A Classification Based Framework For Concept Summarization

Dhruv Mahajan S Sundararajan Subhajit Sanyal Amit Madaan
Yahoo! Labs, Bangalore, India
{dkm,ssrajan,subhajit,amitm}@yahoo-inc.com

ABSTRACT

In this paper we propose a novel classification based framework for finding a small number of images summarizing a concept. Our method exploits metadata information available with the images to get the category information using Latent Dirichlet Allocation. We modify the import vector machine formulation based on kernel logistic regression to solve the underlying classification problem. We show that the import vectors provide a good summary satisfying important properties such as coverage, diversity and balance. Furthermore, the framework allows users to specify desired distributions over category, time etc, that a summary should satisfy. Experimental results show that the proposed method performs better than state-of-the-art summarization methods in terms of satisfying important visual and semantic properties.

Categories and Subject Descriptors: I.4m [Computing Methodologies]: Image Processing and Computer Vision - Miscellaneous
General Terms: Algorithms, Design
Keywords: Concept and Image Summarization

1. INTRODUCTION

In this work, we focus on the *problem of automatic summarization of images for a given concept* (e.g.,Oil Spill) where additional information is available in the form of metadata. The goal is to discover various aspects (e.g.,*environmental impact*, *protests*, *politics*) of the concept and present a set of selected images according to some criteria [3, 2, 1] like likelihood, diversity, balance in terms of both visual and semantic sense.

For image collections, most of the works target applications like presentation of iconic views, image summarization or browsing [3, 1], ranking [2], identification and summarization of landmarks. In contrast, our work is suitable for a wide variety of concepts, made possible by discovering topics through Latent Dirichlet Allocation (LDA). It approaches the problem through a novel classification based strategy rather than the conventional clustering based approach. Furthermore, it is *still unsupervised* because we automatically get the category information of the images through LDA. Also, our method satisfies most of the important properties like likelihood, diversity, balance both visually and semantically, while the other approaches satisfy only a subset of them. Finally, additional user specified properties like preferred topic and temporal distributions can be incorporated more naturally in our framework, and have never been attempted before to the best of our knowledge.

2. OUR APPROACH

Concept Summarization: Effective summarization of a collection is characterized by some important properties that the summary should possess. Based on the properties used in [3, 1, 2], we use the following properties for both semantic and visual aspects. They are: (1) **Diversity -** Two images in the summary should not be similar to each other visually or semantically. (2) **Coverage -** The summary should cover all interesting and important visual and semantic aspects. Visual and semantic aspects with high likelihood should be present in the summary. (3) **Balance -** The various visual and semantic aspects should be present in a balanced way to avoid any misunderstanding of summary. Besides this, the summary should be sparse in the number of images.

Classification Framework: The motivation behind our approach is based on two observations: (1) description of an image often gives important information about the semantic aspects (topics) and (2) image features in a good representation are useful to predict the semantic topics. Then, the basic idea is to build a classifier model (using the image features) specified by a subset of images that predict the semantic aspects well.

Let $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$ and $\mathcal{T} = \{T_1, T_2, \ldots, T_N\}$ denote a collection of images with respective descriptions. Let us assume that $\mathcal{X} = \{\mathbf{x}_i : i = 1, \ldots, N\}$ is a feature representation of $\mathcal{I}$. We discover the semantic topics from $\mathcal{T}$ using LDA. Let us assume that there are M topics. Let us denote the topic distributions of the collection $\mathcal{I}$ discovered by LDA as: $\mathcal{Q} = \{\mathbf{q}_i : i = 1, \ldots, N\}$ where $\mathbf{q}_i = [q_{i,1}, \ldots, q_{i,M}]$ is the topic distribution of i-th image. Usually, each image belongs to very few topics and in the extreme case it belongs to just one topic. Therefore, each $\mathbf{q}_i$ is sparse. Our idea is to treat each topic as a class and $\mathbf{q}_i$ as the class probability distribution over a set of classes $C = \{c_1, c_2, \ldots, c_M\}$, i.e., $P(c_j|\mathbf{x}_i) = q_{i,j}, \forall i = 1, \ldots, N, j = 1, \ldots, M$. Thus, the problem is: *given the image-class distribution pairs for all the examples, find a sparse set of images* $\mathcal{I}_{\mathbf{S}} \subset \mathcal{I}$ *that summarizes the concept for some* $\mathbf{S} \subset \{1, \ldots, N\}$ *with a user specified value for* $|\mathbf{S}|$, *where* $|\mathbf{S}| \ll N$. Our proposal is to design a sparse kernel classifier and output the concept summary comprising of the image summary $\mathcal{I}_{\mathbf{S}} = \{I_i : i \in \mathbf{S}\}$ (that specifies the sparse kernel classifier) with its descriptions $\mathcal{T}_{\mathbf{S}} = \{T_i : i \in \mathbf{S}\}$. We use the import vector machine [4] as the sparse kernel classifier.

Import Vector Machine (IVM): Let $\mathbf{y}_i$ be a binary vector with only one non-zero element, where the components are defined as: $y_{i,m}, m \in 1, 2, \ldots, M$ and let c_i denote the class label of the i-th image. Then, the IVM optimization problem for a multi-class classification problem can be written as [4]: $\min_{\mathbf{a},\mathbf{S}} -\frac{1}{N}\sum_{i=1}^{N} y_{i,c_i} \log(P(c_j|\mathbf{x}_i)) + \frac{\lambda}{2}\sum_{m=1}^{M} \mathbf{a}_{:,m}^T \mathbf{K}\mathbf{a}_{:,m}$, where, $P(c_j|\mathbf{x}) = \frac{e^{f_j(\mathbf{x})}}{\sum_{m=1}^{M} e^{f_m(\mathbf{x})}}$, $f_m(\mathbf{x}) = \sum_{i\in\mathbf{S}} a_{i,m} k(\mathbf{x}, \mathbf{x}_i)$, (i, j)-th entry of the

WWW 2011, March 28–April 1, 2011, Hyderabad, India.
ACM 978-1-4503-0637-9/11/03.

kernel matrix $\mathbf{K}$ is given by $k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\mathbf{x}, \mathbf{x}_i) = \exp(-\frac{||\mathbf{x}-\mathbf{x}_i||^2}{2\sigma^2})$ (σ^2 is the kernel width parameter). $\mathbf{a}_{:,m}$ denotes the coefficient vector of the m-th classifier, $f_m(\mathbf{x})$. In IVM, the examples in $\mathbf{S}$ are called the import vectors. Thus, we design an IVM to find the concept summary $(\mathcal{I}_\mathbf{S}, \mathcal{T}_\mathbf{S})$. In our general setting, we have the probability distribution $\mathbf{q}_i$ (instead of $\mathbf{y}_i$) where more than one class can have nonzero values. Thus, we have:

$$\min_{\mathbf{S},\mathbf{a}} -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} q_{i,j}\log(P(c_j|\mathbf{x}_i)) + \frac{\lambda}{2}\sum_{m=1}^{M}\mathbf{a}_{:,m}^T\mathbf{K}\mathbf{a}_{:,m}. \quad (1)$$

Note that (1) is a combinatorial optimization problem. Therefore, similar to [4] we use a greedy algorithm (Algorithm 1) to find $\mathbf{S}$.

```
Data: Class distributions, $\{q_{i,j}, i = 1,..,N, j = 1,..,M\}$
      Image features $\mathcal{X} = \{\mathbf{x}_i, i = 1,..,N\}$
      Distribution over categories $P_t$, if applicable
      Number of Summary Images $L = |\mathbf{S}|$, parameters $\lambda$ and $\eta$
Result: Subset $\mathbf{S}$ of indexes of the summary images
begin
    $\mathbf{S} \longleftarrow \emptyset$
    for $k \leftarrow 1$ to $L$ do
        $\bar{\mathbf{S}} \longleftarrow \{1, ..., N\}\backslash\mathbf{S}$
        for $i \in \bar{\mathbf{S}}$ do
            Construct $\mathbf{S}_i \longleftarrow \mathbf{S} \cup \{i\}$.
            Set $\mathbf{S}$ to $\mathbf{S}_i$ in Equation 1 and optimize for the
            coefficients $\mathbf{a}$ using gradient descent algorithm
            $E_i \leftarrow$ optimized objective function value (Eq. 1)
            if Distribution constraint is applicable then
                $E_i \longleftarrow E_i + \eta KL(P_t, \frac{1}{k}\sum_{\mathbf{x}_l \in \mathbf{S}_i}\mathbf{q}_l)$.
            end
        end
        $\mathbf{S} \longleftarrow \mathbf{S} \cup \{argmin_{i\in\bar{\mathbf{S}}} E_i\}$.
    end
end
```

Algorithm 1: A Greedy Algorithm for Subset Selection

Distribution Regularization: Our framework has the flexibility in selecting the subset according to some user defined requirements. For example, for the concept Oil Spill, it is appropriate to include more images that represent *political* aspects than others while preparing a slide show for a web page discussing *political* news. Let $\mathbf{P}_t$ denote a target distribution over the categories that the summary should satisfy; that is, we want $P_t(j) = \frac{1}{K}\sum_{i\in\mathbf{S}}\mathbf{q}_{i,j}, \forall j = 1, \ldots, M$. We *add* a KL-divergence based regularization term ηKL $(P_t(c), \frac{1}{K}\sum_{i\in S}\mathbf{q}_i)$ to the objective function (Equation 1). Here, η controls the contribution of this term.

System Implementation: For data acquisition we obtained the collections from *flickr*, internal search indices of mRSS feeds from sites like `http://news.yahoo.com`, `http://omg.yahoo.com`, etc. For each image I_i, we extract a 1024 dimensional feature vector $\mathbf{x}_i$ using convolution neural network (CNN). For text, we use a bag of word representation with standard preprocessing. We derive the topic distribution information $\mathcal{Q}$ automatically using LDA with the parameter values $\alpha = 0.1$ and $\beta = 0.01$. For IVM, we use the Gaussian kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2})$ with σ in the range $2-3$. We set λ to a value in the range $[0.001 - 0.0001]$.

3. EXPERIMENTAL EVALUATION

Figure 1 shows a summary of 20 images for the San Francisco dataset (4602 images collected from *Flickr*) with the uniform category distribution constraint. Notice that our method is able to capture the different aspects well. For example, there are two different views of *golden Gate Bridge* (blue box) in the summary. Moreover, the images of different neighborhoods like *Victorian Houses* and *Chinatown* are also present. We compared our method with the methods proposed in [3, 1, 2]. We observed that each of these methods miss out at least one of these aspects.

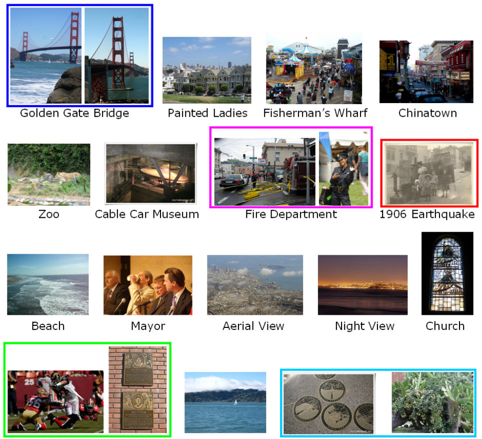

Figure 1: **Our summary for the concept:** San Francisco.

Quantitative Comparison: We also made a quantitative comparison of different methods (Table 2). Table 1 shows the metrics used for different visual and semantic properties.

Visual Likelihood (VL)	$\sum_{i\in\mathbf{S}}\sum_{j=1}^{N} k(\mathbf{x}_i, \mathbf{x}_j) - \lvert\mathbf{S}\rvert$
Visual Diversity (VD)	$\sum_{i\in\mathbf{S}}\sum_{j\in\mathbf{S}} k(\mathbf{x}_i, \mathbf{x}_j) - \lvert\mathbf{S}\rvert$
Semantic Likelihood (SL)	$-\sum_{i\in\mathbf{S}}\sum_{j=1}^{N} KL(\mathbf{q}_i, \mathbf{q}_j)$
Semantic Diversity (SD)	$-\sum_{i\in\mathbf{S}}\sum_{j\in\mathbf{S}} KL(\mathbf{q}_i, \mathbf{q}_j)$
Semantic Balance (SB)	$KL(\frac{1}{N}\sum_{i=1}^{N}\mathbf{q}_\mathbf{i}, \frac{1}{K}\sum_{i\in\mathbf{S}}\mathbf{q}_\mathbf{i})$

Table 1: **Metrics used for different properties.**

In this experiment, we used 10 popular image search queries from three broad types of concepts *Current Affairs*, *Travel* and *Celebrities*. The average number of images was around 400. We computed the metrics in Table 1 for each query and method, and ranked them on each property. The results reported in Table 2 are the ranks averaged over these queries. Columns 2-4 in Table 2 show the (query) average rank of different methods on each property. The last column shows an average rank (row wise average score) for each method. *Note that we compared only methods that use* **both** *image features and metadata* in this experiment; therefore, we excluded Simon et al. [3] method. Our method performs significantly better than the second best method. Note that other methods are able to do well only on a subset of properties.

	VL	VD	SL	SD	SB	Avg.
Our Method	2.0	2.0	1.6	1.9	1.7	**1.84**
Eva et al [2]	3.0	1.0	1.7	3.0	1.5	2.04
Spec. Clustering [1]	1.0	3.0	2.7	1.1	2.8	2.12

Table 2: **Ranks of methods for different properties.**

4. REFERENCES

[1] J. Fan, Y. Gao, H. Luo, D.A. Keim, and Z. Li. A novel approach to enable semantic and visual image summarization for exploratory image search. In *MIR*, 2008.

[2] Eva Hörster, Malcolm Slaney, Marc'Aurelio Ranzato, and Kilian Weinberger. Unsupervised image ranking. In *LS-MMRM*, 2009.

[3] Ian Simon, Noah Snavely, and Steven M. Seitz. Scene summarization for online image collections. In *ICCV*, 2007.

[4] J. Zhu and T. Hastie. Kernel logistic regression and the import vector machine. *JCGS*, 14:185–205, 2005.