



Filtering Online Media Hashtags Using AIA System

¹Dr. DJ. Samatha Naidu, ²M. Sai Narasimha

¹Principial of Annamacharya, PG College of Computer Studies, Rajampet.
Student MCA 4th Semester, Annamacharya PG College of Computer Studies, Rajampet.

ABSTRACT

Instagram is a rich source for mining descriptive tags for images and multimedia in general. The tags–image pairs can be used to train automatic image annotation (AIA) systems in accordance with the learning by example paradigm. Present a novel methodology, based on the principles of collective intelligence that helps in locating those hashtags.

Keywords: FILTERING, WHITEBOX TESTING BLACKBOX TESTING, AIA SYSTEM

INTRODUCTION

In particular, we show that the application of a modified version of the well-known hyperlinkinduced topic search (HITS) algorithm, in a crowdtagging context, 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. As a proof of concept, we used the crowdsourcing platform Figure-eight to allow collective intelligence to be gathered in the form of tag selection (crowdtagging) for Instagram hashtags. The crowdtagging data of Figure-eight are used to form bipartite graphs in which the first type of nodes corresponds to the annotators and the second type to the hashtags they selected. The HITS algorithm is first used to rank the annotators in terms of their effectiveness in the crowdtagging task and then to identify the right hashtags per image.

MOTIVATION

The main problem of the content-based image retrieval is the so-called semantic gap [30], [35], [37], [42]: content-based retrieval is associated with low-level features while humans use high-level concepts for their search. To overcome this problem, automatic image annotation (AIA) methods were developed, that is, processes by which computing systems automatically assign metadata in the form of captions or keywords to images [4]. Among the AIA methods, those based on the learning by example paradigm are probably the most common one [21]. A small set of manually annotated training images are used to train models, which learn the correlation between image features and textual.

EXISTING WORK

instagram is a rich source for mining descriptive tags for images and multimedia in general. The tags–image pairs can be used to train automatic image annotation (AIA) systems in accordance with the learning by example paradigm. In previous studies, we had concluded that, 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 rrelevant hashtags, i.e., stop-hashtags, that are used across totally different images just for gathering clicks and for searchability enhancement

2.1. RELATED WORK

Topic modelling on Instagram hashtags: An alternative way to automatic image annotation

Authors: A. Argyrou, S. Giannoulakis, and N. Tsapatsoulis,

Automatic Image Annotation (AIA) is the process of assigning tags to digital images without the intervention of humans. Most of the modern automatic image annotation methods are based on the learning by example paradigm. In those methods building the training examples, that is, pairs of images and related tags, is the first critical step. We have shown in our previous studies that hashtags accompanying images in social media and especially the Instagram provide a reach source for creating training sets for AIA. However, we concluded that only 20% of the Instagram hashtags describe the actual content of the image they accompany, thus, a series of filtering steps need to apply in order to identify the appropriate hashtags. In this paper we apply topic modelling with Latent Dirichlet Allocation (LDA) on Instagram hashtags in order to predict the subject of the related images. Since a topic is

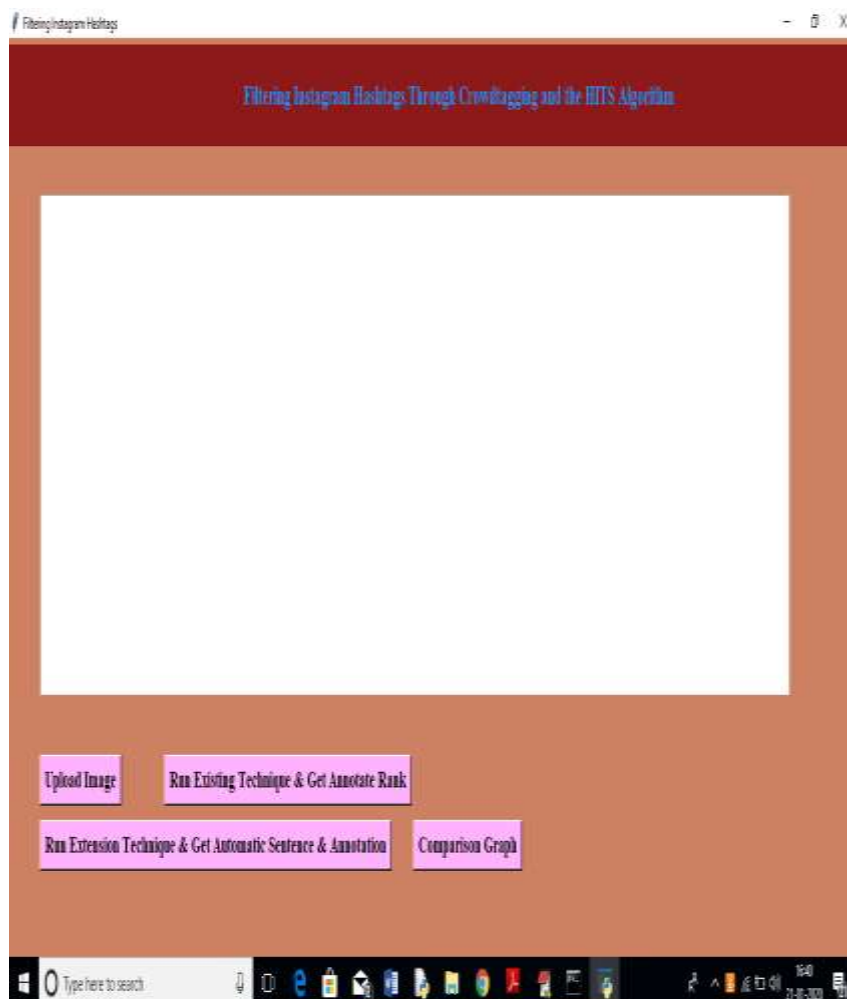
composed by a set of related terms, the identification of the visual topic of an Instagram image, through the proposed method, provides a plausible set of tags to be used in the context of training AIA methods.

PURPOSED WORK

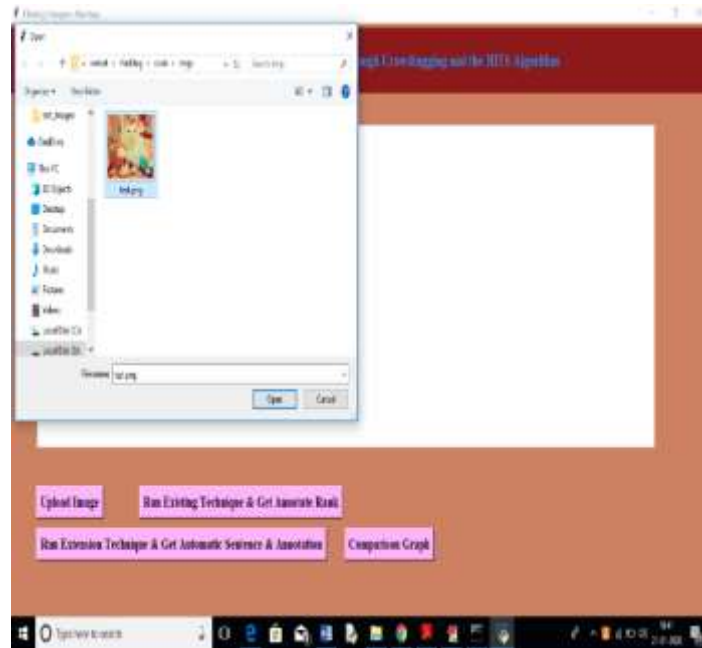
Interaction, content sharing, and collaboration. These media give the users the opportunity to share their content such as text, video, and images [31]. Users usually accompany the content they post with text such as comments or hashtags. This alternative text (comment, hashtags, etc.) provides valuable information about the user posts and other information. Preece et al. [32] construct a Sentinel platform that can enhance social media data in order to understand different situations they based also in Youtube video comments. Sagduyu et al. [33] present a novel system that can present large-scale synthetic data from social media. In their system, they use textual content (hashtags and hyperlinks in tweets) to produce topics and train the n-gram model. The users in several of those media, e.g. Twitter, Instagram, and Facebook, use hashtags to annotate the digital content they upload. Hashtags are, usually, words or nonspaced phrases preceded by the symbol # that allow creators/content contributors to apply tagging that makes it easier for other users to locate their posts. A great portion of the digital content shared on social media platforms consists of images and short videos. Thus, effective retrieval of images from social media and the web, in general, becomes harder and more challenging day by day. Contemporary search engines are basically based on text descriptions to retrieve images; however, inaccurate text descriptions and the plethora of nontextually annotated images led to extended research for content-based image retrieval techniques

Sample Screens

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



In above screen click on 'Upload Image' button to upload image



In above screen I am uploading one image 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 image will get below screen. Now click on 'Run Existing Technique & Get Annotation Rank' button to get below screen

CONCLUSION

In this paper, we have presented an innovative methodology, based on the HITS algorithm and the principles of collective intelligence, for the identification of Instagram hashtags that describe the visual content of the images

they are associated with. We have empirically shown that the application of a two-step HITS algorithm in a crowdtagging context provides an easy and effective way to locate pairs of Instagram images and hashtags that can be used as training sets for content-based image retrieval systems in the learning by example paradigm. As a proof of concept, we have used 25 000 evaluations (500 annotations for each one of 50 images) collected from the Figure-eight crowdsourcing platform to create a bipartite graph composed of users (annotators) and the tags they selected to describe the 50 images. The hub scores of the HITS algorithm applied to this graph, called hereby full bipartite graph, give us a measure of the reliability of the annotators. The aforementioned approach is based on the findings of Theodosiou et al. [39], in which the reliability of annotators is better approximated if we consider all the annotations they have performed rather than the subset of gold test questions. In the second step, a weighted bipartite graph for each image is composed in the same way as the full bipartite graph. The weights of these graphs are the hub scores computed in the previous step. By thresholding the authority scores of the per image graphs, obtained by the application of the HITS algorithm on the weighted graphs, we can rank and then effectively locate the hashtags that are relevant to their visual content as per the annotators evaluation..

References

- [1]. J. Boyle et al., "Predicting emergency department admissions," *Emerg. Med. J.*, vol. 29, pp. 358365, May 2012, doi: 10.1136/emj.2010.103531.
- [2]. S. L. Bernstein et al., "The effect of emergency department crowding on clinically oriented outcomes," *Acad. Emerg. Med.*, vol. 16, no. 1, pp. 110, 2009, doi: 10.1111/j.1553-2712.2008.00295.x.
- [3]. D. M. Fatovich, Y. Nagree, and P. Sprivilis, "Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia," *Emerg. Med. J.*, vol. 22, no. 5, pp. 351354, 2005, doi: 10.1136/emj.2004.018002.
- [4]. M. L. McCarthy et al., "Crowding delays treatment and lengthens emergency department length of stay, even among high-acuity patients," *Ann. Emerg. Med.*, vol. 54, no. 4, pp. 492503, 2009, doi: 10.1016/j.annemergmed.2009.03.006.,