



## LEARNING VISUAL SEMANTIC RELATIONSHIPS FOR PROFICIENT VISUAL RECOVERY

S. Vimalananthi

Assistant Professor, Department of Computer Science and Engineering, SSM College of Engineering, Komarapalayam, Tamilnadu

**Cite This Article:** S. Vimalananthi, "Learning Visual Semantic Relationships for Proficient Visual Recovery", International Journal of Computational Research and Development, Volume 2, Issue 1, Page Number 51-54, 2017.

### Abstract:

In this paper, I investigate how to establish the relationship between semantic concepts based on the large-scale real-world click data from image commercial engine, which is a challenging topic because the click data suffers from the noise such as typos, the same concept with different queries, etc. I first define five specific relationships between concepts. I extract some concept relationship features in textual and visual domain to train the concept relationship models. The relationship of each pair of concepts will thus be classified into one of the five special relationships. I study the efficacy of the conceptual relationships by applying them to augment imperfect image tags, i.e., improve representative power. We further employ a sophisticated hashing approach to transform augmented image tags into binary codes, which are subsequently used for content-based image retrieval task.

**Index Terms:** Visual Concept Relationship & Hashing, Image Retrieval

### 1. Introduction:

In recent years, exploring semantic relationships between concepts has been a promising research topic in this area, since it has been widely applied on natural language processing, object detection, and multimedia retrieval. However, it is extremely time-consuming and labor-extensive to manually to design such network for distinguishing the complex concept relationship and to construct the network by linguistic experts. Although great efforts have been cost to enlarge the scale of concept network, the concept database is still limited and difficult to update compared to the overall information on the Web. In order to cope with the above problems, some approaches have been proposed to automatically explore the relationship by means of measuring the difference of two concepts. These approaches cost little human effort and can cover far more concepts in the Web. Bussmann proposed the componential analysis to distinguish the meaning of concepts through structured sets of semantic features. The performance of componential analysis depends on the definition of sense components, however, to define these sense components needs a high level of language knowledge and to date there is no uniform standard. Lei et al. proposed Flickr distance to measure the similarity between concepts by the squared root of Jensen Shannon divergence between the corresponding visual language models. However, they only measured the distance of pair wise concepts and cannot recognize the category of concept relationship Hashing techniques have been showing its extremely high efficiency and effectiveness in various Large scale applications, such as multimedia search, multimedia annotation, etc. The typical procedure of hashing is to first transform the original representation of data into low-dimensional features (realvalued) using dimensionality reduction techniques; and then the low-dimensional features are binarized using zero thresholding or other sophisticated learning-based approaches. "Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. We propose a novel framework to facilitate large-scale image retrieval with visual semantic relationship modeling as well as sophisticated hashing technique. Specifically, first define five types of specific semantic relationships between concepts, and model them by extracting textual and visual knowledge from a precious click log dataset collected from a commercial image search engine. An 8-dimensional feature is constructed by considering both textual and visual information. Multimedia related applications. Then, employ these learned relationship models to determine which kind of relationship a pair concept belongs to and quantify the estimated relationships. Furthermore, we construct conceptual relationships set, which is used to augment user-provided image tags, which are often imperfect.

The contributions of the paper are summarized as follows: We assume that concept is composed of visual words, which represents the appearance characteristics of concept. We model the concept by the probability distribution of visual words and the similarity between concepts, as well as the difference of probability distribution. To the best of our knowledge, our work is one of the first to give multifold definitions of the concept relationship in visual domain and perform a comprehensive study on the role of concept relationship features in classifying the concept relationship. Different concept similarity metrics have different capacity to describe the relationship between concepts. We extract different concept relation-ship features from three levels, including textual expression of concepts, visual images of concepts and the co-occurrence frequency of concepts labeled on the same image. We design an approach to augment image tag with the derived conceptual relationships, and a sophisticated hashing method is employed to transform images into

binary codes, which can support efficient image search in large scale

## **2. Related Work:**

In this section, we simply review recent research works on measurement of concept relationship (or distance).

**2.1 Google Distance:** Google distance is proposed to compute the distance between two concepts derived from the number of hits returned by Google search engine when querying both concepts. The concept relationship with the same or similar meanings in a natural language sense tend to be close in units of Google distance, while concepts with dissimilar meanings tend to be farther apart. The normalized Google distance derives from the earlier normalized compression distance. NGD can ideally measure the textual conceptual relationship, i.e. the same and similarity, when the pair wise concepts frequently occur in the same web page. However, NGD is insufficient to measure the concept relationship among daily life, since the knowledge of concept relationship stems from human perception and 80% of human cognition comes from visual information.

**2.2 Flickr Distance:** In order to make up for the lack of Normalized Google Distance, Wu et al. [7] proposed the Flickr Distance to explore the relationship between two concepts in visual domain by measuring the square root of Jensen-Shannon divergence between the visual language models corresponding to these two concepts. If two concepts more frequently appear in the same image, the relationship is more similarity and the Flickr distance is small, otherwise dissimilarity and large. Since the Flickr distance measures the similarity between two concepts based on the images in visual domain, it is in accord with the human cognition. In addition that the Flickr distance is established on the Flickr data, However, Flickr data is very noisy. The computational cost of Flickr distance is quite expensive, so the Flickr distance is inadequate in the large-scale environment.

**2.3 Semantic Networks:** Semantic network is a graph structure for representing knowledge in patterns of interconnected nodes, which represent concepts, and arcs, which represent relations. Sowa et al. [16] proposed the semantic network for artificial intelligence and machine translation. According to types of relations, Semantic networks can be divided into definitional network, assertion network, implicational network, executable network, learning network and hybrid network, one of the most commonly used is the definitional network. Definitional network emphasizes “is-a” relation or inheritance relation between concepts, such as WordNet. In the Word Net, nodes represent concept set with synonymous relation and arcs represent inheritance relation between nodes. WordNet is mainly used to help computer emulates the human capability for external knowledge cognition, and is widely used in data mining, expert system and artificial intelligence field.

**Hashing Techniques:** Among data-independent approaches, the family of Locality-Sensitive Hashing (LSH) has been widely employed in reality. The key idea is to create hash functions by random projections. Several follow-up works have been manipulating the distance measures, such as and kernel similarity, p-norm distances and Mahalanobis metric. But they all suffer from the problem that longer code length is needed to achieve good performance. To overcome the above issues, data-dependent (or learning-based) hashing paradigms have been extensively developed. These kinds of methods are capable of capturing intrinsic data characteristics, as well as generate more effective yet shorter binary codes.

## **3. Visual Semantic Relationships:**

In this section, we first define and model relationships between concepts, and then determine specific relationship of each pair of concepts with the assistance of both textual clue and visual characteristics from click log data of image search engine. Further, we design an effective procedure to augment image textual features, which are originally noisy and incomplete, with the derived concept pairs.

### **3.1 Modelling Concept Relationships:**

**3.1.1 Definition of Relationships:** We first define five types of concept relationships as follows:

Complete Similarity (CS) is used to determine whether two concepts have the same meaning, i.e., synonym, such as “Justin Bieber and J.B.”, “Beijing and Peking”, etc.

Type Similarity (TS) is the relationship indicating two concepts belong to the same object. For instance, Husky and Bulldog are both specific types of dog; and Goldfish and Angelfish are both species of fish.

Hypernym Hyponym (HH) is super-subordinate relation, i.e., A is one kind of B, such as Husky and Domestic Dog.

Parallel Relationship (PR) refers to the relationship that two concepts share the same super base class and often appear in the same image.

**3.1.2 Data Set and Features:** Given any pair of concepts, we need to decide which types of relationships they belong to. We propose to model these five types of relationships with an 8-dimensional feature extracted from Clickture dataset, which is a collection of large-scale click log data of a commercial image search engine.

**3.2 Modelling Visual Semantic Relationships:** In this part, we employ the aforementioned conceptual features and Clickture data to train relationship classifiers. For exploring the relationship between these concepts, we extract 8-dimensional conceptual features and manually annotate the type of relationship between two concepts.

**3.3 Augmenting Image Tags with Visual Semantic Relationships:** Given an image  $x$  with imperfect descriptive tag set  $T = \{t_1; t_2; \dots; t_n\}$ , we may use the derived visual concepts relationships to augment it.

Intuitively, if a concept  $c$  is not in the tag set of  $x$ , but  $c$  is of high relevance to the tags in  $T$ , then  $c$  should be probably added to  $T$ . We design a practical algorithm to perform the tag augmentation procedure.

**3.4 Encoding Image Tags into Binary Representation:** Note that image tags may be of high dimensionality and contain unnecessary redundancy. In this part, we employ a sophisticated indexing approach to generate efficient binary representation of image data.

#### 4. Experiments:

In this section, we evaluate the proposed framework in terms of its ability in enhancing representative power of image tags as well as its efficacy in hashing based image retrieval.

**4.1 Data and Experimental Settings:** We used two large-scale real-world datasets for evaluation. The first one is Clickture, which is one of the largest click log datasets from commercial image search engine. Clickture records user search behaviors in forms namely,

< query; image; number of clicks > :

Clickture contains 212.3 million triads, 40 million unique image keys and 73.6 million unique queries. For image retrieval evaluation, we employ NUSWIDE which is one of the largest real-world image databases with manual labels. It consists of 269,648 Flickr images annotated with 81 concepts. We randomly selected 10% test samples from each concept as query examples to search the results in training samples. We extracted 4096D DeCAF. generic feature, which is the activations of the 6-th layer of a deep CNN trained in a fully supervised fashion on ImageNet. It has shown that this feature is very effective for multimedia tasks on various benchmark data sets. All features are  $\ell_2$ -normalized.

**4.2 Comparison Algorithms:** We compare our proposed framework with other state-of-the-art methods in terms of image retrieval performance. Specially, the following methods are employed:

\_ PCAH, which learns hash functions using principle projections.

\_ SH, which is based on the classic spectral method to find the best binary codes.

\_ AGH, which automatically discovers the neighborhood structure inherent in the data to learn appropriate compact codes.

\_ ITQ, which tries to find a rotation of zero centered data so as to minimize the quantization error of mapping this data to the vertices of a zero-centered binary hypercube.

**Results:** The proposed method can consistently achieve the best performance among all comparison methods with different #bits. This phenomenon indicates that the proposed visual semantic relationships component can work more smoothly with the employ hashing technique as compared to other hashing approaches. We choose to use the best results of different numbers of bits for all comparison algorithms. As we can see, in most cases, our method can achieve better performance than other methods, which further proves that our proposed method is superior with the sophisticated hashing techniques as well as the augmented image tags. Furthermore, in order to obtain more detailed insights of the effects of visual semantic relationships, we perform an additional experiment. We compare our proposed method to its variant without augmentation using visual semantic relationships. The results are shown in Figure 4. As can be seen, for all different numbers of hash codes, if we use visual semantic relationships to enhance image tags, we can always obtain better performance than without using visual semantic relationships. This reveals that the relationships mined from the click log indeed helps to provide positive influence on the final performance.

#### 5. Conclusion:

A new approach to establish the conceptual relationships based on the large scale real world click data from image commercial engine. I handled the problem that the click data suffers from the noise such as typos, the same concept with different queries, etc. In order to automatically establish the concept relationship. First defined five types of specific relationships between concepts, then extracted some concept relationship features in textual and visual domain to train the concept relationship models. The relationship of each pair of concepts will thus be classified into one of the five special relationships. The efficacy of the conceptual relationships by applying them to augment imperfect image tags, i.e., improves representative power. By using a sophisticated hashing approach, transformed augmented image tags into binary codes, which are subsequently used for content-based image retrieval task. Experimental results on NUS-WIDE dataset demonstrated the superiority of our proposed.

#### 6. References:

1. R. Datta, D. Joshi, J. Li, , "Image retrieval: Ideas, influences, and trends of the new age," ACM computing Surveys (CSUR), vol. 40, no. 2, p. 5, 2008.
2. L. Leslie, T.-S. Chua, and J. Ramesh, "Annotation of paintings with high-level semantic concepts using transductive inference and ontology-based concept disambiguation," in ACM Multimedia, 2007, pp. 443–452.
3. B. Wang, Z. Li, M. Li, and W.-Y. Ma, "Large-scale duplicate detection for web image search," in ICME. IEEE, 2006, pp.
4. D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," Communications of the ACM, vol. 38, 11, pp. 33–38, 1995.

5. G. A. Miller, "Wordnet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
6. H. Bussmann, *Routledge dictionary of language and linguistics*. Routledge, 2006.
7. L. Wu, X.-S. Hua, N. Yu, W.-Y. Ma, and S. Li, "Flickr distance," in *ACM Multimedia*, 2008, pp. 31–40.
8. C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE TPAMI*, vol. 37, no. 12, pp. 2531–2544, 2015.
9. T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE TPAMI*, vol. PP, no. 99, pp. 1–1, 2015.
10. C. Xu, D. Tao, and C. Xu, "Large-margin multi-View information bottleneck," *IEEE TPAMI*, vol. 36, no. 8, pp.1559–1572, 2014.
11. D. Tao, X. Li, X. Wu, and S. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE TPAMI*, vol. 29, no. 10, pp. 1700–1715, 2007.
12. D. Tao, J. Cheng, M. Song, and X. Lin, "Manifold ranking- based matrix factorization for saliency detection," *IEEE TNNLS*, vol. PP, no. 99, pp. 1–1, 2015.
13. D. Tao, X. Lin, L. Jin, and X. Li, "Principal component 2-dlong short-term memory for font recognition on single Chinese characters," *IEEE TCYB*, vol. PP, no. 99, pp. 1–1, 2015.
14. W. Cao, R. Hong, M. Wang, and X. Hua, "Multifold concept relationships metrics," in *PCM*, 2014, pp. 238–247.