Discovering Family Photo using Discriminative Frequent Subgraph

Changmin Choi[O*], YoonSeok Lee[+], Sung-Eui Yoon[+]
[*] Division of Web Science and Technology, KAIST
[+] Department of Computer Science, KAIST
changmin.choi@kaist.ac.kr, ys0514@kaist.ac.kr, sungeui@kaist.ac.kr

Abstract

We propose a method to discover family photos from group photos using discriminative subgraphs. We represent an image to a graph with social contexts such as age, gender, and face position. We consider frequent subgraphs from all group photos as features for classification. Numerous subgraphs, however, result in the high dimensions, some of which are not discriminative. To address this issue, we adopt a state-of-the-art frequent subgraph mining technique to remove non-discriminative subgraphs. Our method shows approximately 4%~6% higher classification accuracy in lower feature dimensions compared to the previous method.

1. Introduction

Recent studies on image classification have focused on object and scene classification, and they have shown remarkable performance with the improvement of image features such as convolutional neural network. However, classifying images that contain humans needs not only low-level image features, but also social contexts [5] for more advanced classification tasks and higher accuracy.

Several studies have considered gender, age, face size, face position, cloth, and occurrence to understand the demographics of group photos [1, 2, 3]. Furthermore, the social contexts have been widely used to recognize people and groups [4, 5, 6].

Various graph algorithms with social contexts have been employed in the previous works to discover the correlation among the attributes. To apply the graph algorithms to image classification, images first need to be represented to graphs. Next, the feature extraction from the graph is a significant task, which can affect the performance. In graph classification, the graph pattern mining corresponds to the feature extraction.

The problem is that more contextual data cause more complex graph structures. Some previous works in data mining have proposed methods to discover frequent substructure patterns [7, 8]. Accordingly, the frequent subgraphs firstly were used as a feature for image classification [6]. It needs, however, a heuristic threshold to determine the number of feature dimensions. In addition, only considering the frequency can raise the probability to select non-discriminative subgraphs due to repetitive and redundant patterns specifically with a number of vertices and edges. In other words, considering the frequent subgraphs alone as the feature for classification can bring a scalability problem.

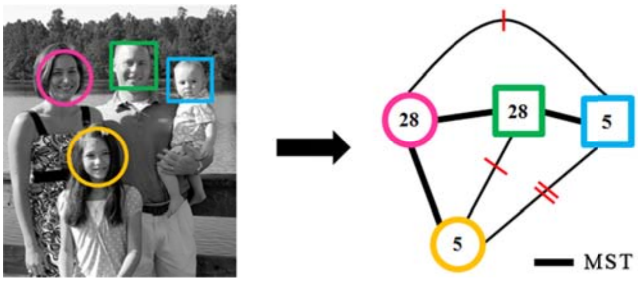

Figure 1. Example of Image to Face Graph

The purpose of this work is to generate more discriminative subgraph features with statistical thresholds for the classification of group photos.

2. Bag-of-Face-sub-Graphs (BoFG) Feature

The attributes of group members enable to discriminate the types of groups, although we do not even know their names or relationships as shown in Fig. 1. This is mainly because human can estimate each person's gender and age and predict their relationship based on those estimations. In addition, understanding each one's position helps to infer the subgroups of a social group. Chen et al. [6] showed that knowing gender, age, and face positions can work effectively for the classification of family and non-family. Thus, we adopted the graph construction of [6] to represent a group photo as a graph.

Fig. 1 illustrates an example of representing a family photo to a face graph. The fully connected graph contains age, gender, and relative positions to each other. It is constructed by the criteria in Sec. 2.1.

2.1 Face Graph

The vertex labels consist of 14 types by age and gender as Fig. 2. The age ranges from 0-year-old to 75-year-old, which are 7 types used in the previous work [1]. The circle means the female, and the square means the male. In total, 14 types are given to each vertex's attribute.

Figure 2. 14 types of vertices by age and gender

Most previous works used the pixel distance using Euclidean distance [1, 4, 9, 10] to measure the closeness. It was, however, not invariant to the scale of image and face, the distance to camera, and the orientation angle of face. Instead, we used the order distance, which was more stable to the pixel scales as identified in [8]. The edge labels consist of integer types by the order distance as Fig. 3. It is the path length among vertices on minimum spanning tree (MST). The MST is initially set by pixel-based distances among face centers.

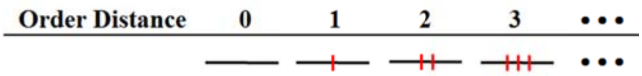

Figure 3. Types of edges by order distance

2.2 Bag-of-Face-sub-Graphs (BoFG)

After the representation of all group photos, the frequent subgraphs extracted from all graphs are regarded as the features for classification. The bag-of-face-subgraph (BoFG) feature was firstly proposed by [6], and it was inspired by the bag-of-word (BoW) [11].

The BoFG is one of the useful features to compare the structure of a group. It helps to infer a type of a group well by using the substructures of groups. For example in Fig. 1, the subgroups of 28(female)-28(male), 28(f)-5(m), and 5(f)-5(m) provides more informative contexts than a simple set of { 28(f), 28(m), 5(f), 5(m) } that does not have any relationship among its elements.

3. Our Approach

The previous work [6] selected the most frequent subgraphs as the BoFG, which were enumerated by gSpan algorithm [7]. In this paper, we use the discriminative subgraphs as the BoFG, which are optimized by CORK algorithm [8].

The method of frequent subgraphs mining based on Apriori approach [12, 13] initially generates candidates and takes a pruning process to remove false positives. The pruning needs the heavy computational cost, because subgraph isomorphism is NP-complete problem. gSpan solved this issue by utilizing two techniques, DFS lexicographic order and minimum DFS code. gSpan introduces DFS code, which represents each edge of DFS tree to a 5-tuple code:

$$(a,\ b,\ l_a,\ l_{(a,b)},\ l_b)\ ,$$

where a and b are visiting order of vertices, l_x is label of a vertex or an edge. Each DFS tree can be represented to a set of DFS codes, which is sorted lexicographically. Finally, we can use minimum DFS tree to check the subgraph isomorphism. With these techniques, gSpan avoids heavy costs of pruning process by blocking false positives in subgraph enumeration.

However, the frequency-based subgraphs have some limitations for graph classification. Most frequent subgraphs hardly show its structural difference among themselves. To overcome the limitations, we usually set low thresholds to generate more subgraphs. CORK [8] considers statistical significance to select discriminative subgraphs instead of only considering the frequency of each subgraph. CORK defines a new measurement counting the number of features that are not helpful for classification among candidate features. This measurement can be integrated into gSpan by simply adding a criterion. It in turn reduces the number of features, while preserving performance in classification.

Figure 4. Our validation dataset includes (a) siblings, (b) single parent, (c) nuclear family, and (d) extended family.

4. Evaluation

We evaluate the effectiveness of the discriminative feature selection with support vector machine (SVM) in a group photo dataset. The classification is conducted with linear kernel and 5-fold cross validation.

4.1 Dataset

We initially asked the dataset of the previous work [6]. Unfortunately, it is not available. Instead, we rearranged the dataset of [1] as the previous work did, and we obtain a soft ground truth containing 1,613 family photos and 1,890 non-family photos. The difference from the previous one is that ours have more 1,073 photos and we comprehensively consider the family types, such as siblings, single parent, nuclear family, and extended family in Fig. 4. Such factors might have caused the lower accuracy (76~ 77% in SVM-Linear) in Table 1 compared to the previous one (88.5% in SVM-Linear) [6] tested with its own smaller dataset.

4.2 Results

Table 1 summarizes that the classification results on the subgraph features by gSpan and by our method. As described in Sec. 3, when extracting the subgraph features by gSpan, we set initially low frequency thresholds to obtain more subgraph feature. On the other hand, when extracting by ours, we do not need to set the threshold. Our method generated maximally 106 subgraphs due to optimizing the discrimination. That means that our feature selection reduced the redundant subgraphs as the dimension reduction techniques do.

Furthermore, the features by our method achieved approximately 4~6% relative improvement to that by gSpan. We also observed that our method outperformed gSpan as increasing dimensions.

The number of the same subgraphs indicates the subgraphs appearing in both features of gSpan and ours. With increasing feature dimensions, the number of discriminative subgraph features increased and the accuracy also increased.

5. Conclusions

Human can coarsely recognize a type of group photos only with contextual information. It motivated us and the previous work [6]. Therefore, we adopted [6]'s method to represent group photos as graphs with age, gender, and face position and extract the frequent subgraphs. However, in the feature selection, we proposed a method to generate more discriminative subgraphs among frequent subgraphs for classification. To verify our assumption, we set a soft ground truth from a public dataset [1] and our method showed higher accuracies in lower feature dimensions.

Acknowledgement

This work was supported by the IT R&D program of MSIP/IITP. [10044970, Development of Core Technology for Human-like Self-taught Learning based on Symbolic Approach]

Table 1. Classification Accuracy of family and non-family in the increasing feature dimensions.

Feature Dimension	The same subgraphs (not)	gSpan	Ours
15	10 (5)	71.82%	71.80%
74	24 (50)	74.11%	**80.67%**
106	26 (80)	76.73%	**81.44%**
212	–	77.76%	–
1175	–	77.50%	–
1961	–	77.28%	–
5104	–	76.28%	–

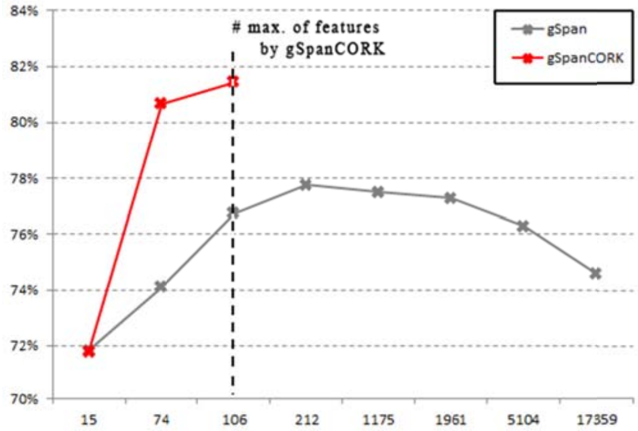

Figure 5. Our validation dataset includes (a) siblings, (b) single parent, (c) nuclear family, and (d) extended family.

References

[1] Gallagher, Andrew C., and Tsuhan Chen. "Understanding images of groups of people." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[2] Chiu, Yi-I., Ru-Yun Hsu, and Chun-Rong Huang. "Spatial Face Context with Gender Information for Group Photo Similarity Assessment." *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014.

[3] Chiu, Yi-I., et al. "Efficient graph based spatial face context representation and matching." Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.

[4] Wang, Gang, et al. "Seeing people in social context: Recognizing people and social relationships." Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010. 169-182.

[5] Murillo, Ana C., et al. "Urban tribes: Analyzing group photos from a social perspective." Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.

[6] Chen, Yan-Ying, Winston H. Hsu, and Hong-Yuan Mark Liao. "Discovering informative social subgraphs and predicting pairwise relationships from group photos." Proceedings of the 20th ACM

international conference on Multimedia. ACM, 2012.
[7] Yan, Xifeng, and Jiawei Han. "gSpan: Graph-based substructure pattern mining." Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE, 2002.
[8] Thoma, Marisa, et al. "Near-optimal Supervised Feature Selection among Frequent Subgraphs." SDM. 2009.
[9] Zhang, Tong, et al. "Consumer image retrieval by estimating relation tree from family photo collections." Proceedings of the ACM International Conference on Image and Video Retrieval. ACM, 2010.
[10] Singla, Parag, et al. "Discovery of social relationships in consumer photo collections using markov logic." Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on. IEEE, 2008.
[11] Sivic, Josef, and Andrew Zisserman. "Video Google: A text retrieval approach to object matching in videos." Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003.
[12] Inokuchi, Akihiro, Takashi Washio, and Hiroshi Motoda. "An apriori-based algorithm for mining frequent substructures from graph data." Principles of Data Mining and Knowledge Discovery. Springer Berlin Heidelberg, 2000. 13-23.
[13] Kuramochi, Michihiro, and George Karypis. "Frequent subgraph discovery." Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on. IEEE, 2001.