석사 학위논문

Master's Thesis

가족사진 발견을 위한 차별적 서브그래프

Discriminative Subgraphs for Discovering Family Photos

최 창 민 (崔 昌 珉 Choi, Changmin)

웹사이언스공학

Graduate School of Web Science Technology

KAIST

2015

가족사진 발견을 위한 차별적 서브그래프

Discriminative Subgraphs for Discovering Family Photos

Discriminative Subgraphs for Discovering Family Photos

Advisor : Professor Yoon, Sung-Eui

by

Choi, Changmin
Graduate School of Web Science Technology
KAIST

A thesis submitted to the faculty of KAIST in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Graduate School of Web Science Technology . The study was conducted in accordance with Code of Research Ethics[1].

2015. 06. 02.

Approved by

Professor Yoon, Sung-Eui

[Advisor]

---

[1]Declaration of Ethical Conduct in Research: I, as a graduate student of KAIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.

가족사진 발견을 위한 차별적 서브그래프

최 창 민

위 논문은 한국과학기술원 석사학위논문으로
학위논문심사위원회에서 심사 통과하였음.

2015년 6월 16일

심사위원장 윤 성 의 (인)

심사위원 이 흥 규 (인)

심사위원 차 미 영 (인)

MWST 20134642

최 창 민. Choi, Changmin. Discriminative Subgraphs for Discovering Family Photos. 가족사진 발견을 위한 차별적 서브그래프. Graduate School of Web Science Technology . 2015. 17p. Advisor Prof. Yoon, Sung-Eui. Text in English.

ABSTRACT

We propose a method to discover family photos from group photos using discriminative subgraphs. Group photos are represented to graphs by social contexts such as age, gender, and face position. The previous work [1] considered the frequent subgraphs from all group photos as features for classification. The feature is a form of bag-of-word model.

However, this approach produces numerous subgraphs, resulting in high dimensions. Furthermore, some of them are not discriminative. To solve this, we adopt a state-of-the-art frequent subgraph mining that removes non-discriminative subgraphs. We also use TF-IDF normalization, which is more suitable for the bag-of-word model. To validate our method, we experiment in two data sets: Chen's [1] and ours. Our method shows consistently better performance, higher accuracy in lower feature dimensions, compared to the previous method.

Contents

List of Tables

List of Figures

Chapter 1. INTRODUCTION

Recent studies on image classification focus on object and scene classification. They show remarkable performance thanks to the improvement of image features, such as convolutional neural network. Most of image features are built from low-level descriptors, which are extracted from pixels. The low-level feature is not enough to describe a group photo since it has more semantic information such as relations, events, or activities. The semantic information on group photo can help to preserve the privacy of end-users in photo-sharing service or in image retrieval. Interestingly, human can roughly recognize those information without prior knowledge because we can estimate a variety of contexts, such as age, gender, closeness, and place, by observing face, position, cloth, and other objects. Accordingly, Chen et al. [1] proposed a method to classify group photos into family and non-family. The system assumes that the annotations about age, gender, and face position are well-estimated beforehand by face detection and statistical estimation from low-level feature. As result, they build a high-level feature named as bag-of-face-sbugraph (BoFG) to represent group photos to graphs. In construction of BoFG, a frequent subgraph mining algorithm is adopted by the assumption that social subgroups in a group resemble subgraphs in a graph.

However, the frequent subgraph mining in the previous work has the limitations to enumerate discriminative subgraphs for classification. First, it needs a hand-tuned threshold to determine the number of feature dimensions in training phase. Second, frequency threshold can raise the probability to select non-discriminative subgraphs due to repetitive and redundant patterns, specifically in case with a number of vertices and edges. In other words, thresholding the number of subgraphs using frequency alone can cause scalability problem. As additional experiment, we validate the effectiveness of term frequency and inverse document frequency (TF–IDF) weighting, which is widely used for bag-of-word (BoW) model.

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

- Revising a method to extract near-optimal and discriminative subgraphs
- Achieving higher classification accuracy with lower feature dimensions
- Achieving higher classification accuracy via TF-IDF normalization
- Experimenting in a new dataset with more images

Chapter 2. RELATED WORK and BACKGROUND

2.1 Social Context in Photographs

Social contexts contain various information such as cloth, age, gender, absolute or relative position, face angle, gesture, body direction, and so on. Social context has been widely used to recognize people and groups [1, 8, 14]. Several works analyzed the context to study locations and other information of groups in photos [5, 6]. Some researchers utilized social context to classify group types, retrieve similar group photos, discover social relations, or predict occupations [1, 4, 9, 8, 10, 11, 12, 14, 16].

To recognize a type of group photos, we need to consider not only low-level, but high-level features [8]. Some of well-known low-level features include SIFT [**?**], GIST [**?**], CNN [**?**], etc. High-level social features can be estimated by face detection, cloth segmentation, or partial body detection. Group structures as high-level features in photographs can be considered by searching similar images, assessing similarity or detecting a certain type of scenes [2, 3, 6].

2.2 Social Subgraphs

Chen et al. [1] proposed a classification system using Bag-of-Face-subGraph (BoFG) feature for group photo classification. The system builds face graphs and utilizes their subgraphs to describe social relationships. BoFG is analogous to the Bag-of-Word (BoW) model in information retrieval and text classification. Specifically, a text corpus corresponds to a group photo album, a document corresponds to an image, and a word corresponds to a subgraph in a face graph. The main difference between these models is that BoW performs clustering over all vectors in order to obtain a codebook, whereas BoFG performs frequent subgraphs mining over all face graphs.

Attributes of group members enable to discriminate the type of groups, although we do not even know their names or relationships, because human can estimate the gender and age by simply looking at photos. In addition, understanding each one's position is informative to recognize the type of a group. Chen et al. showed that only knowing gender, age, and face positions works effectively for a binary classification of family and non-family photos.

Our approach is also based on this method, and represent a group photo into a face graph, elaborated in the subsequent section.

2.2.1 Face Graph

Fig. 2.2 illustrates an example of representing a family photo to a face graph. Each node of the graph corresponds to each person and is associated with a vertex label describing age and gender. Each edge between two nodes encodes relative positions between two persons of those two nodes.

There are 14 different types describing age and gender for each vertex label. The age ranges from 0-year-old to 75-year-old, and is categorized into seven age types. There are two gender types, male and females, and they are visualized with and squres and circles, respectively in Fig. 2.3. These age and gender types result in 14 different types.

Most previous works used the Euclidean distance in the image space, i.e., pixel distance, to measure the

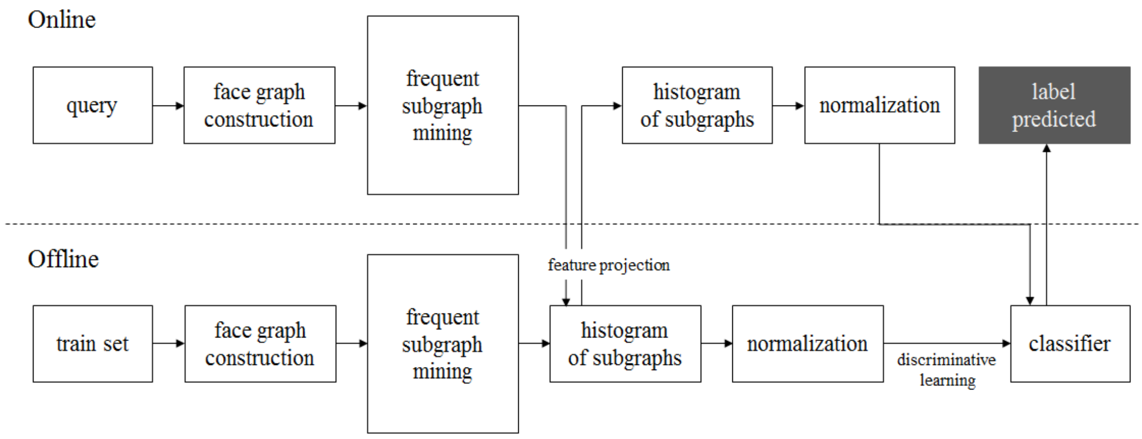


Figure 2.1: The overview of group photo classification system.

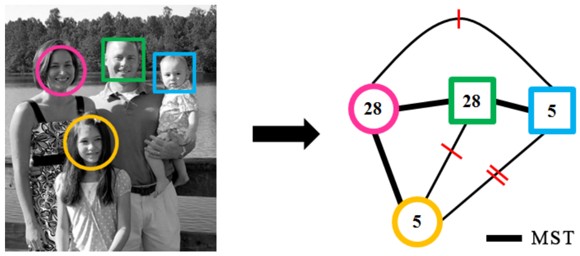


Figure 2.2: Representing a group image to a face graph.

closeness between persons in group photos [6, 14, 16, 11]. Unfortunately, it has been known not to be invariant to scales of images, faces, distance to camera, or the orientation angle of a face.

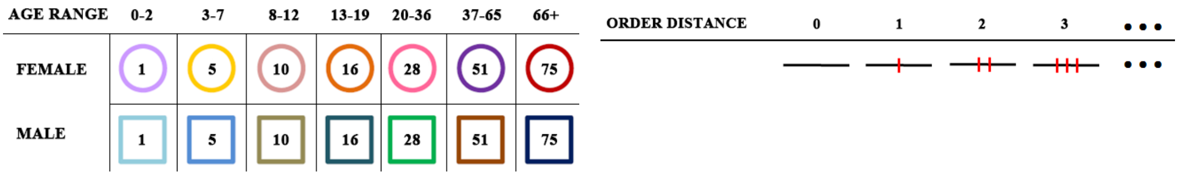
Instead, we use an order distance, **YOON: Define this here** . The order distance has been demonstrated to be more stable over the pixel distance in terms of various factors [1]. The order distance is computed as the path length among vertices on a minimum spanning tree (MST) computed from the face graph. Such order distance is used for each edge label (Fig. 2.2.1 (b)).

Bag-of-Face-subGraphs (BoFG). Once we represent group photos into face graphs, we extract frequent subgraphs and use them features, BoFG, for classification. BoFG has been proposed to be a useful feature to compare structures of group photos. It helps to infer a type of a group by using substructures of groups. For example, in Fig. 2.2, the subsets, $\{28(f)–28(m)\}$, $\{28(f)–5(M)\}$, and $\{5(f)–5(m)\}$ computed from two persons, can provide more information than each node such as $\{28(f), 28(m), 5(f), 5(m)\}$, where f and m represents female and male gender types, respectively.

2.3 Frequent Subgraph Mining

Frequently appearing subgraphs provide important cues on understanding graph structures and similarity between different graphs. As a result, mining frequent subgraphs has been widely studied [7].

For various classification, frequent subgraph mining has been used in training and test phases to build a high-level feature, as used in classifying family and non-family photo types [1]. We have found that extracted subgraphs significantly affect classification accuracy. There are two simple strategies to explore the subgraphs in a



(a) 14 types of vertices

(b) 4 types of edges

Figure 2.3: Types of vertex and edge for face graph.
YOON: Combine this figure with the face graph fig.

database: (1) DFS-based and (2) BFS-based approach [7]. The BFS-based, Apriori-like, algorithm has challenges in candidate generation and pruning false positives **YOON: How about DFS?** More advanced techniques focus on efficient candidate generation, since subgraph isomorphism tests used for many frequent subgraph mining is an NP-complete [7].

YOON: mention what is the novelty of our method briefly here.

Chapter 3. OUR APPROACH

Start from here....

3.1 Overview

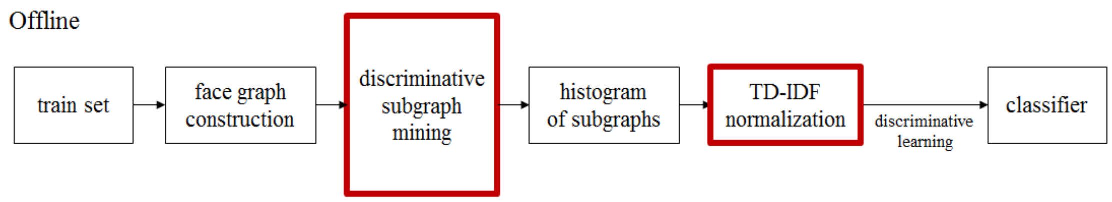


Figure 3.1: The overview of our approach.

We enhance the methods of subgraph enumeration and feature normalization. The previous work [1] adopted the graph-based substructure pattern mining (gSpan) and document frequency (DF) for feature extraction and selection respectively. Among the numerous subgraphs, DF decreased plausibly feature dimensions while minimizing accuracy loss.

Recent studies in graph mining, however, showed more various methods to select discriminative subgraphs. Thoma et al. [13] proposed a near-optimal selection method combined with gSpan, called correspondence-based quality criterion (CORK). We adopt CORK to build more discriminative and fewer features than those by gSpan. Additionally, we normalize BoFG feature using term frequency and inverse document frequency (TF–IDF) that is more appropriate for BoW model.

3.2 Discriminative Subgraphs Mining

3.2.1 Subgraph Enumeration via gSpan

The previous work regarded the most frequent subgraphs as BoFG features, which are enumerated by gSpan. We use the discriminative subgraphs as BoFG, which are optimized by gSpan and CORK. The method of frequent subgraphs mining based on Apriori approach [12, 13] initially generates candidates and takes 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 minimal DFS code. gSpan introduces DFS code, which represents a graph (G) to a 5-tuple code:

$$G = (m,\ n,\ L_m,\ L_{(m,n)},\ L_n) \tag{3.1}$$

where m and n are vertex indices by visiting order, L_m and L_n are vertex labels of v_m and v_n, $L(m,n)$ is a edge label between v_m and v_n. The edge $E_{(m,n)}$ is forward, if $m < n$, otherwise the edge $E_{(m,n)}$ is backward. A single graph, however, can have multiple DFS codes. DFS Lexicographic order enables to have the minimal DFS code of a single graph, which guarantees no duplication in subgraph enumeration. DFS code extends an edge in rightmost path that visits vertices left-to-right in adjacency list. The rightmost path is a shortest path between v_0 and v_n through forward edges. While building a DFS code on DFS code tree as Fig. 3.2, The extension of backward edges must begin from the rightmost vertex and occur before that of forward edges.

Finally, we can use minimum DFS tree to check the subgraph isomorphism. With these techniques, gSpan avoid heavy costs of pruning process by blocking false positive in subgraph enumeration.

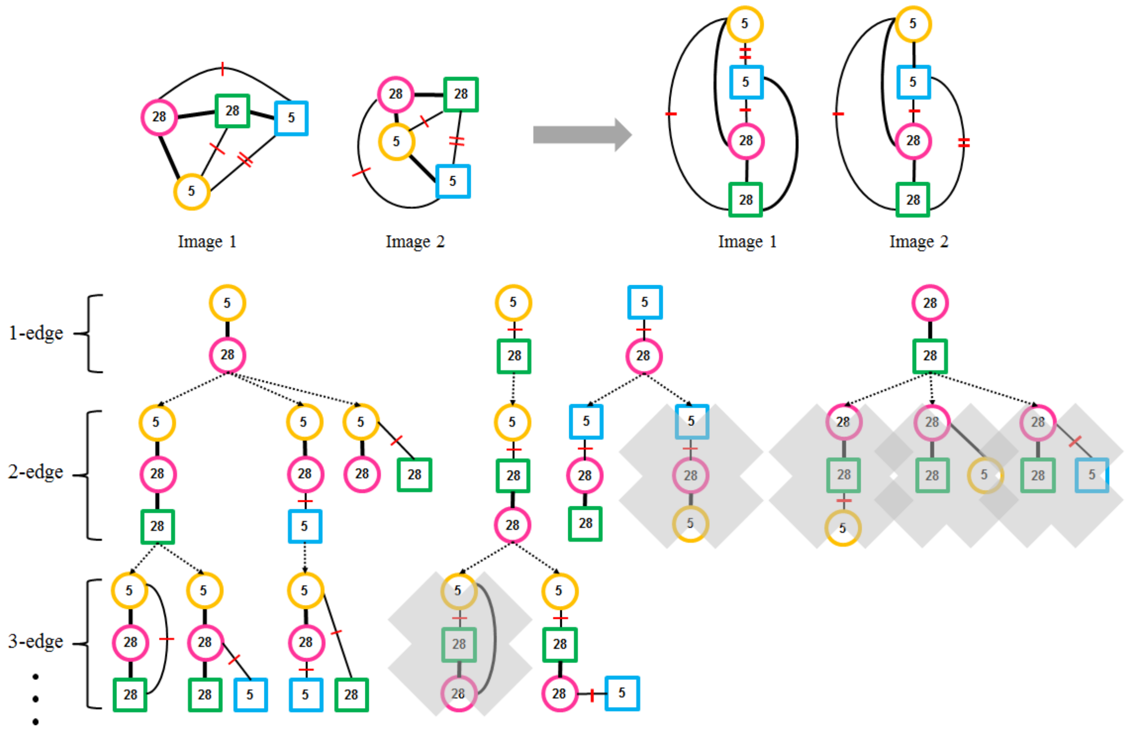


Figure 3.2: It is an example of enumerating all subgraphs. In the enumeration, the search space is pruned by checking if a graph set is a minimal DFS code.

3.2.2 Discriminative Subgraph Selection via CORK

The frequency-based subgraphs have some limitations for graph classification. Most frequent subgraphs hardly show its structural difference among themselves. The minimum frequency of subgraphs and the ratio of DF selection should be picked through several trials. To overcome the limitations, we usually set a low threshold to generate more subgraphs and need to try various ratio values.

CORK 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 which are not helpful for classification among candidate features. This measurement can be integrated into gSpan by adding a criterion. It reduces the number of features while preserving performance in classification and can prune search spaces without a frequency threshold.

A near-optimality of CORK is obtained from a submodular quality function $q(\cdot)$ using greedy forward feature selection. The $q(\cdot)$ considers presence or absence of each subgraph in each class. The $q(\cdot)$ for a single feature is,

$$q(\{S\}) = -(A_{S_0} \cdot B_{S_0} + A_{S_1} \cdot B_{S_1}) \tag{3.2}$$

where S is a subgraph, A and B are classes in data set. A_{S_0} is the number of the subgraphs not-contained in class A. A_{S_1} is the number of the subgraphs contained in class A. Furthermore, Let us assume that we have only two subgraphs S and T as feature candidates. The $q(\cdot)$ for two features is,

$$q(\{S,T\}) = -(\sum_{i,j=0}^{1} A_{S_i,T_j} \cdot B_{S_i,T_j}) \tag{3.3}$$

where T is a supergraph of S, $S \subset T$. S_i and T_j are the possible combinations of two features in each class. If the number of features increases to N, the number of possible feature combinations can increase exponentially to 2^N. Fig 3.3 shows an example of CORK score for two subgraphs in class A and B.

data set

subgraphs

	A			B		
	a_1	a_2	a_3	b_1	b_2	b_3
S	**1**	**0**	**0**	**1**	**0**	**1**
T	**0**	**1**	**1**	**0**	**1**	**1**

Figure 3.3: $a_{1\sim 3}$ and $b_{1\sim 3}$ are images in a given database. The indicator vector is 1, if each subgraph appears in each image, otherwise 0. CORK score by Eq. 3.3 is $-(0\cdot 0+2\cdot 1+1\cdot 1+0\cdot 1) = -3$.

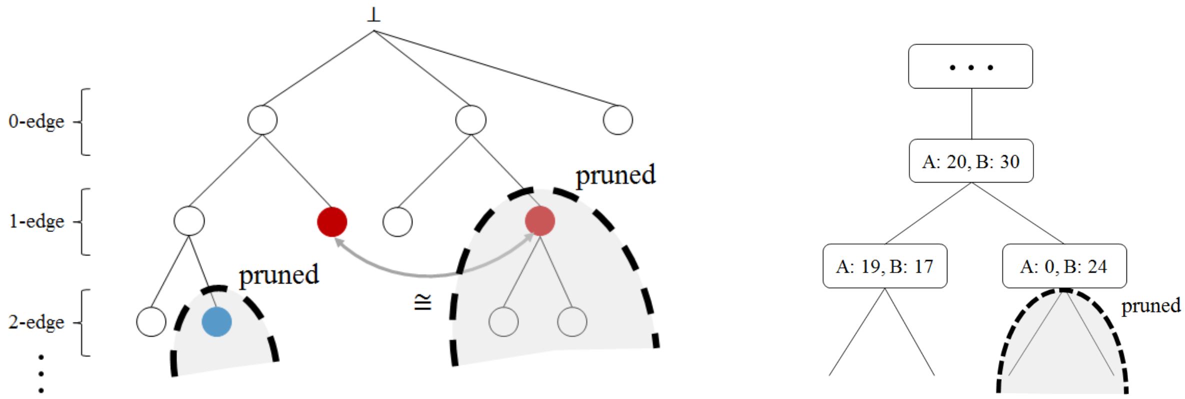
The upper bound of CORK value is derived from three possible cases when the supergraph T is added to its feature set. We need to compute the correspondences that would be removed by T. The best scenario is that T turns all 1-value vectors in one class to 0-value vectors while the vectors in the other class remain unaffected. The first case, thus, is that all 1-value vectors in class A are replaced to 0-value vectors. The second case is that the same scenario is applied to class B. The third case is $q(\{T\}) = q(\{S\})$. In this way, the maximal CORK score of T can be formulated.

$$q(\{T\}) \leq \; q(\{S\}) + max \begin{Bmatrix} A_{S_1} \cdot (B_{S_1} - B_{S_0}) \\ (A_{S_1} - A_{S_0}) \cdot B_{S_1} \\ 0 \end{Bmatrix} \tag{3.4}$$

This equation is applied in the first iteration of greedy forward selection. If the size of feature set grows larger, the graphs which are the only part of a correspondence are considered [13].

gSpan needs a tedious mid-task to obtain sufficient subgraphs before DF selection. It is to set a experimental threshold. It may be determined by either relative ratio to the scale of data set or absolute value such as the number of subgraphs. Neither of them, unfortunately, predicts the total of subgraphs that directly leads to feature dimension. Even if each total of subgraphs in two data sets is equal to each other, the complexity of each face graph is different by the amount and type of vertices and edges.

As a comparative advantage, CORK can work independently without minimum frequency. CORK prunes search space by near-optimal quality score whereas gSpan determines by heuristic minimum frequency.



(a) pruning in gSpan

(b) pruning in CORK

Figure 3.4: (a) is the pruning in gSpan. The blue subgraph is pruned by minimum number of frequency, and the red subgraph s is pruned by minimal DFS code. (b) is the pruning in CORK. The discriminative power of subgraphs reaches to maximal when correspondences in class A become zero.

3.3 TF-IDF Normalization

The previous work [1] used normalization by the sum of term-frequency in each image. We are sure that TF-IDF technique reflects the importance of each subgraph more effectively. As introduced in Sec. 2.2, BoFG feature resembles BoW model. Likewise, the term frequency needs to be de-weighted by the inverse document frequency. To correct the term weights, we adopt TF-IDF with logarithmically scaled frequency.

$$\begin{gathered} TF\text{–}IDF(s,\, i,\, D) = TF(s,\, i) \;\times\; IDF(s,\, D) \\ TF(s,\, i) = \log(1 + f_{s,i}), \qquad IDF(s,\, D) = \log\left(\frac{N}{1 + |\{i \in D:\, s \in i\}|}\right) = \log\left(\frac{N}{1 + n_s}\right) \end{gathered} \tag{3.5}$$

where $f_{s,i}$ is the number of subgraph s occurring in a single image i, N is the number of all images in database D, and n_s is the number of images with subgraph s. If $f_{(s,i)}$ is zero, TF value will be undefined. We add, therefore, 1 to it. If a subgraph s is not in database D, it causes a division-by-zero. We adjust the denominator n_s to $1 + n_s$.

Chapter 4. EVALUATION

We evaluate the effectiveness of the discriminative feature selection and TF-IDF normalization with support vector machine (SVM). The classification is conducted with linear kernel and 5-fold cross validation.

4.1 Dataset

To validate our approach, we experiment in Chen's [1] and our new dataset. The our data set was rearranged from the public data set [6] as the previous work did, and we could 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 has more 1,073 photos and we considered the comprehensive family types, such as siblings, single parent, nuclear family, and extended family in Fig. 4.1. We composed our data set not with prior knowledge, but with human intelligence only.

(a) non-family (b) siblings (c) single parent (d) nuclear family (e) extended family

Figure 4.1: Our data set consists of non-family (a) vs. family (b + c + d + e).

Chen's (2,444) *Ours* (3,503)

	(A)	*(B)*	*(C)*
family	66	1,111	502
non-family	136	1,131	759

Figure 4.2: The composition of Chen's $(A + B)$ and our data set $(B + C)$. (B) indicates the number of co-occurring images in both Chen's and ours. Most images in Chen's were reselected though we did not have preliminary knowledge.

4.2 Results

4.2.1 Effects of CORK

To confirm the effectiveness of ours, we compare the accuracy of the previous one and ours at the same level of feature dimensions. Once 10,000 subgraphs are enumerated approximately, they are sorted by descending order of document frequency and we consider them as BoFG. Note that DF ratios range from 0.01 to 1.0. High-dimensional BoFG increased the difference of the number of subgraphs enumerated. The range of differences was ascended tens to hundreds. The inaccuracy in subgraphs does not significantly affect the prediction result.

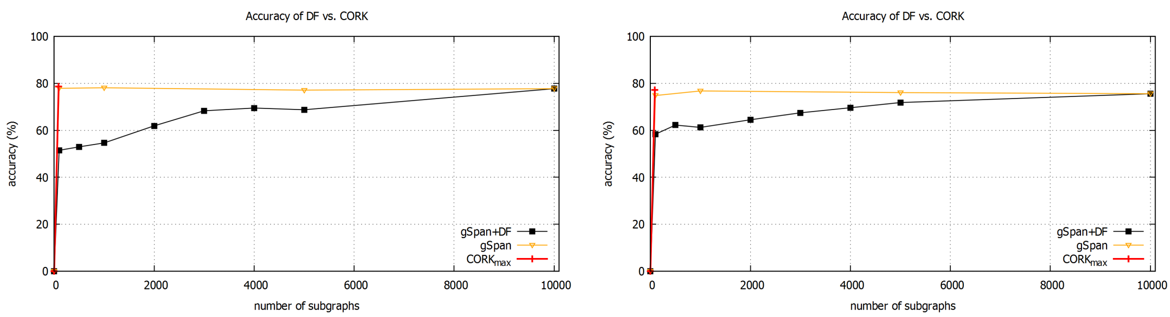
Chen's data set									
selection \ *dimension*	90	100	500	1,000	2,000	3,000	4,000	5,000	10,000
$gSpan+DF$	50.00%	51.51%	52.98%	54.65%	61.92%	68.33%	69.52%	68.78%	77.76%
$gSpan$	78.61%	77.92%	80.12%	78.16%	77.51%	77.31%	76.49%	77.14%	77.76%
$CORK_{max}(Ours)$	78.65%	—							
Our data set									
selection \ *dimension*	90	100	500	1,000	2,000	3,000	4,000	5,000	10,000
$gSpan+DF$	56%	58.37%	62.25%	61.26%	64.51%	67.48%	69.64%	71.84%	75.61%
$gSpan$	74.78%	74.84%	77.43%	76.8%	76.63%	76.83%	76.49%	76.09%	75.61%
$CORK_{max}(Ours)$	77.26%	—							

Table 4.1: The accuracy of gSpan vs. CORK in Chen's and our data set.

The previous work [1] introduced briefly its analogy with text classification. We are not sure that DF is adequate for feature selection because there is an ambiguity about which phase DF is implemented in, either during or after gSpan. Note that $gSpan$ and $gSpan+DF$ correspond respectively to the adaption of DF *during* and *posterior to* gSpan in Table 4.1. $CORK_{max}$ is the maximum number of subgraphs when the quality criterion works as a threshold without minimum frequency.

The former, implementing DF during gSpan, means thresholding by minimum frequency in DFS code tree. As introduced in Sec. 3.2.2, setting minimum frequency requires trial and error to attain the same number of dimensions. If minimum threshold is set too high to take more numerous subgraphs, DFS code tree limits opportunities to traversal sibling vertices. On the other hand, if minimum threshold is too low to obtain a various type of subgraphs, few subgraphs are taken. With the trade-off, CORK does not deeply go into children graphs while approximating a near-optimal quality. Table 4.1 shows that gSpan without DF also results in the accuracy similar to that of CORK. It implies that the criteria of family and non-family by human intelligence are more toward diverse and small subgraphs containing a few vertices and edges.

The latter, implementing DF after gSpan, yields different subgraphs compositions. We checked the number of identical subgraphs between BoFG and test images during 5 cross-validation as Table 4.4 and 4.5. The posterior DF selection was not helpful enough to represent group images with BoFG. It showed poor results even in hundreds of thousands of subgraphs.



(a) Chen's data set

(b) Our data set

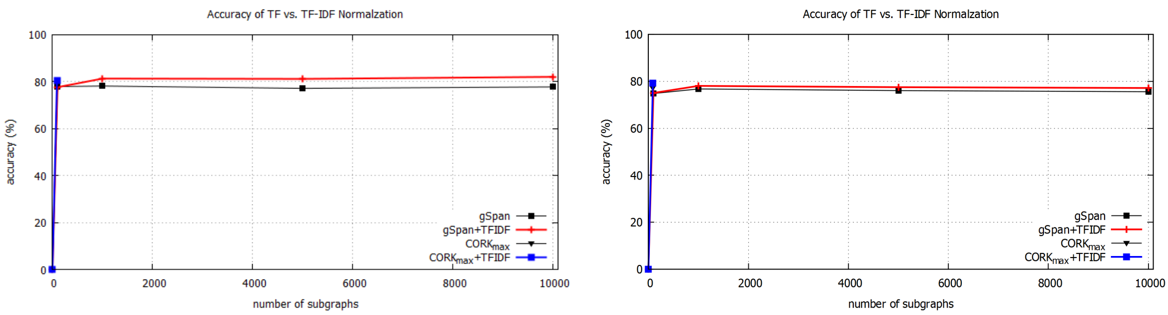
Figure 4.3: The accuracy by gSpan vs. CORK in our data set.

4.2.2 Effects of TF-IDF Normalization

We experimented TF–IDF with gSpan and CORK. TF–IDF does not fit with gSpan+DF since DF posterior to gSpan originally lets BoFG eliminate the characteristic of Bow model. In both gSpan and CORK, TF–IDF improved mostly the classification accuracy and worked better with CORK and high-dimensional gSpan.

Chen's data set					
normalization \ dimension	76	100	1,000	5,000	10,000
gSpan	78.61%	77.92%	78.16%	77.14%	77.76%
gSpan+TF–IDF (Ours)	77.67%	77.63%	81.31%	81.18%	82.04%
$CORK_{max}$	78.65%	—			
$CORK_{max}+TF–IDF$ (Ours)	80.61%	—			
Our data set					
normalization \ dimension	90	100	1,000	5,000	10,000
gSpan	74.78%	74.84%	76.8%	76.09%	75.61%
gSpan+TF–IDF (Ours)	75.40%	75.09%	78.09%	77.55%	77.2%
$CORK_{max}$	77.26%	—			
$CORK_{max}+TF–IDF$ (Ours)	79.34%	—			

Table 4.2: The accuracy of TF vs. TF–IDF in Chen's and our data set.



(a) Chen's data set (b) Our data set

Figure 4.4: The accuracy of TF vs. TF–IDF in Chen's and our data set.

4.2.3 Comparasion of Subgraphs

We check the number of subgraphs co-occurring in the BoFG features generated by both gSpan and CORK. It helps to understand how much different the feature composition by CORK is. For the experiment, we choose the train set which is the closest to the average accuracy among 5-validation sets. Even in hundreds of thousands of dimensions, the same composition of subgraphs by CORK is not achieved using by gSpan.

Chen's data set: 77 subgraphs by $CORK_{max}$									
$gSpan\ subgraphs$ / $CORK\ subgraphs$	78	100	495	997	10,301	⋯	111,764	⋯	560,177
$Identical\ \ Subgraphs$	16	17	21	23	36	⋯	54	⋯	59
Our data set: 85 subgraphs by $CORK_{max}$									
$gSpan\ subgraphs$ / $CORK\ subgraphs$	85	99	505	996	9,748	⋯	84,713	⋯	326,034
$Identical\ \ Subgraphs$	20	21	25	26	29	⋯	50	⋯	59

Table 4.3: The number of identical subgraphs between gSpan vs. CORK in Chen's and our data set.

In test phase, the feature projection by CORK performs better than gSpan or gSpan+DF does. Table 4.4 and 4.5 display that gSpan+DF returns very few identical subgraphs in low dimension. It causes worse performance in projecting test images on BoFG. With gSpan+DF, the fewer subgraphs are, the more test images are without a BoFG subgraph. On the other hand, gSpan shows better performance than gSpan+DF in most dimensions. However, there remains the problem of hand-tuned thresholds. Compared to both methods, CORK represents the maximum number of test images on BoFG with the minimum of subgraphs.

$dimension$ / $selection$	76	100	1,000	5,000	10,000
$Identical\ \ Subgraphs\ \ with\ \ test\ set\ \ (actual\ \ subgraphs)$					
$gSpan+DF$	3 (76)	3 (100)	103 (1,006)	2,240 (5,030)	8,568 (10,061)
$gSpan$	76 (76)	100 (100)	992 (995)	4,669 (4,965)	8,568 (10,061)
$CORK_{max}$	37 (76)	—			
$Test\ \ Images\ \ with\ \ No\ \ Vectors$					
$gSpan+DF$	283	283	189	82	30
$gSpan$	53	52	39	37	30
$CORK_{max}$	26	—			
$Avg.\ \ Amount\ \ of\ \ Subgraphs\ \ from\ \ Test\ \ set:$ 45,686 (Std.: 19,353)					

Table 4.4: The number of identical subgraphs between BoFG and Test set & test images without a BoFG feature in Chen's data set.

It is impractical to execute gSpan many times for numerous subgraphs. In Chen's data set, we regard the minimum frequency as 15 to have subgraphs as many as possible. This threshold does not enumerate the accurate number of subgraphs that we expect. The amount of subgraphs ranges from approximately 20,000 to 80,000 over the same threshold. We round off the number of subgraphs to the nearest integer.

In our data set, the minimum frequency is set to 20. The amount of subgraphs ranges from approximately 19,000 to 180,000 for the threshold.

dimension / *selection*	90	100	1,000	5,000	10,000
Identical Subgraphs with test set (acutal subgraphs)					
$gSpan+DF$	3 (89)	3 (98)	98 (989)	2,355 (4.947)	9,473 (9,894)
$gSpan$	90 (90)	100 (100)	1,001 (1,001)	5,009 (5,040)	9,473 (9,894)
$CORK_{max}$	43 (90)	—			
Test Images with No Vectors					
$gSpan+DF$	351	351	236	163	43
$gSpan$	62	62	50	48	43
$CORK_{max}$	28	—			
Avg. Amount of Subgraphs from Test set : $78,110$ (Std.: 65,973)					

Table 4.5: The number of identical subgraphs and test images without a BoFG feature in our data set.

Chapter 5. CONCLUSION & FUTURE WORKS

Human can roughly recognize a type of group photos only with contextual information. It motivated us and the previous work [1]. We adopt BoFG feature [1] to represent group photos as graphs with age, gender, and face position. The BoFG method, however, needs a hand-tuned threshold to enumerate subgraphs and the select discriminative features. In addition, TF–IDF normalization reflects the characteristic of BoFG better than the previous method since BoFG is analogous to BoW model. Thus, we propose to adopt a quality function with near-optimal guarantees [13] to generate more discriminative subgraphs among frequent subgraphs for classification. The method work independently as a threshold, which contributes to fewer subgraphs. To validate our approach, we set a new data set with soft ground truth by rearranging a public dataset [6]. Our data set includes more 1,059 images than Chen's data set [1]. Interestingly, The proportions of family and non-family images reselected from Chen's are 94.39% and 89.27%. It denotes that human intelligence without prior knowledge enables to determine whether the type of group photo is family. Finally, our method performs similar or higher accuracy in the lowest feature dimension. The feature set is optimally established without multiple executions of subgraph mining algorithm.

We further need to analyze how much the minimum number of subgraph vertices affects to represent group photos; because we set the minimum value as 2 in this work. The single vertex with no edge may be more discriminative and The vertex with more 3-edges may be redundant. We also need to consider other methods of subgraphs enumeration over query. The subgraphs from query affect directly the prediction accuracy. A BFS-based enumeration may be more efficient in consideration of the types of subgraphs.

References

[1] Y.-Y. Chen, W. H. Hsu, and H.-Y. M. Liao. Discovering informative social subgraphs and predicting pairwise relationships from group photos. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 669–678. ACM, 2012.

[2] Y.-I. Chiu, R.-Y. Hsu, and C.-R. Huang. Spatial face context with gender information for group photo similarity assessment. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 2673–2678. IEEE, 2014.

[3] Y.-I. Chiu, C. Li, C.-R. Huang, P.-C. Chung, and T. Chen. Efficient graph based spatial face context representation and matching. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 2001–2005. IEEE, 2013.

[4] W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese. Discovering groups of people in images. In *Computer Vision–ECCV 2014*, pages 417–433. Springer, 2014.

[5] A. C. Gallagher and T. Chen. Finding rows of people in group images. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 602–605. IEEE, 2009.

[6] A. C. Gallagher and T. Chen. Understanding images of groups of people. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 256–263. IEEE, 2009.

[7] C. Jiang, F. Coenen, and M. Zito. A survey of frequent subgraph mining algorithms. *The Knowledge Engineering Review*, 28(01):75–105, 2013.

[8] A. C. Murillo, I. S. Kwak, L. Bourdev, D. Kriegman, and S. Belongie. Urban tribes: Analyzing group photos from a social perspective. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 28–35. IEEE, 2012.

[9] K. Shimizu, N. Nitta, Y. Nakai, and N. Babaguchi. Classification based group photo retrieval with bag of people features. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, page 6. ACM, 2012.

[10] H. Shu, A. Gallagher, H. Chen, and T. Chen. Face-graph matching for classifying groups of people. In *2013 IEEE International Conference on Image Processing*, 2013.

[11] P. Singla, H. Kautz, J. Luo, and A. Gallagher. Discovery of social relationships in consumer photo collections using markov logic. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–7. IEEE, 2008.

[12] Z. Song, M. Wang, X.-s. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1084–1091. IEEE, 2011.

[13] M. Thoma, H. Cheng, A. Gretton, J. Han, H.-P. Kriegel, A. Smola, L. Song, P. S. Yu, X. Yan, and K. M. Borgwardt. Discriminative frequent subgraph mining with optimality guarantees. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 3(5):302–318, 2010.

[14] G. Wang, A. Gallagher, J. Luo, and D. Forsyth. Seeing people in social context: Recognizing people and social relationships. In *Computer Vision–ECCV 2010*, pages 169–182. Springer, 2010.

[15] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pages 721–724. IEEE, 2002.

[16] T. Zhang, H. Chao, C. Willis, and D. Tretter. Consumer image retrieval by estimating relation tree from family photo collections. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, pages 143–150. ACM, 2010.

Summary

Discriminative Subgraphs for Discovering Family Photos

최근 이미지 분류는 물체나 배경분류에 주력하고 있으며, 이미지 특징(feature)도 그에 맞춰 픽셀 정보로부터 추출된다. 그러나 여러 사람이 나타나는 사진일 경우, 촬영목적이 사람에 맞춰져 있어 물체 및 배경분류 뿐만 아니라 관계나 이벤트와 같이 조금 더 높은 수준의 의미론적 정보(semantic)를 내포하고 있을 수 있다. 이전 연구에서는 이를 픽셀 정보에서 추출하는 것보다 성별, 나이, 얼굴 위치 등과 같은 사회적 맥락(social context)에서 더 잘 파악할 수 있다고 주장한다. 그 근거로, 사회학 관점에서 한 그룹 안에도 여러 작은 그룹(subgroup)이 존재하듯이 이를 그래프로 표현해 서브그래프(subgraph)들로 각 이미지를 재표현하는 것이 가능하다고 말한다. 이를 증명하기 위한 실험으로 총 2,444장의 가족과 비가족 그룹 분류를 하였다.

본 논문에서는 동일한 특징 추출방법을 따르되, 서브그래프를 생성하는 과정에서의 몇 가지 한계점들을 극복할 수 있는 방법과 추출된 특징값의 정규화 방식을 개선하는데 주력하여 최종적으로는 더 적은 특징값을 가지고 더 높거나 기존과 동일한 수준의 분류결과를 산출하는데 성공하였다. 먼저, 기존 연구의 서브그래프 추출 방식은 깊이우선탐색(DFS-based)기반의 후보군 생성방식으로 데이터 셋이 나타나는 최소 빈도수를 임계값으로 설정하여 그보다 작은 서브그래프는 생성하지 않는 방식이었다. 이것은 원하는 서브그래프의 양을 정확히 조절할 수 없다는 점과 얼마만큼의 서브그래프를 만들어야 이미지를 제대로 표현하는데 문제가 없는지를 측정할 수 없었다. 훈련 데이터 셋(train set)으로 여러 번의 실험을 수행해야만 어느 정도의 근사값을 구할 수 있었다. 여기에 특징선택(feature selection) 단계가 추가로 수행될 수 있는데, 이는 더 적은 수의 특징들로 동일 혹은 높은 분류결과를 얻는 것이 목표이다. 이전 연구에서는 문서빈도(document frequency)를 적용하였다고 했으나 적용시점이 서브그래프 마이닝 동안인지 직후인지에 대한 설명이 모호하였고, 우리는 두 가지 경우를 모두 실험해 보았다. 두 경우 모두, 이미 앞에서 설명한 문제를 피할 수는 없었지만 전자의 경우 대부분의 특징선택비율(ratio)에서 우리가 채택한 알고리즘(CORK)과 70% 후반대의 비슷한 결과를 보여주었고, 후자의 경우 선택비율이 낮을 수록 50%대의 분류결과를 보여주었다. 본 논문에서 채택한 서브그래프 생성알고리즘은 기존 연구에서 제안한 것과 동일한 전개구조를 가졌지만, 서브그래프가 생성될 때마다 탐욕적 전방탐색(greedy forward selection) 하에서 분류 정확도를 높일 수 있도록 거의 최적(near-optimal)을 보장하는 계산방식이 더해졌으므로, 차별적 그래프들을 자동으로 선별해낼 수 있다. 이 계산 방식을 설명하자면, A, B 두 개의 클래스가 존재할 때, A와 B에 둘 다 존재하거나 둘 다 존재하는 않는 이진 벡터(binary vector)가 많은 서브그래프는 품질점수가 낮아져 제거된다. 흥미로운 사실은 최소 출현빈도수라는 임계값 설정 없이도 이 품질 계산 부등식에 의해 그래프의 가지치기(pruning)가 가능하였다.

또 하나, 본 논문에서는 기존 연구에서 제시한 단어빈도(term frequency)에 의한 정규화보다 백오브워드(bag-of-word) 모델의 특징을 더 잘 나타내는 문서빈도 × 역문서빈도(TF–IDF) 가중치에 의한 정규화를 제안하였고, 대부분 실험결과에서 약1%∼4% 정도 더 높은 분류결과를 보여주었다.

감 사 의 글

이 력 서

이　　　름 : 최 창 민

생 년 월 일 : 1984년 11월 01일

주　　　소 : 대전 유성구 대학로 291 한국과학기술원 전산학과 3440호

E-mail 주 소 : changmin.choi@kaist.ac.kr

학　　력

2003. 3.. – 2011. 8.　한양대학교 경영학부 (B.A.)

2000. 2.. – 2003. 2.　마산중앙고등학교

경　　력

2012. 6. – 2013. 5.　Quality Assurance Tester, Google Korea

2010. 6. – 2010. 9.　Intern, Valtech Consulting