Disaster Image Filtering and Summarization Based on Multi-layered Affinity Propagation

Yimin Yang, Shu-Ching Chen
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, USA
{yyang010, chens}@cs.fiu.edu

***Abstract*—In this paper, a disaster image filtering and summarization (DIFS) framework is proposed based on multi-layered affinity propagation. The proposed framework is able to automatically identify and summarize latent semantic themes (scenes) in a disaster topic and filter junk images at the same time. Specifically, the images belonging to a disaster topic are first clustered into different groups based on visual descriptors using affinity propagation (AP). Then the typical instances within each cluster are collected to perform the second-layer clustering for identifying final positive clusters by utilizing both visual and textual similarities concurrently. At both layers, the proposed curve fitting function is applied to select appropriate preference values for the AP algorithm. The experimental results on the real world Flickr data set demonstrate the effectiveness of the proposed framework.**

***Keywords*-disaster topics; image filtering; image summarization; multi-layered affinity propagation**

I. Introduction

With the proliferation of mobile devices, emergency responders, supporting agencies and even private citizens are capturing pictures of disaster events as they unfold. This kind of imagery information is of great value for understanding and evaluating disaster situations and therefore supporting efficient decision-making process. However it is a daunting task for emergency managers to collect, organize, and present the imagery data in an efficient and effective way.

The traditional way of accessing and collecting images is keyword-based search, which mostly relies on textual information, such as in Flickr [1]. There are two main problems with the retrieved results using the keyword-based search method. The first one is the well-known semantic gap issue. For example, a query using the keyword "avalanche" may return results both containing images describing the disaster event avalanche or the ones depicting cars with the brand "Avalanche". To the users intended to search for images regarding the topic of disaster "avalanche", the images tagged by the same keyword but with different semantics are considered as junk images, together with the ones mistagged by the users. The other main concern is the lack of organization and summarization of the images within one topic. For example, there may be different themes (scenes), such as building collapse and evacuation, for the keyword "earthquake". Without the well structured and summarized search results, it is difficult to identify those scenes under each topic for efficient browsing.

In our previous work [2], [3], we presented a hierarchical disaster image categorization framework, which classifies images in a supervised manner. In this paper, we focused on the unsupervised filtering and summarization of disaster images collected from Flickr [1]. To solve the aforementioned two problems, we develop a disaster image filtering and summarization (DIFS) framework based on multi-layered affinity propagation (AP) [4]. The proposed framework first clusters the initial collection into visually differentiated groups. Next, the top-ranked instances within each group are selected to build a typical subset of the data, followed by the second layer of clustering using both visual and textual similarities concurrently. Finally, the distribution of the primary clusters will be analyzed to determine the final positive clusters generated in the first layer and filter out the junk images at the same time.

The rest of the paper is organized as follows. Section II briefly reviews the related work on image filtering and summarization. Section III describes the proposed DIFS framework in details. Section IV presents the experimental results and Section V concludes the paper.

II. Related Work

Many pioneer studies have been done for image filtering and summarization respectively. Xie *et al.* [5] propose a K-way min-max cut clustering algorithm for filtering out junk images for Google Image search results. The limitation is that the number of clusters has to be preset, which lacks flexibility and may not match the semantic distribution for an image topic. In [6], the TSI-pLSA method is presented for image categorization based on a visual vocabulary, where the performance heavily relies on the quality of the training data. Wnuk *et al.* [7] propose a nonparametric measure of strangeness based on visual characteristics of images. It neglects the role of textual features in capturing image semantics. Recently, many researchers have proved the effectiveness of AP-based methods in automatic image summarization [8], [9], [10]. Jia *et al.* [8] present a hierarchical affinity propagation approach to image collection summarization based on visual features. Later, the authors incorporate the textual information to update the AP algorithm and build a hybrid image summarization scheme [9]; however, the hybrid AP algorithm does not outperform the original version [4] in general. In [10], Liu *et al.* utilize both the temporally consistent and constrained AP algorithms to select exemplars for performing semi-automatic tagging of photo albums. None of the existing approaches has addressed the image filtering and summarization tasks at the same time automatically.

III. DIFS Framework

A. Visual Similarity Construction

The appropriateness of the similarity matrix greatly affects the performance of image clustering. In this paper, we propose to construct a similarity matrix using visual descriptors, such as Histogram of Oriented Gradient (HOG) [11], color and edge directivity descriptor (CEDD) [12], as well as other low-level visual features, including color histogram, color moment and texture wavelet.

The combination of the above three types of features forms a 707-dimensional feature vector for each image instance. To perform efficient clustering in later stage, the Principle Component Analysis (PCA)-based feature reduction is performed. We keep the top Q feature components having the individual energy distribution larger than a preset threshold.

Finally, the similarity between an image pair $(I_{c,j}, I_{c,k})$ for disaster topic c is represented by the negative square of Euclidean distance as shown below:

$$s(I_{c,j}, I_{c,k}) = -\left\|\overrightarrow{I_{c,j}} - \overrightarrow{I_{c,k}}\right\|^2,\ j \neq k \tag{1}$$

B. First-layer Affinity Propagation

The AP clustering algorithm propagates affinities by passing two types of messages between two data points (images) [4]: the "responsibility" $r(I_{c,j}, I_{c,k})$ sent from image $I_{c,j}$ to image $I_{c,k}$, representing how well $I_{c,k}$ serves as the exemplar of $I_{c,j}$ considering other potential exemplars for $I_{c,j}$; and the "availability" $a(I_{c,j}, I_{c,k})$ sent from image $I_{c,k}$ to image $I_{c,j}$, reflecting how appropriate $I_{c,j}$ chooses $I_{c,k}$ as its exemplar considering other potential images that may choose $I_{c,k}$ as their exemplar. The responsibility and availability are updated iteratively using the following equations:

$$\begin{aligned} r(I_{c,j}, I_{c,k}) \leftarrow & s(I_{c,j}, I_{c,k}) \\ & - \underset{l:l\neq k}{max}(a(I_{c,l}, I_{c,j}) + s(I_{c,j}, I_{c,l})), \end{aligned} \tag{2}$$

$$\begin{aligned} a(I_{c,k}, & I_{c,j}) \leftarrow min(0, \\ & r(I_{c,k}, I_{c,k}) + \sum_{l:l\notin\{k,j\}} max\,\{0, r(I_{c,l}, I_{c,k})\}). \end{aligned} \tag{3}$$

The self-availability is updated as

$$a(I_{c,k}, I_{c,k}) \leftarrow \sum_{l:l\neq k} max\,\{0, r(I_{c,l}, I_{c,k})\}\,. \tag{4}$$

This message reflects an accumulated confidence that image $I_{c,k}$ is an exemplar, based on the positive responsibilities sent to the candidate exemplar k from other images.

Finally, the exemplar for image $I_{c,j}$ is chosen as follows.

$$e^*_{c,j} \leftarrow \underset{I_{c,k}}{argmax}(r(I_{c,j}, I_{c,k}) + a(I_{c,k}, I_{c,j})). \tag{5}$$

C. Textual Similarity Construction

To explore the semantic context within a specific disaster topic, we perform latent semantic analysis utilizing the textual information such as tags, titles, and available descriptions for each image. Specifically, the term-document matrix X is first constructed. The top W words with maximum term frequencies are selected. The standard TF-IDF weight is used to transform the term-document matrix. The term frequency is normalized by log-frequency weighting as follows:

$$w_{t,d} = \begin{cases} 1 + log(TF_{t,d}), & if\ TF_{t,d} > 0 \\ 0, & otherwise \end{cases} \tag{6}$$

where $TF_{t,d}$ and $w_{t,d}$ denote the term frequency and log-frequency of term t in document d respectively. The similarity matrix is built based on cosine measurement shown below:

$$s(D_{c,j}, D_{c,k}) = \frac{\vec{D_{c,j}} \cdot \vec{D_{c,k}}}{\|D_{c,j}\| \cdot \|D_{c,k}\|}, \tag{7}$$

where $D_{c,j}$ and $D_{c,k}$ represent the normalized document vector for image j and image k in disaster topic c respectively. Finally, PCA is applied.

D. Second-layer Affinity Propagation

At the second layer, both visual clustering and textual clustering are performed based on the selected typical instances. Next, the distribution of the primary clusters is analyzed, i.e., to determine which original clusters are included in the primary cluster produced at the second layer. Finally, the intersection of the visual and textual cluster distributions identifies the final positive clusters. Based on our experimental observation, most of the clusters in the first layer are both visually and semantically related to the disaster topic, especially the top-ranked instances (called *typical instances*) within each cluster. Therefore, it is reasonable to expect the *primary cluster* (with the largest number of instances) in the second layer to accumulate most of the relevant instances, which can be used to trace back the relevant clusters (called *positive clusters*) in the first layer. We use the intersection of the identified positive clusters from visual and textual clustering respectively to ensure the pureness and accuracy of the positive clusters. The second layer affinity propagation and filtering procedure is summarized as follows:

Algorithm 1 Second Layer Affinity Propagation

Input: typical instance set A, visual similarity matrix S_V, and textual similarity matrix S_T for all topics.
Output: recognized *positive clusters*.
for each topic c **do**
  procedure SECLAYAP(A^c, S_V^c, S_T^c)
    perform AP clustering based on S_V^c;
    $B_V^c \leftarrow \textit{the primary cluster}$;
    $G_V^c \leftarrow \textit{group IDs in } B_V^c$;
    perform AP clustering based on S_T^c;
    $B_T^c \leftarrow \textit{the primary cluster}$;
    $G_T^c \leftarrow \textit{group IDs in } B_T^c$;
    return $G_V^c \bigcap G_T^c$;
  end procedure
end for

IV. Experimental Results

In this section, the effectiveness of the proposed DIFS framework will be demonstrated from different aspects. First, the relationship between the preference value (i.e., the parameter for AP) and the number of clusters is explored and represented by a curve fitting function for evaluating and selecting a proper input for the AP algorithm; and then the clustering results at the first layer and second layer are presented and analyzed in details respectively. Over 110,000 images as well as their tags and descriptions covering 28 disaster topics are crawled from Flickr [1] as the test set.

A. Preference Selection

The AP algorithm has a heuristic parameter P, called preference, which indicates the preference that an image is chosen as an exemplar. Although the AP algorithm can automatically determine the number of clusters, i.e., Y, based on the P value, there is no explicit relationship between Y and P. Usually, it is suggested to set P as the median similarity (S_{med}) or minimum similarity (S_{min}). However, it is not always a good choice, especially for our image summarization task. To explore the underlying relationship between Y and P, the following experiment is conducted (based on the visual similarity). 100 runs of AP clustering are performed with P values ranging from $10 * S_{min}$ to S_{med} with an equal footstep for each of the 28 disaster topics. The evolution of Y as a function of P is illustrated in Figure 1. The P value is normalized using the scaling factor $1/(10 * S_{min})$ to diminish the effect of different numbers of images in each disaster topic. As shown in the figure, the P−Y curves follow a similar pattern, i.e., Y is almost monotonically increasing with P polynomially. Therefore, we use the least-square fitting method to capture the relationship between P and Y, where Y is expressed as a polynomial function for P as shown below. The fitting curve is highlighted in red-dot circles.

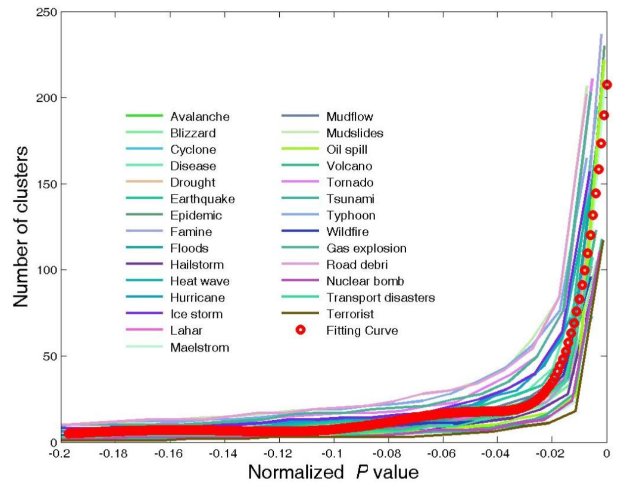

Figure 1. Number of clusters (Y) as a function of the preference (P) values.

$$Y = a_n P^n + a_{n-1} P^{n-1} + ... + a_1 P^1 + a_0 P^0 = \sum_{i=0}^{n} a_i P^i \quad (8)$$

It is worth noting that the fitting function is similarity sensitive, i.e., different similarity matrices may adapt to distinct fitting functions. For example, the visual and textual similarity matrices in our framework may result in two versions of fitting functions. Furthermore, extra (P, Y) points may be added to better approximate the curve near S_{min}. Once the P–Y curve fitting functions are constructed, we may estimate and select the P values without actually running the AP clustering algorithm as done in most existing approaches.

B. First-layer Clustering Results

Figure 2 illustrates the first-layer clustering results for the disaster topic “avalanche”. Specifically, it shows the exemplars together with the top 3 images ranked by similarity within each cluster when the number of clusters reaches 16. As implied from the figure, the AP clustering procedure reasonably captures the distribution of image instances in the feature space. The clustering results are satisfactory in the sense that different clusters depict distinct scenes (possibly different events) related to the disaster topic; those relevant clusters are defined as *positive clusters* (see IV-C) to be identified in the second layer. It is worth noting that there exists some non-relevant clusters, which are to be filtered. In our experiments, we also discard the clusters with too few instances, i.e., less than 5.

C. Second-layer Clustering and Filtering Results

By performing the second-layer clustering, we can identify most of the positive clusters generated in the first layer and filter out the non-relevant clusters. Specifically, for the disaster topic “avalanche”, 5 out of 6 true positive clusters

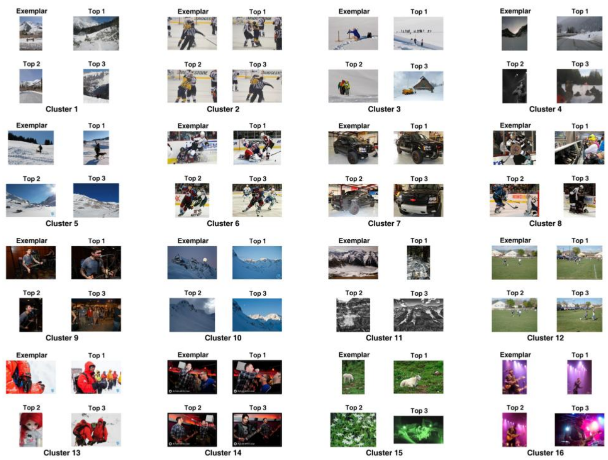

Figure 2. Clustering results for the disaster topic “avalanche” with 16 clusters. There are four images in each cluster, where the top-left one is the exemplar and the rest are the top three images ranked by similarity.

are identified with one false positive cluster. To further investigate the distribution and filtering of the non-relevant instances within each cluster, we perform the average precision analysis for each recognized positive cluster as shown in Table I, where the first column lists the positive clusters and columns 2 through 7 display the average precisions with top $T\%$ of instances in a descending similarity order. The last row calculates the mean average precisions (MAP) for all positive clusters. As indicated by the evaluation results, the positive instances dominant over 90% of the positive clusters, demonstrating the relative accuracy of the clustering results. Finally, we select the top 4 images (including the exemplar) in each positive cluster as the summarization results, and filter out the last 30% instances considered as junk images to further improve the pureness.

Table I
MEAN AVERAGE PRECISION FOR TOPIC “AVALANCHE”.

Cluster ID	**Top 10%**	**Top 30%**	**Top 50%**	**Top 70%**	**Top 90%**	**All**
1	1.000	0.991	0.968	0.948	0.933	0.928
3	1.000	0.998	0.967	0.947	0.934	0.930
5	0.982	0.860	0.840	0.829	0.820	0.816
10	1.000	0.995	0.964	0.950	0.929	0.923
11	1.000	1.000	1.000	0.993	0.975	0.966
MAP	0.996	0.969	0.948	0.933	0.918	0.912

V. CONCLUSION

In this paper, we proposed a multi-layered DIFS framework, where AP was first applied to the original image collections to build the initial clusters for each disaster topic; then both the visual similarity and textual similarity were utilized in the second layer to identify the positive clusters and filter out junk images. We also presented the curve fitting method for selecting the P value appropriately. In the future, we will further investigate the general relationship between the preference value and the number of clusters under various similarity construction strategies, and study the interrelationship between visual and textual similarities to refine the clustering results.

ACKNOWLEDGMENT

This research was supported by the U.S. Department of Homeland Security under grant Award Number 2010-ST-062-000039, the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001, and NSF HRD-0833093.

REFERENCES

[1] http://www.flickr.com/

[2] Y. Yang, H.-Y. Ha, F. C. Fleites, S.-C. Chen, and S. Luis, “Hierarchical disaster image classification for situation report enhancement,” in *IEEE International Conference on Information Integration and Reuse*, pp. 3-5, 2011.

[3] S. Luis, F. C. Fleites, Y. Yang, H.-Y. Ha, and S.-C. Chen, “A visual analytics multimedia mobile system for emergency response,” in *IEEE International Symposium on Multimedia*, pp. 337-338, 2011.

[4] B. J. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, 315:972-976, 2007.

[5] F. Xie, Y. Shen, and X. He, “K-way min-max cut for image clustering and junk images filtering from Google image,” in *ACM Multimedia*, pp. 803-806, 2010.

[6] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, “Learning object categories from google's image search,” in *IEEE International Conference on Computer Vision*, pp. 1816-1823, 2005.

[7] K. Wnuk and S. Soatto, “Filtering Internet image search results towards keyword based category recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[8] Y. Jia, J. Wang, C. Zhang, and X.-S. Hua, “Finding image exemplars using fast sparse affinity propagation,” in *ACM Multimedia*, pp. 639-642, 2008.

[9] H. Xu, J. Wang, X.-S. Hua, and S. Li, “Hybrid image summarization,” in *ACM Multimedia*, pp. 1217-1220, 2011.

[10] D. Liu, M. Wang, X.-S. Hua, and H.-J. Zhang, “Semi-automatic tagging of photo albums via exemplar selection and tag inference,” *IEEE Trans. on Multimedia*, 13:82-91, 2011.

[11] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.

[12] S. Chatzichristofis and Y. Boutalis, “Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval,” in *IEEE International Conference on Computer Vision*, pp. 312-322, 2008.