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

**Authors:** [**Argyris Argyrou**](https://ieeexplore.ieee.org/author/37086487579)**; [Stamatios Giannoulakis](https://ieeexplore.ieee.org/author/37086358643);**[**Nicolas Tsapatsoulis**](https://ieeexplore.ieee.org/author/37271573200)

**Abstract:**

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.

**2. Crowdsourcing for multiple-choice question answering**

**Authors: Bahadir Ismail Aydin, Yavuz Selim Yilmaz, Yaliang Li, Qi Li, Jing Gao and Murat Demirbas**

**Abstract:**

We leverage crowd wisdom for multiple-choice question answering, and employ lightweight machine learning techniques to improve the aggregation accuracy of crowdsourced answers to these questions. In order to develop more effective aggregation methods and evaluate them empirically, we developed and deployed a crowdsourced system for playing the “Who wants to be a millionaire?” quiz show. Analyzing our data (which consist of more than 200,000 answers), we find that by just going with the most selected answer in the aggregation, we can answer over 90% of the questions correctly, but the success rate of this technique plunges to 60% for the later/harder questions in the quiz show. To improve the success rates of these later/harder questions, we investigate novel weighted aggregation schemes for aggregating the answers obtained from the crowd. By using weights optimized for reliability of participants (derived from the participants’ confidence), we show that we can pull up the accuracy rate for the harder questions by 15%, and to overall 95% average accuracy. Our results provide a good case for the benefits of applying machine learning techniques for building more accurate crowdsourced question answering systems.

**3. Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing**

**Abstract**

A common challenge with processing naturalistic driving data is that humans may need to categorize great volumes of recorded visual information. By means of the online platform CrowdFlower, we investigated the potential of crowdsourcing to categorize driving scene features (i.e., presence of other road users, straight road segments, etc.) at greater scale than a single person or a small team of researchers would be capable of. In total, 200 workers from 46 different countries participated in 1.5 days. Validity and reliability were examined, both with and without embedding researcher generated control questions via the CrowdFlower mechanism known as Gold Test Questions (GTQs).

By employing GTQs, we found significantly more valid (accurate) and reliable (consistent) identification of driving scene items from external workers. Specifically, at a small scale CrowdFlower Job of 48 three-second video segments, an accuracy (i.e., relative to the ratings of a confederate researcher) of 91% on items was found with GTQs compared to 78% without. A difference in bias was found, where without GTQs, external workers returned more false positives than with GTQs. At a larger scale CrowdFlower Job making exclusive use of GTQs, 12,862 three-second video segments were released for annotation. Infeasible (and self-defeating) to check the accuracy of each at this scale, a random subset of 1012 categorizations was validated and returned similar levels of accuracy (95%).

**4. A survey and analysis on automatic image annotation**

**Abstract:**

In recent years, image annotation has attracted extensive attention due to the explosive growth of image data. With the capability of describing images at the semantic level, image annotation has many applications not only in image analysis and understanding but also in some relative disciplines, such as urban management and biomedical engineering. Because of the inherent weaknesses of manual image annotation, Automatic Image Annotation (AIA) has been raised since the late 1990s. In this paper, a deep review of state-of-the-art AIA methods is presented by synthesizing 138 literatures published during the past two decades. We classify AIA methods into five categories: 1) Generative model-based image annotation, 2) Nearest neighbor-based image annotation, 3) Discriminative model-based image annotation, and 4) Tag completion-based image annotation, 5) Deep Learning-based image annotation. Comparisons of the five types of AIA methods are made on the basis of the underlying idea, main contribution, model framework, computational complexity, computation time, and annotation accuracy. We also give an overview of five publicly available image datasets and four standard evaluation metrics commonly used as benchmarks for evaluating AIA methods. Then the performance of some typical or well-behaved models is assessed based on benchmark dataset and standard evaluation metrics. Finally, we share our viewpoints on the open issues and challenges in AIA as well as research trends in the future.