

Sorting Instagram Hashtags all the Way throw Mass Tagging using HITS Algorithm

D.Vishnu Vardhan ¹, Mrs. Dr.CH.Aparna²

M.Tech Scholar CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India. vv.devineni9999@gmail.com

² Professor, CSE Department, R.V.R & J.C College of Engineering,
Guntur, Andhra Pradesh, India. aparna@rvrjc.ac.in

Abstract

Instagram is one of the fastest-growing online photo social web services where users share their life images and videos with other users. Image tagging is an essential step for developing Automatic Image Annotation (AIA) methods that are based on the learning by example paradigm. Hashtags can be used on just about any social media platform, but they're most popular on Twitter and Instagram. Using hashtags is essentially a way to group together conversations or content around a certain topic, making it easy for people to find content that interests them. Practically on average, 20% of the Instagram hashtags are related to the actual visual content of the image they accompany, i.e., they are descriptive hashtags, while there are many irrelevant hashtags, i.e., stop-hashtags, that are used across totally different images just for gathering clicks and for search ability enhancement. Hence in this work, Sorting instagram hashtags all the way through mass tagging using HITS (Hyperlink-Induced Topic Search) algorithm is presented. The hashtags can sorted to several groups according to Jensen-Shannon divergence between any two hashtags. This approach provides an effective and consistent way for finding pairs of Instagram images and hashtags, which lead to representative and noise-free training sets for content-based image retrieval. The HITS algorithm is first used to rank the annotators in terms of their effectiveness in the crowd tagging task and then to identify the right hashtags per image.

Keywords:

Instagram, Automatic Image Annotation (AIA), Hashtags, Hyperlinked-Induced Topic Search (HITS) algorithm.

1. Introduction

Social media is an interactive communication technology that eases the creation and sharing of information, ideas, and other forms of expression via virtual communities and networks..Instagram is a free social networking platform specially for young sters , allow- ing users to upload and share photos/videos through a mobile app [1]. Instagram is one of the most used social media platforms where around 500 million users interact with content daily, which creates excellent marketing opportunities. Instagram is a social media that makes it easy for people to tell stories and engage with readers through visual displays.

Instagram's marketing growth is primarily focused on the page's content, but suitable hashtags are equally essential to traffic. Hash tag research is one of the most complex parts of an Instagram marketing campaign, and usually, online tools are used to get a set of hashtags from one niche-specific hashtag These tools are suitable for initial days, but their recommendations are often less in number, outdated, and lack connectivity among hashtags [2].

In the last decades, the use of social media platforms has been constantly growing, and so has been the need to categorize huge amounts of information and to develop tools for finding what users are searching for. Hashtags became an everyday reality of the frequenters of social networks: Twitter, Face book and Instagram, and even messengers like WhatsApp now offer the feature. But a brief glance into the web environment reveals that hashtags have become much more than just topic markers – people use hashtags for anything, from proselytizing to informing the readers In fact, hashtags have become an integral part of social networks [3]. AutoCaption is a caption generating system for custom photos. After uploading a photo to the cloud service, several modules (face recognition, proxemics, landmarks, as well as scene classification and metadata processing) work in parallel, and then the output data goes into the text generator. The user can choose the captions he likes, as well as change the order, add or remove words. In most cases, this method of generating captions allows you to accurately determine what is happening in the photo, but a significant disadvantage of this method is that the caption in this case is the actual description of the photo and is completely unsuitable for a social network, especially for Instagram [4].

Instagram is a relatively new form of communication where users can easily share their updates by taking photos and adjusting them using filters. Many companies and business actors are placing advertisements as promotional media on Instagram to build consumer buying desires [5]. The use of hashtags is undoubtedly a part of our digital life. There is a hashtag for almost every social interest, for example, political causes or protests, branding or advertising campaigns, genre representation, the awareness of illness, erotic conten, tourism, gastronomy, memories, and so on as natively digital objects. Hashtags may serve as indexes for their functions, meanings, and practices [6].

Hashtags are the best means of attracting attention to content. In posts that have at least one hashtag, engagement rates are on average 12.6% higher than in publications without hashtags. Hashtags are an important part of discovery on Instagram, allowing brands to gain exposure to niche groups and specific areas of interest. While they may not drive exponential engagement growth, they give audiences an organic way to discover branded content through the topics and forums that interest them. Proper use of hashtags is a key condition for successful promotion of their content in the most popular social network today which is Instagram, where millions of users daily post new photos, leave comments and likes, follow the photos of other people's lives, promote products and services [7]. Social media, and especially the Instagram, provide a rich source of image-tag pairs. Mining the right ones, automatically or semi automatically, so as to be used as training examples is extremely important. We have to consider, however, that, in many cases, hashtags that accompany images in social media are not related with the image's content but serve several other purposes such as the expression of user's emotional state, the increase in user's clicks and findability, and the beginning of a new communication or discussion. Hence in this work, sorting instagram hashtags all the way throw mass tagging using HITS algorithm is presented. The rest of the work is organized as follows: The section II describes the different researches related to hashtags sorting. The section III demonstrates sorting instagram hashtags all the way throw mass tagging using HITS algorithm. The result analysis of hashtags sorting is discussed in section IV. Finally this work is concluded in section V.

2. LITERATURE SURVEY:

Nartlada Bhakdisuparit, Iwao Fujino, et. al., [8] demonstrates Understanding and Clustering Hashtags According to their Word Distributions. The purpose of this study is to understanding hashtags used in twitter and instagram and clustering hashtags according to their word distribution, so that we can discover the public trend of user's topic in real time and bring benefit to marketing management. Fei-Fei Kou, Jun-Ping Du, Cong-Xian Yang, Yan-Song Shi, Wan-Qiu Cui Mei-Yu Liang, and Yue Geng et. al., [9] presents Hash tag Recommendation Based on Multi-Features (HRMF) of Microblogs. First, the HRMF expands short text into long text, and then it simultaneously models multi-features (i.e., user, hashtag, text) of microblogs by designing a new topic model. To further alleviate the data sparsity problem, HRMF exploits hashtags of both similar users and similar microblogs as the candidate hashtags Experimental results on a real-world dataset crawled from Sina Weibo demonstrate the effectiveness of this HRMF for hashtag recommendation. Kelley Cotter et. al., [10] discussed about Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. Through a thematic analysis of online discussions among Instagram influencers, the author observed that influencers' pursuit of influence resembles a game constructed around "rules" encoded in algorithms. Within the "visibility game," influencers' interpretations of Instagram's algorithmic architecture and the "game" more broadly act as a lens through which to view and mechanize the rules of the game. Illustrating this point, this article describes two prominent interpretations, which combine

information influencers glean about Instagram's algorithms with preexisting discourses within influencer communities on authenticity and entrepreneurship. Aleksandra Laucuka et. al., [11] demonstrates Communicative Functions of Hashtags. Different uses of hashtags were subjected to semantic analysis in order to disclose generalizable trends. As a result, ten communicative functions were identified: topic-marking, aggregation, socializing, excuse, irony, providing metadata, expressing attitudes, initiating movements, propaganda and brand marketing. These findings would help to better understand modern online discourse and to prove that hashtags are to be considered as a meaningful part of the message. Yuyun Gong, Qi Zhang et. al., [12] suggests Hash tag Recommendation Using Attention-Based Convolutional Neural Network. They presents a novel architecture with an attention mechanism. The results of experiments on the data collected from a real world microblogging service demonstrated that the this model outperforms state-of-the-art methods. By incorporating trigger words into the consideration, the relative improvement of the proposed method over the state-of-the art method. Nicholas Carah, Michelle Shaul et. al., [13] demonstrates Brands and Instagram: Point, tap, swipe, glance. They conceptualize Instagram as an image machine that captures and calibrates attention. Instagram expands the terrain upon which brands operate by dispersing the work of creating and engaging with images into consumers' everyday lives. The efforts made by brands to experiment with mobile media demonstrate the need to critically examine how participatory, discursive, and algorithmic modes of control are interrelated. Stamatios Giannoulakis, Nicolas Tsapatsoulis et. al., [14] Evaluated the descriptive power of Instagram hashtags. Authors investigate whether tags accompanying photos in the Instagram can be considered as image annotation metadata. If such a claim is proved then Instagram could be used as a very rich, easy to collect automatically, source of training data for the development of AIA techniques. Hypothesis is that Instagram hashtags, and especially those provided by the photo owner/creator, express more accurately the content of a photo compared to the tags assigned to a photo during explicit image annotation processes like crowd sourcing. Tim Highfield and Tama Leaver et. al., [15] presents A methodology for mapping instagram hashtags. This paper proposes a methodology for studying Instagram activity, building on established methods for Twitter research by initially examining hashtags, as common structural features to both platforms. In doing so, the authors outline methodological challenges to studying Instagram, especially in comparison to Twitter.

3. SORTING INSTAGRAM HASHTAGS:

In this work, sorting instagram hashtags all the way throw mass tagging using HITS algorithm is presented. Instagram hashtags given by crowds are filtered to detect whether hash tag is correct or not which is given by crowds. Here HITS algorithm is used to identify correctness of tags. Now-a-days online social network users are posting messages with related pictures and the hash tags will be assigning to that picture. This related hash tags make other users to search that image easily.

Sometime some user's assigns unrelated hash tags to images which make searching process difficult. To overcome from this

issue hash tags sorting technique is used that will filter hash tags to determine whether hash tag is relevant or irrelevant by matching content of both main hash tag and the annotator hash tags. If annotator assigns related hash tags then it will be relevant and supervisor will give good score to that annotator. HITS algorithm is used to determine whether that hash tags is used more frequently or not, if it's less frequent or unrelated hash tag then a stop has tag will be considered. The Fig. 1 shows the subgraph of user-tag bipartite network for image #7.

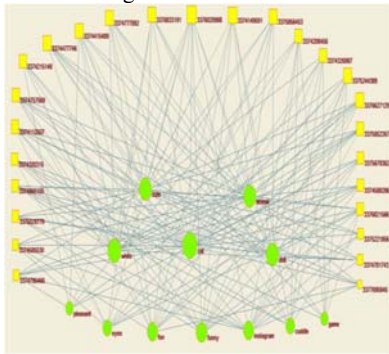


Fig. 1: SUBGRAPH OF USER-TAG BIPARTITE NETWORK FOR IMAGE #7

In Fig. 1 the circles represent the tags, Boxes represent the annotators that selected those tags.

Fig. 2 shows a flow chart of sorting hashtags process using hits algorithm. At first, for each hashtag, a term document matrix is generated from cleared instagram posts data and the result is saved as a data frame. Then the word probability is calculated using the following expression. The appearance number of each word from the document of hashtag is counted and Probability is calculated as follows:

$$P(A) = \frac{\text{number of term}}{\text{number of all terms in the document}} \quad (1)$$

The results of this procedure are stored in 3 column data frame names of word, frequency and probability. This process will provide the typical words or important words for each hashtag and guide us to understanding what is the meaning or content contained in each hashtag. This is how the strongly related words are analyzed for hashtag and use the proceeds data for advertisers and online marketing. Clustering procedure is to cluster and merge each hashtag to several groups according to its word distribution. Here we use data frame from the last procedure. Jensen-Shannon divergence is a distance measure used for clustering hashtags.

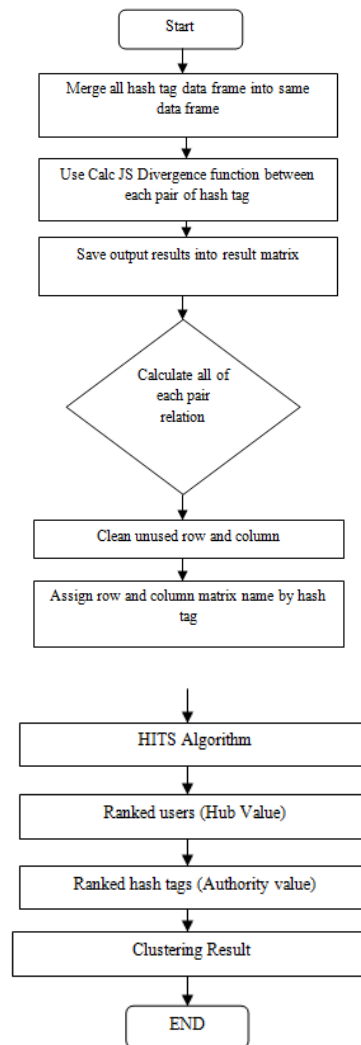


Fig. 2: THE FLOW CHART OF SORTING HASHTAGS PROCESS USING HITS ALGORITHM

For two given probability distributions P and Q, the Jensen-Shannon divergence is calculated according to its definition as

$$D_{JS}(P||Q) = \frac{1}{2}(D_{KL}(P||R)) + D_{KL}(Q||R) \quad (2)$$

Where, $R = \frac{1}{2}(P + Q)$, It is a measure of how much one probability distribution diverges from another probability distribution, or in reverse words, it is a measure of how much one probability distribution are similar with another probability distribution, However, Jensen-Shannon divergence is a symmetrized and smoothed version of the Kullback-Leibler divergence. The expression $D_{KL}(P||R)$ in Eq. (2), which is Kullback-Leibler divergence between two probability distributions P and R, can be calculated as

$$D_{KL}(P||R) = \sum_i P(i) \log \frac{P(i)}{R(i)} \quad (3)$$

The Input data for this procedure are a set of hashtags term document matrix. Words distribution of each paired hashtag is merged and calculated Jensen- Shannon divergence by calling Calc JS Divergence function and store all result to matrix. From this result, each hashtag in the sample is grouped which hashtags similarity and can be group to another hashtag and which one can be related to another.

Hyperlink-induced topic search (HITS) is a ranking algorithm. It is used to filter Instagram hashtags and locate the most relevant. The purpose of the HITS algorithm is to rate webpages. The basic idea is that a webpage can provide information about a topic and also relevant links for a topic. Thus, webpages belong to two groups: pages that provide good information about a topic ("authoritative") and those that give to the user good links about a topic ("hubs"). Authority pages are pages that are relevant to a specific topic and are linked to by many other pages. On the other hand, hubs are pages that link to many related authorities. The authority $a(p)$ and hub value $h(p)$ of a page p (or network node in general) are (iteratively) computed with the aid of the following equations

$$a(p) = \sum_{i=1}^n h(i) \quad (4)$$

$$h(p) = \sum_{i=1}^n a(i) \quad (5)$$

Where n is the total number of pages that the page p connects to and i is one of these pages. Thus a page's hub value is the sum of the authority scores of all its linking pages.

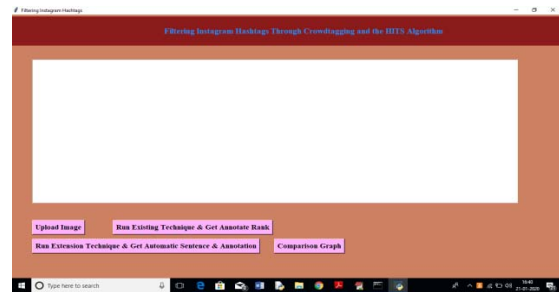
The final hub-authority values of nodes are determined after infinite repetitions of the algorithm. As directly and iteratively applying the above equations leads to diverging values, it is necessary to normalise these values after every iteration so as to sum to 1, i.e., $\sum_p h(p) = 1$, $\sum_p a(p) = 1$. By definition the initial values of $a(p)$ and $h(p)$ are set to 1. In simple words the main principle of the HITS algorithm is that it gives high weights to those hubs that are linking to pages that other hubs are linking to as well, while it gives high authority values to those nodes that are indexed (linked to) by hubs that have high hub values.

The HITS algorithm gives to each webpage both a hub and an authoritative value. In the presented approach, the HITS algorithm in a real crowd tagging environment facilitated by the Figure-eight, formerly known as Crowd flower, crowd sourcing platform. In addition, the number of annotations per image is increased to 500, the bipartite graphs are formed for all images, and the performance of annotators is measured across all those images. Moreover, FolkRank is used as a baseline to evaluate the performance of the presented method. Last step will be clustering calculated result.

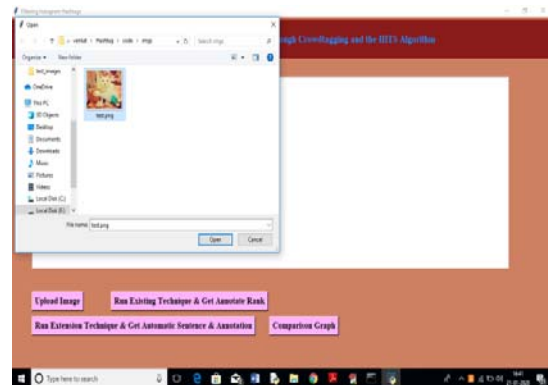
4. RESULT ANALYSIS:

In this work, sorting the instagram hashtags all the way through mass tagging using HITS algorithm is implemented using python. The following steps are followed for the implementation of presented approach. Add Module Details: to build a project I some sample products are used to train the identification models of product. Train Module: In this Module screen train model generated with 100% accuracy and now show product to web cam. Add/Remove Product from basket: To allow application to identify product image and then show in text area and if same product is shown then application will remove from text area. The result analysis of presented approach is demonstrated in this section.

Double click on 'run.bat' file to get below screen.



Click on 'Upload image' option on above screen for uploading an image.

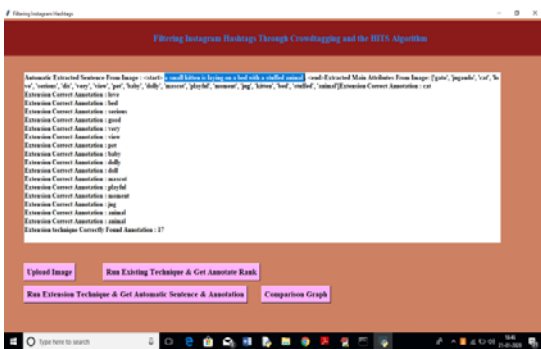


One image is uploaded on the above screen and by seeing that image anybody can say that cat or kitten sitting on a bed with some stuff and our extension will describe same sentence or extract same data from image but existing technique just will check whether given hash tag and annotator tags are similar or relevant or not relevant. After uploading the image the below screen will appear.

Now click on 'Run Existing Technique & Get Annotation Rank' button to get below screen.



In the above screen, the loaded images which are above the annotation are correct as image contains doll, cat, cute, etc. the earlier technique is able for the extraction of 8 correct annotations from all the annotated text. Next click on Run Extension Technique & get Automatic Sentence & Annotation' button to describe image in sentence and to check extracted words are matching with annotators words or not.

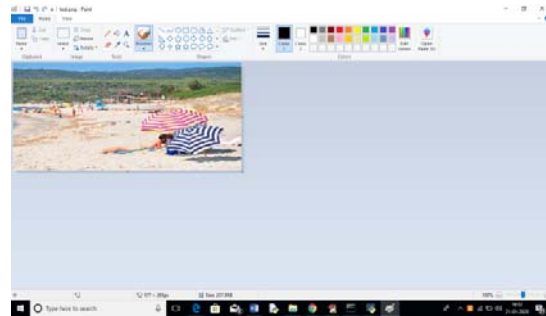


In above screen in selected text one can see this extension technique describing image in a sentence and then extracting words from image and compare with annotator's tags to get relevant details. Extension technique able to extract 17 related annotations. Now click on 'Comparison Graph' button to get below graph.



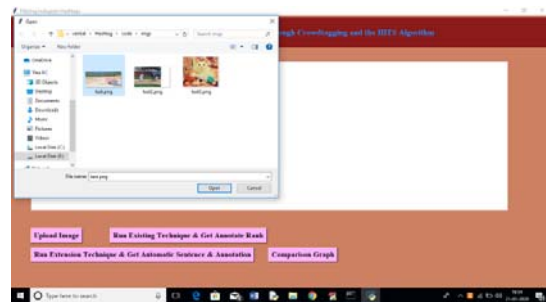
In above graph x-axis represents technique name and y-axis represents count of extracted matching annotations and the

extension technique is able to extract more related words compare to existing technique.



Earlier technique can able to check with only one image whereas presented technique will work with any image. Another example for sorting the instagram hashtags is shown in below figures.

In above image the people are on beach with umbrellas and extension technique can extract this information but cannot compare with earlier technique as it not include this image in its annotation dataset.



In above screen upload the same image and then click on 'Run Extension Technique & get Automatic Sentence & Annotation' button to get below results.



In above screen in selected text the sentence is shown which describes image and its related attributes or hashtag also displayed.

5. CONCLUSION:

In this work, sorting instagram hashtags all the way through mass tagging using HITS algorithm is presented. An innovative methodology is implemented using HITS algorithm and based on the principles of collective intelligence to identify the hashtags which describe the visual contents of image they are associated with. The word probability distribution for understanding the meaning contained in hashtag expression and Jensen-Shannon divergence is used for clustering hashtags into meaningful groups. The application of a two-step HITS algorithm in a crowd tagging context provides an easy and effective way to locate pairs of Instagram images and hashtags. By thresholding the authority scores of the image, obtained by the application of the HITS algorithm, ranking is given and then effectively locate the hashtags that are relevant to their visual content as per the evaluation.

References

- [1] Aparna Sharma, Kavish Sanghvi, Prathamesh Churi, "The impact of Instagram on young Adult's social comparison, colourism and mental health: Indian perspective", *International Journal of Information Management Data Insights* 2 (2022) 100057, Elsevier 2022, doi.org/10.1016/j.jjime.2022.100057
- [2] Implementing Prioritized-Breadth-First-Search for Instagram Hashtag Recommendation, "2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)", doi: 10.1109/Confluence52989.2022.9734217
- [3] Sofia Almeida, Ana Ramires and Joao Marinho, "Hashtags-a Keystone of Instagram in the hotel design strategy-an Iberian case study, *Journal of Quality Assurance in Hospitality & Tourism*, doi.org/10.1080/1528008X.2022.2050876
- [4] Vladimir Volobuev, Polina Y. Afonichkina, "Generating Photo Captions for Instagram", 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), 978-1-6654-0476-1/20, 2021 IEEE, DOI: 10.1109/ElConRus51938.2021.9396441
- [5] Rafki Chandra Wibawa, Chairani Putri Pratiwi, Hasna Larasati, "The Role of Nano Influencers Through Instagram as an Effective Digital Marketing Strategy", *Advances in Economics, Business and Management Research*, 2021, volume 198, Conference Towards ASEAN Chairmanship 2023 (T-A-C 23 2021)
- [6] Janna Joceli Omena, Elaine Teixeira Rabello, and André Goes Mintz, "Digital Methods for Hash tag Engagement Research", *Social Media + Society*, July-September 2020: 1-18, DOI: 10.1177/2056305120940697
- [7] Solomiia Fedushko, Sofia Kolos, "Effective Strategies for using Hashtags in Online Communication", *International Journal of Computing and Related Technologies*, Volume 2, issue 2, 2019, doi:researchaget.net/publication/332875974
- [8] Nartlada Bhakdisuparit, Iwao Fujino, "Understanding and Clustering Hashtags According to their Word Distributions", 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 978-1-5386-5254-1/18
- [9] Fei-Fei Kou, Jun-Ping Du, Cong-Xian Yang, Yan-Song Shi, Wan-Qiu Cui
- [10] Kelley Cotter, "Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram", *new media & society* 1-19, 2018, sagepub.com/journals-permissions
- [11] Yuyun Gong, Qi Zhang, "Hashtag Recommendation Using Attention-Based
- [12] Nicholas Carah, Michelle Shaul, "Brands and Instagram: Point, tap, swipe, glance", *Mobile Media & Communication*
- [13] Stamatios Giannoulakis, Nicolas Tsapatsoulis, "Evaluating the descriptive power of Instagram hashtags", *Journal of Innovation in Digital Eco Systems* 3(2016), 114-129, .doi.org/10.1016/j.jides.2016.10.001
- [14] Tim Highfield and Tama Leaver, "A methodology for mapping instagram hashtags", volume 20, number1, *Peer reviewed Journal on the Internet*, 2015, DOI: 10.5210/fm.v20i1.5563



D.Vishnu Vardhan

M.Tech Scholar-CSE Department, R.V.R & J.C College of Engineering, Guntur, Andhra Pradesh, India. vv.devineni9999@gmail.com



Mrs. Dr.CH.Aparna

Professor, CSE Department, R.V.R & J.C College of Engineering, Guntur, Andhra Pradesh, India. aparna @rvrjc.ac.in