Image Hashtag Generation using Deep Learning

Aditya Verma[1] , Anukrity Varshney[2] , Devesh Mittal[3]
Delhi Technological University
Delhi , India

Abstract **-** Reading an image is easy to remember by the human mind. Human nature always has a curiosity to develop an intelligent model which does almost all works or activities performed by them. A model is said to be artificially intelligent when some algorithms applied onto works efficiently and effectively. Image Captioning is also an artificial intelligence topic in which use of NLP and CNN makes it automated and can create picture definition with proper prediction. A need for neural networks and analysis of data may create this work a better approach. Model architecture may also be helpful in image processing and caption prediction.
Automatically describing the content of an image is a fundamental problem in artificial intelligence that connects computer vision and natural language processing.
The purpose of this research is to develop a safe and efficient hashtag generating solution for social media users which generates relevant hashtags for user image content in order to get a broad reach of the target audience.

Keywords— Hashtags; social media; NLP; machine learning;deep learning; CNN

I. INTRODUCTION

A hashtag is a name or an identifier that resolves to a description of its referent. In other words, a hashtag is a keyword or phrase preceded by the hash symbol, written within a post or comment to highlight it and facilitate a search for it. In the present day, hashtags are immensely used for brand promotion and social media discussions. In principle hashtags facilitate powerful identification functionality to any kind of HTTP based services .
Image Hashtag Generator refers to the process to generate hashtags related to the image based on actions or objects shown in the image. The project uses Deep Neural Networks and Natural Language Processing.

ResNet50 architecture and sequential model with optimisers such as ‘adam’ and ‘RMSProp’ have been used in the project along with Flickr 8K dataset to train and test the algorithm . Flickr8k dataset is a collection for sentence-based image description and search, consisting of 8,000 images that are each paired with five different captions which provide clear descriptions of the salient entities and events. NLP refers to natural language processing and is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language . Stopwords have been used to remove the frequently occurring words in every sentence such as ‘the’ , ‘in ‘, etc. and are filtered out before or after processing of natural language data (text). The unique keywords from the caption that best describes the image is used to create the hashtag .

II. LITERATURE REVIEW

This paper presents how convolutional neural network based architectures can be used to create hashtags based on the contents of an image. Deep convolutional neural networks based machine learning solutions are nowadays dominating for such image annotation problems . Recent research has proposed solutions that automatically generate human-like descriptions of any image. This problem is of significance in practical applications and moreover it links two artificial intelligence areas i.e. NLP (Natural Language Processing) ,CNN and Computer Vision.

Researchers have studied and developed multiple NLP related solutions over time to solve problems in various domains. They can be categorized as follows.

● Lexical and morphological analysis, noun phrase

generation, word segmentation, etc.

● Semantic and discourse analysis, word meaning and knowledge representation.

● Knowledge-based approaches and tools for NLP.

Noun phrasing is considered to be an important NLP

technique used in information retrieval. One of the major goals of noun phrasing research is to investigate the possibility of combining traditional keyword and syntactic approaches with semantic approaches to text processing in order to improve the quality of information retrieval.

The growing technology of NLP suggests that there are two possible scenarios for the future interactions between computers and humans: in the user-friendliness scenario, computers become smart enough to communicate in natural language, and in the computer friendliness scenario humans adapt their practices in order to communicate with, and make use of, computers.

Topic modeling is a type of statistical modeling for

discovering topics that occur in a collection of documents. Latent Dirichlet Allocation (LDA) is an example of topic model and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.

Recently, a number of studies have shown that the use of deep learning and text mining methods to automatically identify relevant studies has the potential to drastically decrease the workload.

Topic analysis is currently gaining popularity in both deep learning and text mining applications.

Since the emergence of topic models, researchers have introduced this approach into the fields of biological and medical document mining. Such experiments proved LDA could be successfully applied to text classification. In the present day, LDA modeling is being developed for machine based communication purposes.

The trending hashtag recommendation problem addresses suggesting hashtags to explicitly tag a post made on a given social media platform, based upon the content and the context of the post. The issue of trending hashtag recommendation has emerged as a mainstream area of research overtime.

All these research efforts are specifically targeted at microblogs which is a highly specific area of content.

In present most of the research work utilize the

advancements of ML to achieve their objectives. “User Conditional Hashtag prediction for Images” by E. Denton, et al. [12] is an approach that used ML along with the hashtags & contextual information about the user to perform hashtag prediction for user given image. Simply how user meta-data combined with images derived from a CNN can be used to

predict hashtags. With the data, the researchers developed a user model which could be applied for a large dataset that is taken from Facebook. The user model primarily predicted hashtags, but the predicted hashtags were not “trending hashtags”. In this approach a hashtag embedding model will be used that trains with the collected data. This method would be very practical because of the availability of data.

Data extraction from social media platforms comes under the categorization of social media data mining. Mainly there are three different ways to harvest data from social media platforms. Those are through APIs, personal archives and scraping. Since most of the social media platforms have updated their restrictions on data extraction due to various privacy related reasons, personal archiving method is not practical.

Therefore, the preferred method for this proposed solution is scraping.

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability of a given sample belonging to a particular class. Bayesian classifier is based on Bayes‟ theorem. Naive Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computation involved and, in this sense, is considered “naive”. Naïve Bayes method recommends hashtags by observing the content produced by the target user. In this paper it proposed to use Bayes model to estimate the probabilities of using different hashtags. Using this method, hashtags which are used by posts that has similar content can be identified.

The Natural Language Toolkit is a suite of program

modules, data sets, tutorials and exercises, covering symbolic and statistical natural language processing. NLTK is written in Python and distributed under the GPL open source license. Over the past three years, NLTK has become popular in teaching and research In the proposed research application, Naïve Bayes classifier allows to classify the generated hashtags based on the analysis that given by the specific algorithms that trained using large training data-sets.

III. ABSTRACT

A recent study on Deep Learning shows that it is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. Deep Learning (DL) and Neural Network (NN) is currently driving some of the most ingenious inventions in todays century. Their incredible ability to learn from data and environment makes them the first choice of machine learning scientists. Deep Learning and Neural Network lies in the heart of products such as self-driving cars, image recognition software, recommender systems etc. Evidently, being a powerful algorithm, it is highly adaptive to various data types as well.

Image annotation is a process by which a computer system assigns metadata in the form of captioning or keywords to a digital image. It is a Type of multi-class image classification with a very large number of classes. It is used in image retrieval systems to organize and locate images of interest from the database. The goal of image captioning research is to annotate and caption an image which describes the image using a sentence. To train a network to accurately describe an input image by outputting a natural language sentence. The task of describing any image sits on a continuum of difficulty. Some images, such as a picture of a dog, an empty beach, or a bowl

of fruit, may be on the easier end of the spectrum. While describing images of complex scenes which require specific contextual understanding and to do this well, not just possibly proves to be a much greater captioning challenge. Providing contextual information to networks has been both a sticking point, and a clear goal for researchers to strive for.

Our model to caption images are built on multimodal recurrent and convolutional neural networks. A Convolutional Neural Network is used to extract the features from an image which is then along with the captions is fed into an Recurrent Neural Network. The architecture of the image captioning model is shown in figure 1.

Image captioning is interesting because it concerns what we understand and about perception with

respect to machines. The problem setting requires both an understanding of what features (or pixel context) represent which objects, and the creation of a semantic construction grounded to those objects.

IV. SPECIFICATIONS

Libraries used:

· *Pandas*

pandas is a software library written for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and time series.

· *Numpy*

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

· *Matplotlib*

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

· *Keras*

Keras allows users to productize deep models and allows use of distributed training of deep-learning models on clusters of Graphics processing units (GPU) and tensor processing units (TPU).

· *Tensorflow*

TensorFlow is a Python library for fast numerical computing created. It is a foundation library that can be used to create Deep Learning models directly or by using wrapper libraries that simplify the process built on top of TensorFlow.

· *NLTK*

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning

· *DATASET- Flickr 8K*

For training our model we are using Flickr8K dataset. It consists of 8000 unique images and each image will be mapped to five different sentences which will describe the image.

V. COMPONENTS

· *CNN*

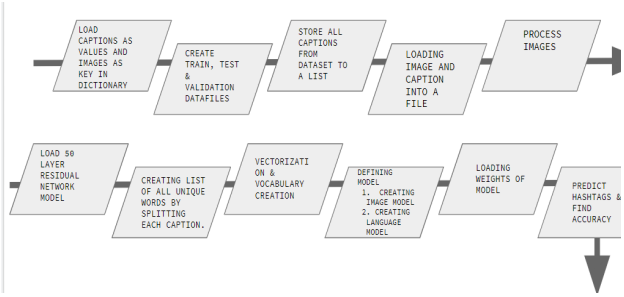
A Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

· *NLP*

Natural language processing is a subfield of linguistics, computer science, and artificial intelligence concerned with interactions between computers and human language. Using image processing by convolution neural networks we can process the image hashtags and represent them accordingly to the image pixels. In other words we can clearly define a natural language such as english in which humans talk to each other can be adjusted accordingly by convolution neural networks.

VI. PROPOSED ALGORITHM AND RESEARCH WORK



Creating testing dataset:

A good dataset that we used for image captioning is the Flickr8K dataset.

The reason is that it is realistic and relatively small so that you can download it and build models on your workstation using a CPU.

Within a short time, you will receive an email that contains links to two files:

· Flickr8k_Dataset.zip (1 Gigabyte) An archive of all photographs.

· Flickr8k_text.zip (2.2 Megabytes) An archive of all text descriptions for photographs.

The image file names are unique image identifiers. For example, here is a sample of image file names:

Keras provides the load_img() function that can be used to load the image files directly as an array of pixels.

We tie all of this together and develop a function that, given the name of the directory containing the photos, will load and pre-process all of the photos for the VGG model and return them in a dictionary keyed on their unique image identifiers.

Next step is Pre-calculation of Photo features . This is an efficiency that means that the language part of the model that turns features extracted from the photo into textual descriptions can be trained standalone from the feature extraction model. The benefit is that the very large pre-trained models do not need to be loaded, held in memory, and used to process each photo while training the language model.

Later, the feature extraction model and language model can be put back together for making predictions on new photos. The first step is to load the VGG model. This model is provided directly in Keras. This downloads the 500-megabyte model weights to the system, which may take a few minutes.

Next, we can see directory of images and call predict() function on the model for each prepared image to get the extracted features. The features can then be stored in a dictionary keyed on the image id.

Next step is loading Descriptions. The file Flickr8k.token.txt contains a list of image identifiers (used in the image filenames) and tokenized descriptions. Each image has multiple descriptions.

In each we use the first caption for all the images. we used the function load_doc() to load the entire annotations file (‘Flickr8k.token.txt‘) into memory. This function when given a filename, returns the document as a string. Then we use load_descriptions() function that take the loaded file, process it line-by-line, and return dictionary of image identifiers to their first description.

Next step is preparing Description Text:

The descriptions are tokenized which means that each token is comprised of words separated by white space. We implemented these simple cleaning operations by function clean_descriptions() that cleans each description in the loaded dictionary from the previous section.

Whole Description Sequence Model:

The image may be encoded using a pre-trained model used for image classification, such as the VGG trained on the ImageNet model. The output of the model is a probability distribution over each word in the vocabulary. The sequence formed is as long as the longest photo description.

The descriptions needed to be first integer encoded where each word in the vocabulary is assigned a unique integer and sequences of words would be replaced with sequences of integers. We used tools in Keras to prepare the descriptions for this type of model.

Loading ResNet50

ResNet50 is a variant of ResNet model which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8 x 10^9 Floating points operations. It is a widely used ResNet model and we have explored ResNet50 architecture in depth.

Step involved to load ResNet50 are:

· Loading of training/testing ids and depths: Reading the training data and the depths, store them in a DataFrame.

· Reading images and masks:

Load the images and masks into the DataFrame and divide the pixel values by 255.

· Calculating the salt coverage and salt coverage classes:

Counting the number of salt pixels in the masks and dividing them by the image size. Also create 11 coverage classes, -0.1 having no salt at all to 1.0 being salt only. Plotting the distribution of coverages and coverage classes, and the class against the raw coverage.

· Creating train/validation split stratified by salt coverage:

Using the salt coverage as a stratification criterion. Also show an image to check for correct upsampling.

VII. APPLICATIONS AND FUTURE WORK

Future work:

· Larger datasets means more training means more accuracy to the resultant photo or image.

· Advanced architecture model (like Attention module)use may increase the value of the project.

· More hyperparameter tuning may lead more accuracy.

Applications:

· For visually impaired people object can be easily determined.

· Advertisements and posters

· Creating ideas

· Can be used in modern applications like twitter, facebook, snapchat and instagram.

VIII. REFERENCES

- · https://www.ijert.org/image-captioning-using-deep-learning#:~:text=A%20Convolutional%20Neural%20Network%20is,is%20shown%20in%20figure
- · cvpr2015.pdf(stanford.edu)
- · https://en.wikipedia.org/wiki/Natural_language_processing
- · https://searchenterpriseai.techtarget.com/definition/convolutional-neural-network
- · (175)Andrej Karpathy - Automated Image Captioning with ConvNets and Recurrent Nets
- · https://www.geeksforgeeks.org/image-caption-generator-using-deep-learning-on-flickr8k-dataset/
- · Elements of Machine Learning, Pat Langley Morgan Kaufmann Publishers, Inc. 1995. ISBN 1-55860-301-8

· The elements of statistical learning, Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. Vol. 1. Springer, Berlin: Springer series in statistics, 2001.