21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Utilizing Co-occurrence Patterns for Semantic Concept Detection in Images

Linan Feng and Bir Bhanu
Center for Research in Intelligent Systems, University of California, Riverside, USA
{fengl, bhanu}@cris.ucr.edu

Abstract

Semantic concept detection is an important open problem in concept-based image understanding. In this paper, we develop a method inspired by social network analysis to solve the semantic concept detection problem. The novel idea proposed is the detection and utilization of concept co-occurrence patterns as contextual clues for improving individual concept detection. We detect the patterns as hierarchical communities by graph modularity optimization in a network with nodes and edges representing individual concepts and co-occurrence relationships. We evaluate the effect of detected co-occurrence patterns in the application scenario of automatic image annotation. Experimental results on SUN'09 and OSR datasets demonstrate our approach achieves significant improvements over popular baselines.

1. Introduction

Semantic concept detection is the fundamental step for many image based applications, such as automated annotation [1] and semantic retrieval [2]. However, approaches based on visual-semantic correlations only have limited effects due to the well-known *semantic gap* problem [3]. Recent research efforts have demonstrated that using correlations between concept as contextual clues can help narrow the semantic gap [4]. Unlike other correlations may measure the semantic meaning differences (e.g., synonymy, meronymy), co-occurrence measures the frequency of concurrent appearance has emerged as an important topic in the research community [5].

To the best of our knowledge, this paper presents the first methodology attempting to explore the semantic co-occurrence patterns from a social network analysis point of view. One of the most important feature of social network is the community structure, i.e. the group of nodes in clusters with large edge weights and dense connections joining nodes of the same community and loose connections to the other communities.

We propose to detect the concept co-occurrence patterns as hierarchical communities in a similar network structure with nodes representing the individual concepts and weighted edges denoting the significances of co-occurrence correlations between concepts. Our intuition behind this approach is based on the fact that concepts in a pattern tend to have denser connections to each other than the connections to the other patterns. These co-occurrence patterns or communities, play similar roles like underlying scene concepts at a higher level of semantics, which are more informative than individual concepts for concept learning.

2. Approach

2.1. Co-occurrence Network Construction

In order to detect the concept co-occurrence patterns, we first select concepts from a given vocabulary as nodes in the co-occurrence network. We build up the vocabulary by including all the labels in the dataset. Then the appearance frequency of concept is calculated. The concepts have frequency lower than 1% of the total number of the images are eliminated. We further exclude the tags that are too general (e.g. "mammal", "tool", "geological formation") or too specific (e.g. "Phoenix tree", "Davy Jones") in semantics which are less relevant to Folksonomy-style tags used in daily life (e.g. Flickr tags) and remove the abstract concepts (e.g. "commencement", "Olympic Games").

We model the remaining concepts as nodes in the network. The connecting edge and the relative weight denote the co-occurrence relationship and its significance. We measure two types of co-occurrences between concepts, namely the *semantic co-occurrence* and *visual co-occurrence* which are calculated by normalized Google distance [6] based on the web page counts and tag distance [7] using the images in the dataset. **Algorithm 1** for constructing the co-occurrence network is described below. Figure 1

978-4-9906441-1-6 ©2012 IAPR

shows the co-occurrence network with line thickness representing the significance of co-occurrence.

Algorithm 1: Co-occurrence network construction

Symbol	Description
G(c)	The number of pages containing concept c reported by Google search engine
G(c_1, c_2)	The number of pages containing both c_1 and c_2
Ω	The number of pages indexed by Google
T(c)	The number of images containing concept c in Flickr
T(c_1, c_2)	The number of images containing both c_1 and c_2 in Flickr
Ψ	The number of images indexed by Flickr

1. Initialize a $N \times N$ concept adjacency matrix A for recording edge weights with every element setting to 0, N is the size of the vocabulary.

2. Measure the semantic co-occurrence between each pair of the concepts $\{c_i, \{c_j\}, i \in 1, ..., N, j \neq i$ by
$\varphi_G(c_i, c_j) = exp(-\frac{max\{logG(c_i), logG(c_j)\} - logG(c_i, c_j)}{log\Omega - min\{logG(c_i), logG(c_j)\}})$

3. Measure the visual co-occurrence by
$\varphi_T(c_i, c_j) = exp(-\frac{max\{logT(c_i), logT(c_j)\} - logT(c_i, c_j)}{log\Psi - min\{logT(c_i), logT(c_j)\}})$

4. Combine the two measures into the final co-occurrence significance by
$A(c_i, c_j) = \lambda \cdot \varphi_G(c_i, c_j) + (1 - \lambda) \cdot \varphi_T(c_i, c_j)$. In our setting we put equal importance on the two measurements, so $\lambda = 0.5$

5. Traverse all the elements in A, add c_i as node in the network, connect two nodes c_i, c_j with edge weight according to the value of A_{ij}.

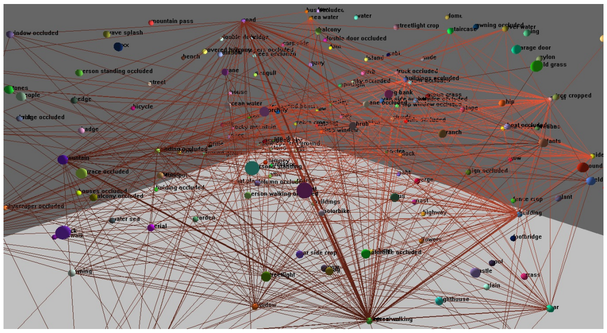

Figure 1. Three-dimension illustration of the co-occurrence network generated from SUN'09 dataset.

2.2. Co-occurrence Pattern Detection

Finding the co-occurrence patterns of interconnected nodes corresponds to uncover organizations from randomness of the topology which is close to graph clustering or partition. We propose a method based on Newman-Girvan modularity [8] optimization. The *modularity* measures the strength of a community or partition by comparing the density of links inside a community with the links to the other communities. Usually high values of modularity suggests good partitions. In the case of weighted network, we define the modularity of community C as:

$$Q_C = \frac{1}{2T}\sum_{i,j}[A_{ij} - \frac{w_i w_j}{2T}]\delta(ID_i, ID_j) \qquad (1)$$

The value of Q is in the range of $[-1, 1]$, and in practice a value greater than 0.3 indicates a significant community. The modularity is calculated over all the pairs of nodes in the network, where T equals half of the summation of all the edge weights in the adjacency matrix. A_{ij} represents the edge weight between node i and j, w_i (w_j) equals the summation of the weights of the edges attached to node i (j), ID_i and ID_j are their community IDs, $\delta(ID_i, ID_j) = 1$ if $ID_i = ID_j$, otherwise $= 0$. We consider iteratively merging the nodes into a hierarchical community structure with different levels of resolution by maximizing the modularity gain in each iteration. The *modularity gain* of moving an outside node i into a community C is evaluated by:

$$\begin{aligned}\Delta Q = [\frac{\Sigma_{in} + w_{i,C}}{2T} - (\frac{\Sigma_{out} + w_i}{2T})^2] - \\ [\frac{\Sigma_{in}}{2T} - (\frac{\Sigma_{out}}{2T})^2 - (\frac{w_i}{2T})^2]\end{aligned} \qquad (2)$$

where Σ_{in} represents the sum of edge weights inside C, $w_{i,C}$ equals the sum of weights of edges that link i to C, T is the same as defined in equation (1), Σ_{out} is the sum of weights of edges that link outside nodes to nodes in C, w_i is the sum of weights of the edges incident to i. Based on the modularity, we propose **Algorithm 2** for detecting the co-occurrence patterns.

As most of the existing methods consider only non-overlapping communities, we expand our algorithm to address the share of nodes between communities. The result of the proposed algorithm working on SUN'09 dataset is in shown in Fig 2. We observe different levels of communities in Fig 2. The highest level of the hierarchy shows two discriminative patterns as "outdoor" and "indoor" scenes indicated by the rightmost horizontal lines. We also observe that individual concepts "person" and "flower" are copied and split into two patterns which hints at overlaps between communities.

2.3. Concept Detection Refinement

We integrate the detected concept co-occurrence patterns into a individual concept detection framework proposed in previous work [8] for image retrieval. In this paper, we evaluate the effect of concept detection in the setting of automated image annotation, there is no overlap between [8] and this paper. In the framework, we use a probabilistic inference model to build the correspondence between semantic concepts and regional visual features. The output of the model is a vector of concepts with relevant probabilistic scores. The concepts have highest scores are used as annotations.

The co-occurrence patterns are utilized for refining the annotations by performing a random walk process over the hierarchical community structure. Suppose the hierarchy has L levels. We measure the distance between concept c_i and c_j as $d_{ij} = l_{ij}/L$, where l_{ij} is the level of the common ancestor of c_i and c_j in the hierarchy. Suppose initially the concept c_i has the probabilistic score $S(c_i)$ given by the inference model, in the k-iteration the score is formulated by the random walk process based on:

$$S_k(c_i) = \alpha \sum_j S_{k-1}(c_j) \cdot d_{ij} + (1 - \alpha) \cdot S(c_i) \quad (3)$$

where α is a weight parameter that belongs to $(0, 1)$. The above formula can strengthen the concepts in closely related communities in the hierarchy and weaken the isolated ones.

Algorithm 2: Finding co-occurrence patterns

```
Partitioning phase:
  1. Assign each node a different community tag $C_i, i = 1, ..., N$.
  2. For each node $V_i$, remove it from its original community $C_i$ and add it into each of its neighboring nodes $V_j$'s community $C_j, j = 1, ...n$.
  2.1 If placing $V_i$ from $C_i$ to $C_j$ produces a positive maximum modularity gain in equation (2), examine the value of $Q_{C_i}$ and $Q_{C_j}$ with $V_i$ assigned to each community by equation (1).
      2.1.1 If both $Q_{C_i}$ and $Q_{C_j}$ are $\geq$ 0.3 which imply a potential share of individual concept between scenes, split node $V_i$ into $V_i$ and $V_i^{'}$ and put into $C_i$ and $C_j$ separately, the edges incident to other nodes are copied between them.
      2.1.2 Else place node $V_i$ into $C_j$.
  2.2 Otherwise, no node will be moved.
3. The first phase stops when every node is traversed and no further improvement can be achieved.
Coarsening phase:
  1. Replace each of the uncovered communities by a single node and replace the edges between communities by a single edge with summed edge weights. 2. Also represent the edges in the same community as a self-looped edge with weight equaling to the sum of the weights of the inside edges.
Iteration:
  Repeat above two phases until no modularity gain given by eq.(2) can be achieved.
```

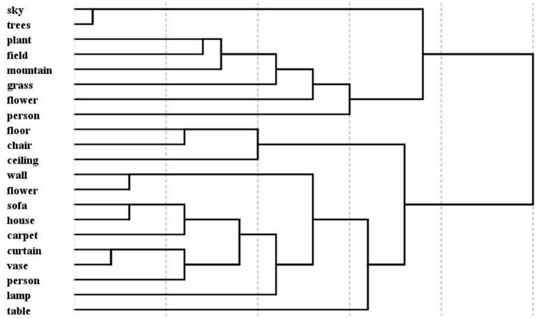

Figure 2. Part of the hierarchical community structure detected for SUN'09 dataset.

3. Experiment Evaluation

Datasets: 1) **Scene Understanding (SUN'09)** [3] dataset contains 12,000 images and more than 5,800 individual concepts covering a variety of indoor and outdoor scene categories. The total number of annotated label is 85,456 which results in average 7 labels/image. SUN'09 images are collected from multiple sources (Google, Flickr, Altavista, LabelMe) and are labeled by single annotator using LabelMe annotation tool. The labels are manually verified for consistency. 2) **Outdoor Scene Recognition (OSR)** [9] dataset has 2,682 images with 520 individual concepts across 8 outdoor scene categories including: coast, forest, highway, inside-city, mountain, open-country, street, tall-building. All the images are in the same resolution with concepts labeled with bounding boxes manually.

Visual Features: We select the same features as in [8].

Baseline Approaches: We compare the automatic annotation accuracy with two baseline approaches: 1) CRM [10], a probabilistic model based on automatic segmentation that makes no assumption about the correspondence between concepts and regions, and 2) MBRM [11], a multiple-Bernoulli model for inferring the degree of presence of concepts. We refer our model as region-based concept detection with co-occurrence patterns, or RCD-CP.

Evaluation Criteria: We measure 1) the overall top-5 annotation accuracy over all the images with different training set sizes and 2) the individual concept detection accuracy for each concept category.

Experimental Results: We split the dataset for training and testing in different sizes as shown in Table 1. We compare the average annotation accuracy over all the images in the datasets, Table 1 shows the results. Our approach shows great improvements over both of the baseline models. And we obtain significant leap of overall accuracy when the training data size exceeds

	SUN'09 Dataset				OSR Dataset			
Methods / % of train	20%	40%	60%	80%	40%	55%	70%	85%
CRM [10]	13.17	15.62	22.01	28.09	15.96	21.75	25.74	29.72
MBRM [11]	14.54	16.71	23.89	29.76	16.03	22.46	26.05	30.61
RCD-CP (Ours)	18.82	19.53	26.62	33.49	19.94	25.61	29.01	34.08
% gain over CRM	42.90%	25.03%	20.95%	19.22%	24.94%	17.75%	12.70%	14.67%
% gain over MBRM	29.44%	20.10%	11.43%	12.53%	24.39%	14.02%	11.36%	11.34%

Table 1. Averaged overall concept detection accuracy as a function of training set size.

the testing data size for both of the datasets. The last two rows conclude % gains of approach over the baselines. Our improvements is meaningful since our approach considers modeling the relationship between semantic concepts and provides more contextual information. The impact of training set size is clear and consistent. More training data results in more accurate annotation models for all the three approaches. However, our approach can have a significant performance increase even when the training set size is small (e.g. 20% for SUN'09 and 40% for OSR).

For individual concept detection, we evaluate the performance by detection accuracy which equals to the number of correctly annotated images in the concept category divided by the number of images containing the concept in the dataset. The results for OSR dataset are shown in Fig 3. Clearly, our RCD-CP model which considers co-occurrences as contextual information has the best performance among the methods. The detection results of some of the concept categories (e.g. "sign", "sidewalk", "skyscraper") is less accurate compared to other concepts from all the three methods. It is understandable since these semantic concepts have larger visual variations than the other concepts. However, it is interesting to see that our approach brings a performance gain over all the concepts with large visual variabilities. This is because the random walk process on the co-occurrence patterns helps boost the detection of difficult concepts from the easy ones based on the fact that they co-occur frequently.

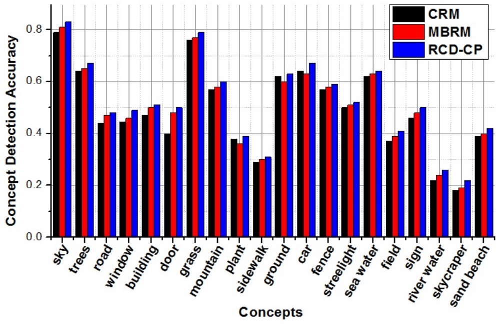

Figure 3. Individual concept detection accuracy of 20 concepts evaluated on OSR.

4. Conclusions

In this paper, we presented a novel approach for detecting concept co-occurrence patterns based on hierarchical community structure in networks. We also proposed a random walk process based approach to integrate the co-occurrence patterns into the automatic image annotation framework. Experimental results on OSR and SUN09 datasets show that the proposed approach has significant performance improvement in concept detection.

5 Acknowledgment

This work was supported in part by NSF grants 0641076 and 0727129.

References

[1] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos. Supervised learning of semantic classes for image annotation and retrieval. *PAMI*, 29(3):394410, 2007.

[2] J. Deng, A. Berg, and F. Li. Hierarchical Semantic Indexing for Large Scale Image Retrieval. In *CVPR* 2011.

[3] C. Dorai and S. Venkatesh. Bridging the Semantic Gap with Computational Media Aesthetics. IEEE *MultiMedia*, 10(2):15-17, 2003.

[4] M. J. Choi, J. J. Lim, A. Torralba, and A. S. Willsky. Exploiting hierarchical context on a large database of object categories. In *CVPR* 2010.

[5] J. Yuan, M. Yang, and Y. Wu. Mining Discriminative Co-occurrence Patterns for Visual Recognition. In *CVPR* 2011.

[6] R. Cilibrasi and P. Vitanyi. The Google Similarity Distance. *KDE*, 19(3):370-383, 2008.

[7] D. Liu, X. Hua, L. Yang, M. Wang and H. Zhang. Tag Ranking. In Proc *WWW* 2009.

[8] L. Feng and B. Bhanu. Concept learning with co-occurrence network for image retrieval. In *ISM* 2011.

[9] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 42(3):145-175, 2001.

[10] V. Lavrenko, R. Manmatha and J. Jeon. A model for learning the semantics of pictures. In Proc *NIPS* 2003.

[11] S. L. Feng, R. Manmatha and V. Lavrenko. Multiple Bernoulli relevance models for image and video annotation. In *CVPR* 2004.