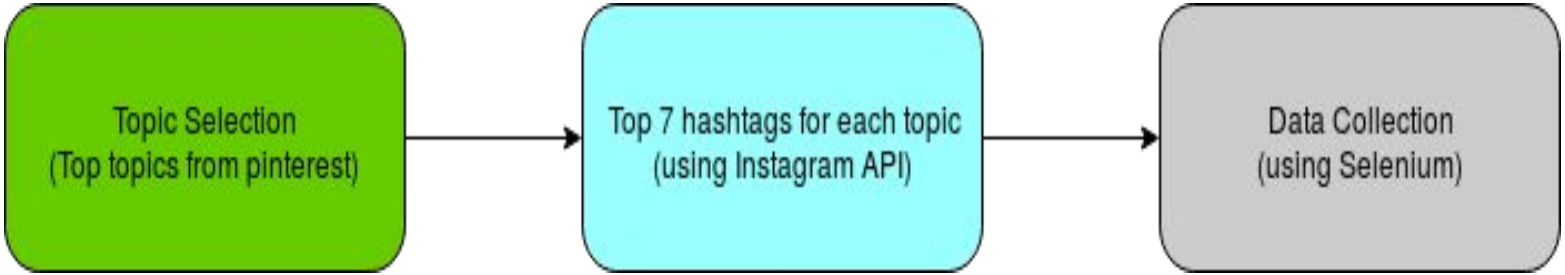
Major Components of Projects

- **Identifying prior art :** Some amount of work has already been done on utilizing images and multi-modal content for hashtag generation. [1] and [2] utilize CNN and attention to make predictions for hashtags. [3] on the other hand uses zero-short learning for the same. Understanding these approaches and incorporating their ideas in our project is the first task.
- **Data Collection:** The data collection process would consist of
    - Generating or creating a list of seed hashtags
    - Segmenting it to generate more hashtags (Using Viterbi algorithm)
    - Use this larger set to scrape more hashtags from Instagram. This process will be repeated until there is enough data for training and testing.
- **Modeling:** The problem would most likely require us to utilize deep learning methods like CNNs and LSTMs. The task is slightly similar to image captioning tasks that mostly utilize pretrained Imagenet models and an LSTM for generating the captions. We would also be using methods similar to the ones presented in prior works that combine visual (CNN) and textual (LSTM and Word2Vec) information via attention modules.

Milestones

- Collect a large dataset of Instagram posts and hashtags for a specific domain. Create a data collection pipeline that can be utilized for collecting data for other topics as well given different seed list of hashtags.[Till end of Jan]
- Implement few different models (Based on prior art) for hashtag generation using multimodal data. Evaluate them on our dataset and compare with simple baselines like image only / text only classifiers or non machine learning methods like synonym finder.[First week of march]
- Implement different models for image classification, evaluation of the models, results and analysis. [Till End may now]

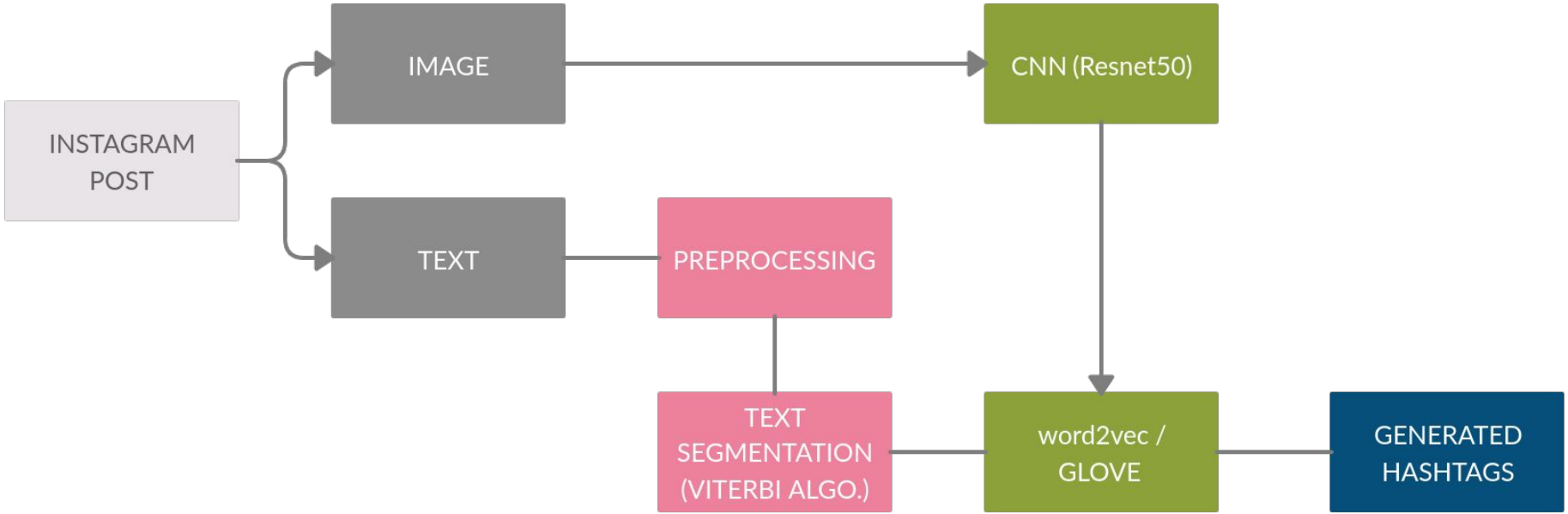
Building Dataset



Continue...

- Our dataset consists of hashtags collected from a diverse set of 8 topics mentioned below:
    - Pets , Art
    - Jewellery, Food
    - Travel, Architecture
    - Babies, Nature
- We have figure out 5 most popular hashtag for above topics and Wrote script to scrap instagram posts for given a hashtag.
    - Ex : For Food : #foodaddict, #foodinstagram, #foodstyling, #foodblog, #foodphotography

High level Architecture



Text Model

Text Preprocessing

Hashtag Segmentation

Text Classifier : We have trained a word embedding approach like **Word2Vec** and **Glove embeddings** on our corpus and used distance measures such as cosine similarity to get the most relevant hashtags. As one of the baselines, we also aim to implement a simple topic modelling based approach for the task of hashtag generation. The approach is detailed below.

Typically a post consists of the following:

- A photo
- A caption of 2-3 lines
- A bunch of hashtags

Continue ...

1. Topic based Glove Model:- We have used glove embedding to find similarity between words. Here for each topic, we have one model that is trained on its corresponding topic's corpus.

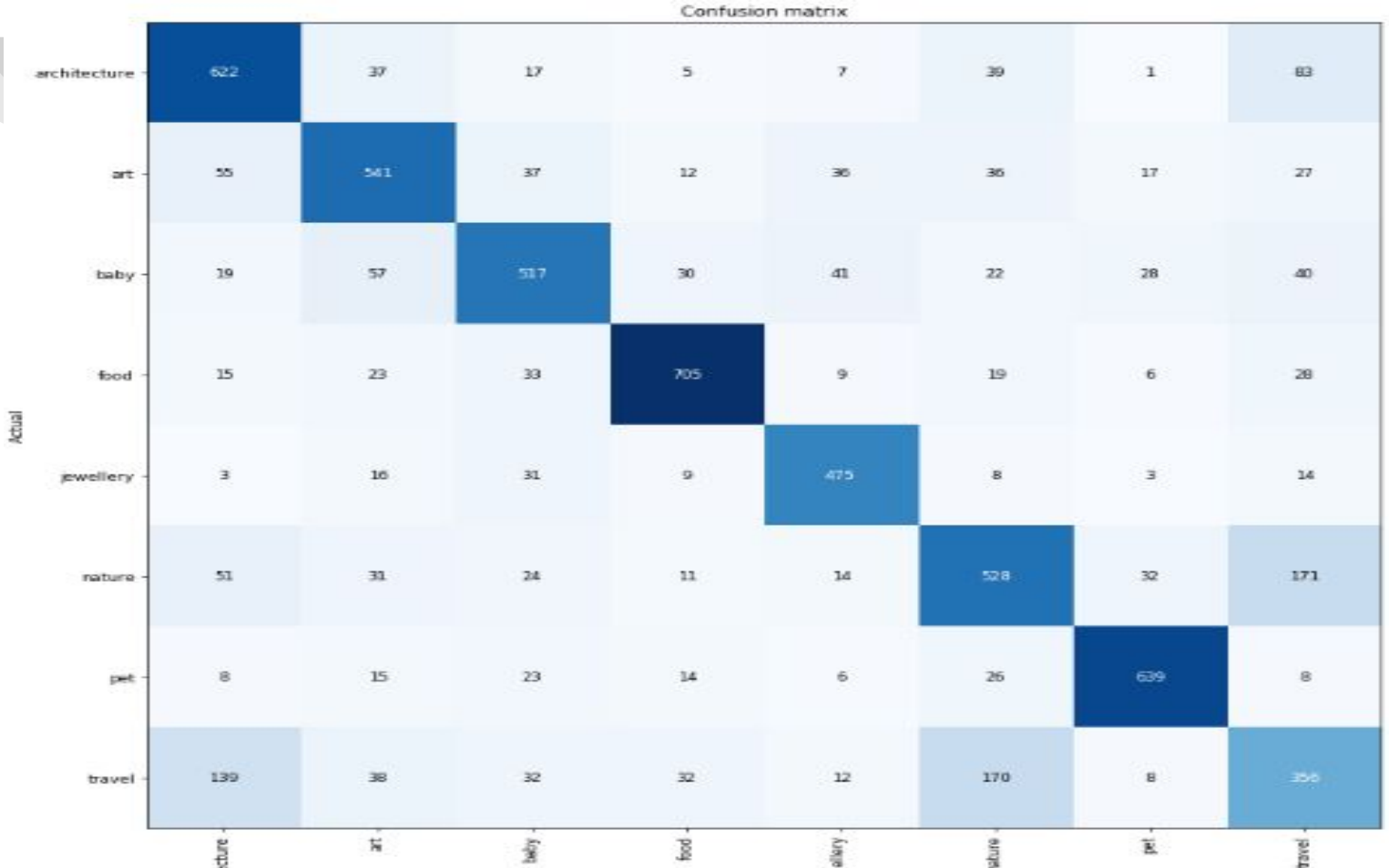
2. Topic based Word2vec Model:- We have used word2vec embedding to find similarity between words. Here for each topic we have one model that is trained on corresponding topic's corpus plus pre trained wikipedia glove embedding.

3. Global word embeddings:- We have trained one global model irrespective of topic that is trained on complete corpus plus pre trained wikipedia glove embedding.

Image Classifier

We train an image classifier such as Resnet-50 and Resnet-34 for the task of image classification. We model our problem as a multi-label classification problem and leverage transfer learning to finetune the model on our dataset. We train all our models using PyTorch and utilize the negative cross entropy loss for training. The optimizer used for training is the Adam optimizer with an initial learning rate of 0.01. We also leverage certain other approaches such as early stopping and cosine annealing along with data augmentations to improve the accuracy of the model. Our task is slightly ill posed as the labels provided for training is not annotated by humans. Rather, the labels are based on the topics provided by the user for an Instagram posts and are bound to have several mis-classifications.

Model	Accuracy
resnet-34	71.2 %
resnet-50	72.8 %


Confusion matrix
Actual
architecture
art
baby
food
jewellery
nature
pet
travel
622 37 17 5 7 39 1 83
55 541 37 12 36 36 17 27
19 57 517 30 41 22 28 40
15 23 33 705 9 19 6 28
3 16 31 9 475 8 3 14
51 31 24 11 14 528 32 171
8 15 23 14 6 26 639 8
139 38 32 32 12 170 8 356

Example

Predicted Hashtags: #view #mountain #world #amazing #ig

Evaluation and Results

```
like-------------follow
fun--------------enjoy
life-------------life
smile------------thing
friends----------thing
cute-------------cool
love-------------good
holidays---------good
view-------------view
happiness--------daili
```

Predicted	life	good	thing	follow	cool	enjoy	view	daili
like	0.3	0.56	0.38	0.6	0.33	0.33	0.29	0.41
fun	0.29	0.47	0.33	0.04	0.36	0.48	0.11	0.14
life	1.0	0.4	0.12	0.26	0.07	0.27	0.25	0.43
smile	0.09	0.21	0.33	0.13	0.31	0.23	0.06	0.1
friends	0.16	0.28	0.36	0.32	0.16	0.21	0.11	0.13
cute	0.03	0.18	0.36	0.02	0.39	0.18	-0.0	0.01
love	0.54	0.55	0.16	0.37	0.2	0.33	0.24	0.4
holidays	0.24	0.38	0.12	0.2	0.18	0.33	0.24	0.19
view	0.25	0.3	0.13	0.26	0.13	0.31	1.0	0.22
happiness	0.3	0.27	0.14	0.2	0.12	0.29	0.13	0.35

```
0.9
```

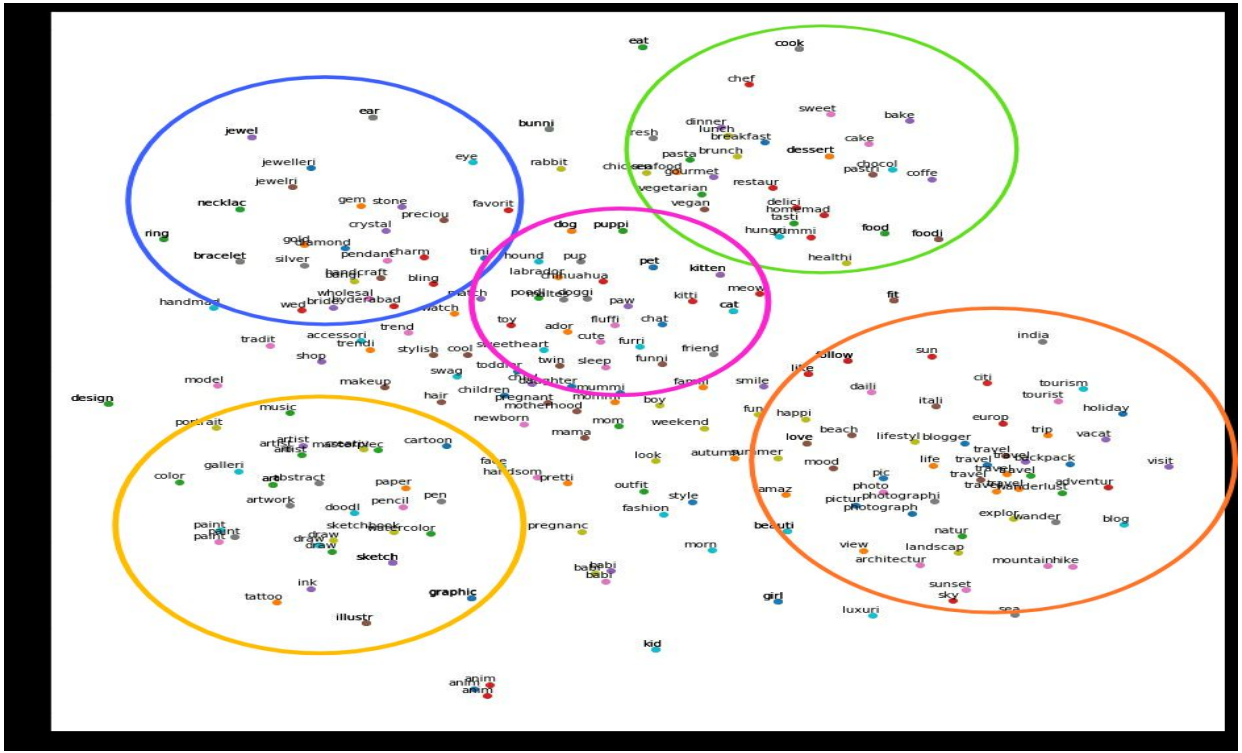
Continue...

Model	Accuracy
Top K Baseline	41.829%
word2vec + wikipedia glove embeddings	87.558%
Global model (word2vec)	63.040%
Global model (glove)	74.18%
Our approach (glove with topics)	96.172%
Oracle	96.525%

Table 3: Results for our proposed approach

Description	Accuracy
Our approach without segmentation	65.34%
Our approach with segmentation	96.172%

Analysis


eat
cook
chef
sweet
bake
dinner
lunch
breakfast
cake
dessert
brunch
pasta
chocol
coffe
vegetarian
restaur
vegan
delici
homemad
tasti
food
foodi
healthi
ear
jewel
bunni
jewelleri
jewelri
eye
rabbit
necklac
gem
stone
favorit
precious
crystal
ring
gold
diamond
bracelet
silver
pendant
charm
tini
handcraft
bling
handmad
wholesal
bridesmaid
hyderabad
watch
dog
puppi
hound
pup
labrador
chihuahua
pet
kitten
meow
kitti
paw
cat
fit
toy
ador
fluffi
chat
trend
tradit
accessori
trendi
stylish
cool
sweetheart
cute
furri
friend
shop
twin
sleep
funni
swag
toddler
daughter
mummi
family
smile
like
follow
sun
india
model
makeup
design
hair
children
pregnant
mommi
boy
daili
citi
tourism
tourist
holiday
music
portrait
mother
newborn
mom
weekend
fun
happi
italian
europ
mama
beach
lifestyl
blogger
travel
trip
vacat
artist
cartoon
look
autumn
summer
love
mood
life
visit
galleri
color
abstract
art
paper
pretti
outfit
style
amaz
pictur
photographi
photograph
photo
pic
adventur
artwork
pencil
pen
doodl
fashion
explor
wander
blog
paint
sketchbook
watercolor
pregnanc
beauti
natur
draw
sketch
morn
view
landscap
architectur
mountainhike
baby
babi
ink
tattoo
graphic
illustr
girl
sunset
sky
luxuri
sea
kid
anim

References

1. Cesc Chunseong Park, Byeongchang Kim, and Gunhee Kim. Towards personalized image captioning via multimodal memory networks. IEEE transactions on pattern analysis and machine intelligence, 41(4):999–1012, 2018.

2. Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. Hashtag recommendation for multimodal microblog using co-attention network. In IJCAI, pages 3420–3426,2017.

3. Shivam Gaur. Generation of a short narrative caption for an image using the suggested hashtag.In 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), pages 331–337. IEEE, 2019.

4. Yue Wang, Jing Li, Irwin King, Michael R Lyu, and Shuming Shi. Microblog hashtag generation via encoding conversation contexts. arXiv preprint arXiv:1905.07584, 2019.

5. Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. Hashtag recommendation for multimodal microblog using co-attention network. In IJCAI, pages 3420–3426, 2017.

Thank You