Prediction of Hashtags for Images

Prajwal Bharadwaj N [1], Taurunika Shivashankaran [1], Madhushree S [1], Prajwal K [1], Sachin D N [2]

[1,2] Department of Computer Science and Engineering
Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

Abstract: **Hashtags, usually, are one among the everyday patterns in web-based locale lives. They're often used with pictures or writings through web-based networking social media. They are used to simplify the venture of categorizing any photo that has been uploaded on social media. But, however, manual annotations for images, additionally for figuring out training sets for the machine learning algorithms, wishes introduced effort, work and might also incorporate human judgmental blunders or subjectivity. Hence, alternate methods to routinely generate training sets, like pairs of pictures and also tags are grasped. Choosing or wondering about a realistic hashtag for a photograph is an unwielded procedure. Machine learning helps in making this method easier. In this paper, we have labored on constructing our personal dataset of images that can be used to predict appropriate hashtags for images. The dataset carries pictures under various categories that have been trained and tested in order to be classified. This helps us become aware of and segregate photos and to predict their splendid hashtags.**

Keywords—Hashtags; machine learning; predict; social media

I.   INTRODUCTION

Hashtags are single words, word concatenations or abbreviations that are prefixed by the symbol "#". They commonly accompany the   images   online, most possibly on the social media platforms. Hashtags, most regularly are utilized to condense whole of the substance of a client's post as well as catch the consideration or notice of their followers. On Instagram, for instance, straightforward hashtags like #cat and #hill depict basic items or areas in a photograph. Passionate hashtags, for example, #love express a client's sentiments, unique hashtags, for example, #itsfashion and #autumn order themes and inferable hashtags such as #colorful and #occupied speak to circumstantial or relevant data. There are also a wide vary of popular hashtags that supply summary ideas, and are not always associated to a specific picture content material such as #nomakeup, #lazysunday [6].

Modern structures for understanding net content material make broad use of machine learning algorithms for image recognition. In definite, deep learning techniques such as convolutional networks have grown to be very famous due to their amazing overall performance [2]. Training such fashions has generally depended on giant sets of manually annotated data, which maybe time-consuming and onerous to fetch. Furthermore, such records avoid countless aspects of image appreciation that are of distinct hobby to web users: (i) their aim is on specific physical description so elements such as sentiment are now not mentioned; (ii) the records distribution varies from online information and it is additionally not going to adapt shortly to changing consumer interests and (iii) labels are impartial of the users who

sincerely authored the snapshots or posts [10].

In this work, we considered the sizeable number of image content on the web where customers have supplied hashtags as a powerful, alternate training statistics source. And also, with producing very massive amounts of labelled dataset in comparison to a manually generated one, we can also at once instruct on the authentic data we want to capture, as a substitute to the one whose distribution shall differ from consumer's interests. We hence outline our training job as how we want to predict our hashtags for a given picture uploaded by way of a precise user. The learning techniques that we use as a result contains characteristic representations modeling each the image by way of pixels and the user by the way of metadata [12]. Thus, the hypothesis is that the mixture of these sources presents useful data.

Among the principal barriers in perceiving things from pictures is the quantity of various classes which the picture may contain is huge. The information assortment and explanation process for every one of these classes might be excessively wasteful and impracticable. The other method to perceive these articles is to structure calculations that re-enact how people conquer this issue. A person can recognize the article being referred to regardless of whether it might be the first occasion when he sees it. This derivation is finished by pulling data about this article from an alternate source and afterwards utilizing this data to attempt to distinguish the item [4].

To discover popular hashtags and predict their relevance in the near future, big companies can dispose the time and resources to discovering these popular hashtags. They can work their advertising plans around it and create social media especially to a trend of that time. However, small businesses do not have the resources to use this kind of marketing strategy, since many times they are limited to what their product represent. Traditional content creators like photographers and videographers that try to sell their media by taking advantage of trends and reaching out to their niche market, are good examples of those cases [3]. This means that small businesses don't have funds to find out trends and fit to their product since their only strategy is to work around their product. For that, there is a need for an automated tool that analyses their content and returns the proper hashtags to that context and provides a metric for the popularity effect that each hashtag will have. Consider that, we are not looking toward to if an hashtag is popular or not, - which can be translated by counting how many times that hashtag is searched or even with other metrics - but the effect that the hashtag has on social media where it's used [5].

Our developed methods are widespread, strong and scalable. Similarly, they can be mixed with a range of

machine learning algorithms and are also used in large-scale real-time circumstances. Image hashtag prediction, as such, has a range of applications. For example, social networking sites may want to use such a system to promote hashtags to customers while they add photo content. They may also be used to search for photos or recommendations and notation of images mainly according to the content. Hashtags can also perform different functions such as removing synonyms (e.g. pool #swimming vs pool #snooker) or determining characteristics (#blue). In our experiments, we follow our ways to a giant data set of de-identified Instagram messages and reveal that photo and consumer modelling can significantly improve first-class label prediction in today's new techniques.

II. RELATED WORK

Given that millions of images are posted every minute on all social networking platforms, it's hard to stay unique in the crowd. If you are not one of your followers, there is no hope that your posts will be seen. That's what the hashtag is for. A hashtag is a keyword before a hash symbol "#" written in a post title or comment that highlights it and allows it to be searched. Basically, by adding a hashtag to a message, one can index it via social media to make it available to everyone, even if they're not your followers. For example, if your business is about extreme sports, you can add a hashtag, say #adventurous, to your Instagram post to attract these people with an interest for adventure and amuse. Because of the hashtags, your posts are not limited to your followers. Simply add one to access the post for all other users who are interested in finding similar topics for the hashtags. Choosing the proper hashtag can widely extend the outreach of your social media posts to millions of the dedicated followers.

A. Zero Shot Classification by Generating Artificial Visual Features

For addressing Zero Shot Classification and Generalized Zero Shot Classification by trying to learn the restrictive generator utilizing seen classes and to create counterfeit training models for classifications without any duplicates [1]. ZSC has hence become a supervised standard learning problem. General analysis with four generative models and six data sets definitely approve the method as giving a cutting edge that gets about both ZSC and GZSC. In addition, when few categories are determined by just semantic descriptions (visual properties), learn how to sort images, while others also consist of some images. This task is generally known as zero-shot classification (ZSC). Most of these methodologies are based on learning embedded spaces, which allow you to compare the visual characteristics of unidentified categories with the semantic descriptions.

B. Supervised Machine Learning Algorithms

The archive breaks down the adequacy of supervised machine learning algorithms as far as precision, learning rate, complexity and risk of extreme change measures [8].

The primary goal of this article is to furnish a general correlation with cutting edge Machine learning algorithms. Shows how every algorithm varies as indicated by application zone and it isn't the situation that just one is prevalent in every situation. These algorithms address the impact of monitored machine learning algorithms on accuracy, learning speed, complexity, and overfitting operational risks. The main purpose of the method is to provide a general distinction with a variety of state-of-the-art machine learning algorithms.

C. Zero-shot learning by convex combination of semantic embeddings

For demonstrating that this basic and straight forward strategy presents a considerable lot of the points of interest related with progressively complex image embedding schemes and in truth surpasses the cutting-edge strategies in Image Net zero-Shot learning task [9]. The accomplishment of the strategy is its capacity to exploit the qualities inborn in the best in class image classifier, the cutting-edge textual embedding system from where it has been deployed. Here, a simple technique is proposed to implement an image integration system from an existing n-way image classification and a semantic word integration model, which already has n-class tags in its vocabulary. This method maps the images in the semantic integration space by a combination of class label integration vectors and requires zero extra training. This simple and straightforward technique is a series of professionals associated with more complex image integration projects, and the course accidents various cutting-edge methods in ImageNet's zero-take learning process.

D. Harrison Dataset

The dataset accounts that the task of hashtag prediction requires Extensive and relevant comprehension in the circumstance to be transmitted in the picture [7]. This work is the main endeavor at vision in the hashtag suggestion for the world in interpersonal organizations. This benchmark is required to quicken the hashtag advance recommendation. It laid out testing issues of hashtags suggestion frameworks wide understanding scope of visual data, utilization of conditions between hashtag classes and understanding logical data.

The Harrison dataset is a virtual dataset that comprises of 57,383 images from Instagram and a mean of 4.5 related hashtags per image. To evaluate this dataset, the basic configuration is designed with the convolutional neural network (CNN) and visual aspect extraction based on the multi-label classification. Based on this framework, two uniquely distinctive models, material-based models, visual models and an integrated model for them will be evaluated upon this particular data set.

E. User Conditional Hashtag Prediction for Images

Investigation of two different ways to consolidate the heterogeneous qualities in a learning structure: Simple fit and 3-way amplifier gate, on which the image data is

modelled on the user's metadata. Learning the content of user-image posts are a specifically delighting task on social networking and online media. Current machine learning methods aim primarily on the training set of labelled pairs for image marking, and making image classification easier for the pixels given in the image. Instead, the work takes advantage of the depth of data fetchable by the user, i.e. first, the user uses the hashtag to obtain a description of the image content, second, the user uses the hashtag to obtain a description of the image content, and second, the user uses the user-defined context information. The hashtag can be used to predict how user metadata, such as age, gender, is used in conjunction with image features captured from convolutional neural networks.

F. ImageNet Classification with Deep Convolutional Neural Networks

A large deep-neural network [6] Imagenet was created to standardize 1.2 million high-resolution images in 1000 individual classes in the LSVRC-2010 competition. According to test data, it reached the top-1 and top-5 of error level of 37.5% and 17.0% which is certainly better than the preceeded state of art. This deep network has about 60 million parameters and 6.5 million neurons, which include five complex layers, some of which are replaced by the maximum pooling layers, and three fully connected layers are replaced with the ultimate smooth maximum 1000 pathway. To accelerate the training, the use of unsaturated neurons and complex methods is a very effective implementation of the GPU. The recently developed regulatory system, which is not fully connected, is a "regulatory method" to reduce additional changes to fully connected layers. "Dropout, used recently, has proven very efficient.

III. BACKGROUND

Before we get into a detailed description of our project, we would like to discuss a few topics related to our study in order to have a better conceptual understanding on our project. The methodologies and tools used in our project are: Instaloader, Atom (Editor), TensorFlow.js, Transfer learning with the VGG-19 model.

A. Instaloader

Instaloader is an online tool used to download photos or other media including their titles/ captions and other metadata related to them from Instagram. It downloads various private and public profiles, related hashtags, highlights and user stories, recorded feeds and media. In addition, it downloads specific relevant captions, geotags and comments for each release, and automatically identifies profile name changes and renames the destination directory appropriately, and allows you to store the downloaded media and customization of filters. This is an open source free software that is driven in Python.

The following are the targets of Instaloader:

a) Profile
b) #hashtags
c) %location id
d) Stories
e) Feed
f) Saved
g) @profile
h) -post

Instaloader drives along all the medium that matches the given destinations and downloads the images, videos and the associated captions. In addition to this, you can specify and download geotags on every post and store them as Google Map links (needs –sign-in) and download comments on every post.

B. Atom Editor

Atom is an independently licensed open sourced text and source code editor that is reliable for the operating systems such as macOS, Microsoft Windows and Linux along an added control for various plug-ins that are driven in Node.js, and integrated with Git Control, which is originally constituted by GitHub.com. It is a desktop application constructed with different web methodologies. Many of the expansion packages include free licensed softwares and are created and maintained by the community. Atom is based on Electrons (previously referred as Atom Shell), a framework that facilitates multiplatform desktop applications with Chromium and Node.js. It was driven by Coffee Script and Less. Atom allows the users to install all third-party packages including looks required to manipulate the features and themes of the editor. Packages may be downloaded, maintained and scripted via Atom package manager (APM).

C. TensorFlow

TensorFlow.js is a free and open sourced, hardware-driven JavaScript library suitable to train and deploy machine learning systems. It is used to develop machine learning in the browser. It makes use of adaptable and intuitioned APIs to create the ones from scratch employing the JavaScript low-level linear algebra library or the layers of higher-level API. It automatically provides support for WebGL, and will drive the code at the back scenes whenever there is availability of a GPU. Users can additionally view the webpage from a remote device, in this case the system can be advantageous of the sensor data, that is from an accelerometer or a gyroscope.

D. Transfer learning with VGG-19

Based on transfer learning, the information of an already trained system gaining knowledge of model is applied to another however associated problem. Say, for example, in case you trained a not so complex classifier to predict if or not a photograph consists of a handbag, you may utilize the information which was gained by the model at some point of its training to understand other things such as the umbrella or sunglasses.

VGG-19 is a CNN based neural network with a depth of 19 layers. We can load an initial version of a trained network with greater than a million images from the ImageNet database [5]. The initial network images can be classified into 1,000 categories of classes such as the fashion, sports, nature and many more. As a result, the network has learned rich

representations of functionality for a huge variety of images. The input size of the image network is about 224 x 224.

With transfer learning and VGG-19, we essentially try to make the most what has been found out in one project to enhance generalization in another. We switch the weights that a network has discovered at "mission A" to a new "mission B."

IV. METHODOLOGY

Here, we shall have a look at our data collection procedure and ground statistics of our collected data. We shall also have a walk-through of data cleaning methods and the method of predicting hashtags.

A. Data Collection

In this experiment, we have tried collecting images from various social media sites using Instaloader. These data are to be grouped in particular categories in order to be able to be classified. These categories are the classes in which we have grouped the images. Considering we have ten classes of the datasets, where each class comprises of around 800 images each. A proper data collection method is very much essential in its own way since it assures that the statistics collected are each described and also found accurate. Likewise, subsequent decisions primarily constituted on parameters embodied within the revealing is made the use of decisive information. The system gives each a base to measure from and in positive cases a demonstration of what improvements are required. These classes are named from most popularly found topics on Instagram. The most popular classes are named as: animals, art, fashion, fitness, flowers, food, instagood, nature, selfie, sports.

B. Data Cleaning

Once the data has been collected, it is supposed to be cleaned in order to ignore the unnecessary or noisy data. This was done by surveying each class of the dataset and examining for data that has not been grouped properly or is unwanted. After cleaning, a statistic collection need to be constant with different similar datasets in the device. The inconsistency found or declined can also be originally aimed by user input mistakes, by corrupted data being transmitted and storage, and by way of one-of-a-kind statistics and dictionary based definitions of comparable values in various stores. Data cleaning differs with respect to records validating in the validation nearly continuously implies facts is declined from the device at access and is achieved at the entry time, rather than on sets of data. Here in our assignment we have accumulated the data from diverse assets and wiped clean the facts set based totally on the classes as mentioned earlier.

We have made use of transfer learning as a classification method for creating clusters of images in every class. Images that are very long way from the centroids are usually found noisy, later this was determined by the results we had got. We opted an uncontrolled threshold of two standard deviations of the mean, and also every image with space from centroid higher than the threshold found to be noisy. Generally, we were capable of deleting about 50 noisy pictures from every class by using this procedure.

After cleaning our data, we were left with around 700 images in each class which makes our classes appropriate for use.

C. Data Processing

During this level, the data which has been collected at some point of the preceding degree is honestly processed for interpretation. Processing is done using Convolutional neural network algorithm, known as transfer learning with VGG-19, though the technique itself might also vary slightly depending on the supply of data being processed and its intended use. In transfer learning with VGG-19, the classes of deep neural networks that have shown much dominance in different creative and prophetic tasks in computers, attracts interest in many areas. It is created to routinely and flexibly take a look at spatial hierarchies of features through lower back propagation by the use of multiple constructing blocks, which includes convolutional layer, fully related layers, and pooling layers [7].

VGG-19 is one of the foremost classes to do photographs popularity, photographs classifications. Objects detections, popularity face etc., are some of the areas wherein VGG-19 is extensively used. Transfer learning photo classifications take an input photo, process it and classify it underneath certain classes (E.g., selfie, Nature, Animals, and Fashion). By the use of this algorithm, we have processed and wiped clean the records based totally at the given classes.

In addition, an extension to CNN, we have also employed transfer learning method for training the classes of data accordingly. Transfer learning along with VGG-19 is one of the methods which can be used for classifying problems and regression problems in image recognition. Among these cases, the entry includes the nearest training sets inside the characteristic area.

It is a good classification framework for various other data sets and can be used when authors make the models reachable to the public as they can be utilized as it is or along modification for other similar models as well. Along with transfer learning, it can be used for facial recognition models as well. The weights are readily reachable with other architectures like keras as it can be modified with and used as per one's will. The style and content loss does not occur while using VGG-19 framework.

This technique was used by us to train the data that was previously collected. The images from all classes of the dataset were passed on to training model which was classified into a class which contained on of the hashtags related to it. It is a supervised learning method which helps train data in a more exquisite way.

Our current training model includes 3 parameters namely: epoch, batch size and learning rate.

Epoch is a hyperparameter that has been described earlier than training a model. One epoch is while a whole dataset is passed both forward and backward via the neural community most effective exactly once. One epoch is too massive to feed to the system without any delay. Thus, epoch cannot be applied to a very large dataset, it is divided to form few batches.

Batch size is known as the number of training examples in 1 Forward/1 Backward pass. (With increase in Batch size, required memory space increases.)

Learning rate is defined as a hyper-parameter which defines how much adjusting of weights in the network through the loss gradient is required. Further more there is a decrease in the entity, the lower we tour alongside the down slope. Also this is probably an excellent concept (the use of a slow learning rate) among phrases of ensuring that we do no longer omit all local minima, this can mean that we shall be making a long while to constrict, mainly if we are caught on a plateau-like vicinity.

D. Prediction of Hashtags

For the prediction of hashtags, we have used Transfer learning where TenserFlow.js and HTML are used to connect to the front end to display the result.

Firstly, we start off by acquiring images from Instagram through instaloader. This way we have quite a huge dataset of images which are around 5000. Then we begin preprocessing the photographs by means of which we have around 10 classes of images with about 700 images in each of the classes. Once the required assets and libraries are imported, we include the data and start preprocessing them. The scope of pre processing is a development of the photo records that surpasses unwillingly distorted or improvises a few image characteristics vital to similarly process, even though geometrical transformations of images (like Rotating, scaling, translating) are classified amongst pre processing techniques right there considering similar methods are recommended.

Then, we intend to train the model by using transfer learning. We prefer transfer learning as it can train with datasets of limited sizes and is very much similar to them. Hence we make use of these pretrained models. We have used the model of VGG-19 in our project which left us with a better accuracy and improved efficiency.

Then we shall convert or obtained model into Tensorflow.js. It is a famous Machine learning library for Python. TensorFlow.js is that it lets in us to load pre-trained fashions. TensorFlow.js also lets us to include already trained models within the Python model of TensorFlow.js. Which means we may describe a system and teach it with the help of Python, then store it in a location to be on the net and cargo it for your script. The approach may extensively improvise performance as we don't need to educate this version within our browser [11]. Tensorflow.js will be used

along with HTML and CSS to generate the actual webpage following which the hashtags will be predicted.

The final stage was that of combining all of these functions to a modelled framework. Here, we integrate all the listed functions into a single framework. We shall utilize small batches of the transfer learning to train and test.

Thus, using transfer learning method, similar technique can be followed to predict hashtags for various other images as well.

On performing the above-mentioned operations, we are now able to predict the hashtags. On the homepage of our website, we are supposed to uploading an image by choosing any file. After that when we click on predict, the hashtags are fetched from the datasets and are predicted with appropriately absolute accuracy. The output would look like it is shown in the below figure (fig. 3).

Fig. 1. Figure showing the predicted hashtags for the uploaded image.

V. EXPERIMENTS AND RESULTS

We have conducted various experiments in order to compare various methods for an accurate prediction of hashtags. We are glad to have found and implemented one with an impressive accuracy score. Since our aim was to predict hashtags for images with an improved accuracy, I have stated the results as follows:

Based on this training data, our test results status for each class of data is as shown in the figure below (fig. 2).
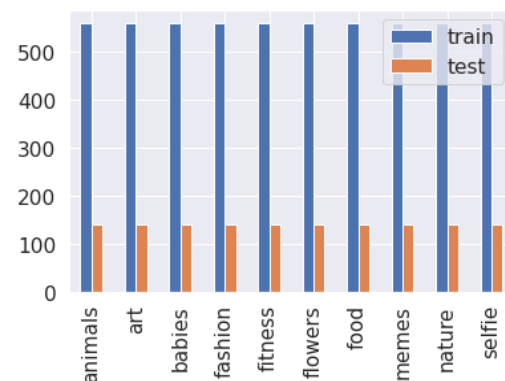

Fig. 2. Figure showing the obtained results of training and testing each class of data.

As seen from the graph below (fig. 3), the confusion matrix for non-normalized VGG-19 prediction has a definitely high accuracy that ranges to about 68%.

The loss as we can observe has always been kept minimal to ensure accurate predictions.
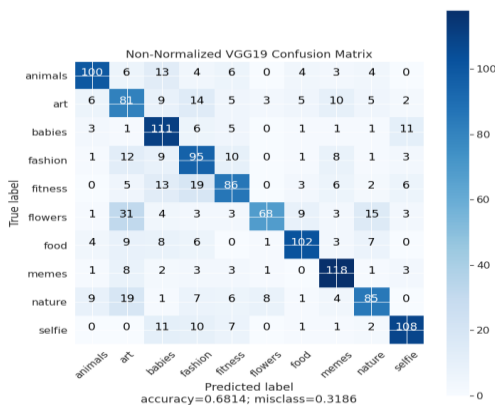

Fig. 3. Figure showing the confusion matrix for non-normalized VGG-19.

The confusion matrix for normalized VGG-19 (fig. 4) is quite similar to the non-normalized one with respect to accuracy and misclass and also matching true and predicted labels.
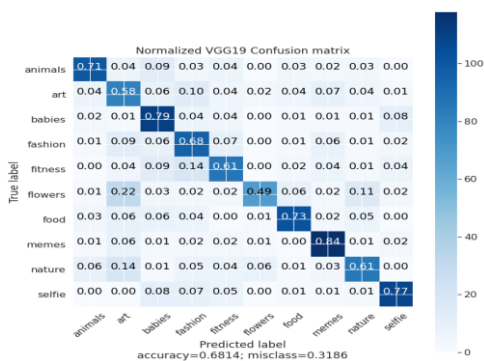

Fig. 4. Figure showing the confusion matrix for normalized VGG-19.

When we tried to test our model for loss based on the trained data, we obtained the graph as shown in fig. 5. The model loss for our test data has been limited to be as low as possible.
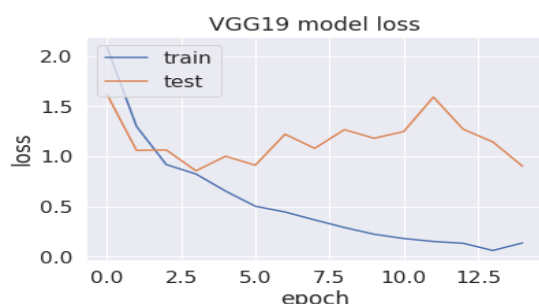

Fig. 5. Figure showing the train and test curve for model loss using VGG-19.

When we finally tested our model for accuracy, it gave us the following result as shown in fig. 6. The accuracy score has been maintained to be as high as possible upto 68.14%.
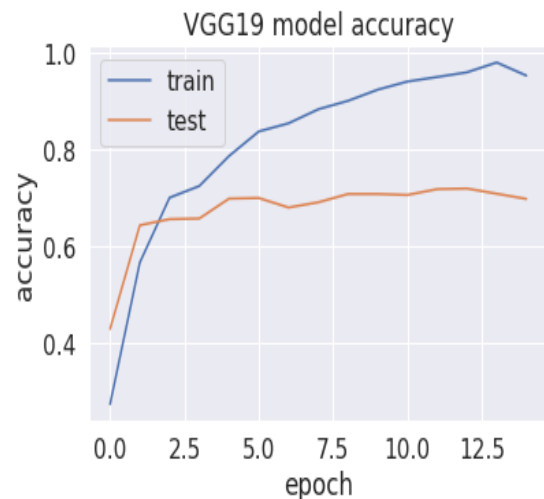

Fig. 6. Figure showing the train and test curve for model accuracy using VGG-19.

VI.   CONCLUSION

From our tests, it has turned out that the data collection and cleansing leads to very excellent objective results, even if they are extensively computed and time-consuming. Also, to develop an end-to-end image recognition system was a great learning experience and had a lot of excitement in implementing the ideas we learned in the classroom. With the advent of machine learning, emerging methods such as CNN, KNN and transfer learning are checked by our experiments. But deep learning methods sometimes need a huge amount of data to hook up. The next step is to improve our classification accuracy and compare it in fully supervised ways. You can also explore the common problem where test classes include training labels. So, we can create a more promising and better model in the future.

REFERENCES

[1]   Stéphane Herbin, Maxime Bucher, Frédéric Jurie. Zero-Shot Classification by Generating Artificial Visual Features. RFIAP, Jun 2018, Paris, France.hal-01796440

[2]   Paredesand Philip HS Torr - Bernardino Romera. An embarrassingly simple approach to zero-shot learning. In ICML, pages 2152–2161, 2015.

[3]   Shreyash Pandey, Abhijith Pathak. Predicting Instagram tags with and without data.

[4]   Stephane Herbin, Maxime Bucher, Frederic Jurie. Generating Visual Representations for Zero – Shot Classification, arXiv:1708.06975v3 [cs.CV] 11 Dec 2017.

[5]   Mohammad Norouzi et al. "Zero-shot learning by convex combination of semantic embeddings". In: arXiv preprint arXiv:1312.5650 (2013).

[6]   Ilya Sutskever, Geoffrey E Hinton, and Alex Krizhevsky. "ImageNet Classification with Deep Convolutional Neural Networks". In: Advances in Neural Information Processing Systems Twenty-five. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. url: http://papers.nips.cc/paper/ 4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[7] Hanxiang Lee, Minseok Park, Junmo Kim. HARRISON: A Benchmark on Hashtag Recommendation for Real-world Images in Social Networks, arXiv:1605.05054v1 [cs.CV] 17 May 2016.

[8] Narina Thakur, Amanpreet Singh and Aakanksha Sharma. A Review of Supervised Machine Learning Algorithms, 978-9-3805-4421-2/16/$31.00 c 2016 IEEE.

[9] Tomas Mikolov, Yoram Singer, Mohammad Norouzi, Andrea Frome, Jonathon Shlens, Samy Bengio, Greg S. Corrado, Jeffrey Dean. Zero-Shot Learning by Convex Combination of Semantic Embeddings. arXiv:1312.5650v3 [cs.LG] Mar 2014.

[10] Jason Weston, Emily Denton, Manohar Paluri, Lubomir Bourdev, Rob Fergus. User Conditional Hashtag Prediction for Images, 2015 ACM. ISBN 978-1-4503-3664-2/15/08..$15.00.DOI: http://dx.doi.org/10.1145/2783258.2788576.

[11] O. Tsur, and A. Rappoport, "What's in a Hashtag? A Content Based Prediction of the Spread of Ideas in Microblogging Communities". In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, 2012, pp. 643-652.

[12] Y.Y. Ahn, F. Menczer, and L. Weng, "Virality Prediction and Community Structure in Social Networks", Scientific Reports, 2013, 3 (2522).

[13] Hanbury, A survey of methods for image annotation, J. Vis. Lang. Comput. 19 (5) (2008) 617–627.

[14] L. Fei-Fei, A. Karpathy, "Deep visual-semantic alignments for generating image descriptions", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'15 and IEEE Computer Society, 2015, pp. 3128–3137