Revista Cubana de Ciencias Informáticas
Vol. 9, No. 1, Enero-Marzo, 2015
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 1-5

Tipo de artículo: Artículo corto
Temática: Inteligencia artificial, Minería de datos, Procesamiento de imágenes, Reconocimiento de patrones
Recibido: 12/09/1014 | Aceptado: 19/01/2015

A Nectar of Frequent Approximate Subgraph Mining for Image Classification

Un nectar sobre la minería de subgrafos frecuentes aproximados en clasificación de imágenes

Niusvel Acosta Mendoza [1*], Andrés Gago Alonso [1], Jesús Ariel Carrasco Ochoa [2], José Francisco Martínez Trinidad [2], José Eladio Medina Pagola [1].

[1] Advanced Technologies Application Center (CENATAV). 7ma A #21406 e/ 214 y 216, Rpto. Siboney, Playa. C.P. 12200. La Habana, Cuba. Correo-e: {agago, jmedina}@cenatav.co.cu

[2] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE). Luis Enrique Erro # 1, Santa María Tonatzintla, 72840 Puebla, PUE, México. Correo-e: {ariel, fmartine}@ccc.inaoep.mx

* Autor para la correspondencia: nacosta@cenatav.co.cu

Abstract

Frequent approximate subgraph mining has emerged as an important research topic where graphs are used for modeling entities and their relations including some distortions in the data. In the last years, there has been a considerable growth in the application of this kind of mining on image classification; achieving competitive results against other approaches. In this nectar, a review of recent contributions on image classification based on frequent approximate subgraph mining is presented. We highlight the usefulness of this type of mining, as well as the improvements achieved in terms of efficiency and efficacy of the proposed frameworks.

Keywords: approximate graph mining, frequent approximate subgraph mining, graph-based image classification.

Resumen

La minería de subgrafos frecuentes aproximados ha emergido como un importante tópico de investigación donde los grafos son usados para modelar entidades y sus relaciones incluyendo distorsiones en los datos. En los últimos años,

se ha observado un considerable crecimiento en la aplicación de este tipo de minería en clasificación de imágenes, donde se han alcanzado resultados competitivos comparados con otros enfoques. En este néctar se presenta una revisión de las contribuciones más recientes en clasificación de imágenes basada en la minería de subgrafos frecuentes aproximados. Se resalta la utilidad de este tipo de minería, así como las mejoras alcanzadas en términos de eficiencia y eficacia del esquema propuesto.
Palabras clave: *clasificación basada en grafos, minería de grafos aproximados, minería de subgrafos frecuentes aproximados.*

Introduction

In practical applications, exact matching between objects is unusual, since distortions, as a general rule, must be taken into account. Thus, in real world applications, where data is represented as graphs, the use of frequent approximate subgraphs (*FASs*), instead of exact ones, can enhance data modeling.

Taking this into account, there have been some approaches (Acosta *et al.*, 2012b; Jiang, *et al.*, 2012) for developing frequent approximate subgraph (*FAS*) miners, considering different kinds of approximation criteria. However, over graph collections, only *VEAM* (***V****ertex and* ***E****dge* ***A****pproximate graph* ***M****iner*) algorithm (Acosta *et al.*, 2012b) allows semantic variations between the labels of vertices and edges; preserving graph topology. In this nectar, an overview of FAS mining for image classification is presented, specifically the approach based on VEAM. In this paper, the results and contributions reached by FAS mining for image classification are summarized, including some of the possible applications where relevant results could be obtained.

The organization of this paper is the following. In Section 2, the VEAM algorithm is briefly described. Later, in Section 3, a review of recent contributions on image classification based on FAS mining is presented. Finally, our conclusions and future work directions are discussed in Section 4.

Methods and Materials

There are many algorithms for computing FASs on graph collections (Acosta *et al.*, 2012b; Jiang *et al.*, 2012), considering several heuristics for graph matching. However, as we have already commented in Section 1, only VEAM (Acosta *et al.*, 2012b) follows the idea that vertex or edge labels sometimes can be replaced by others, considering in this way, data distortions. In order to show the usefulness of VEAM, several applications on image classification tasks have been reported (see Section 3). VEAM processes a graph collection using a depth-first search

Revista Cubana de Ciencias Informáticas
Vol. 9, No. 1, Enero-Marzo, 2015
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 1-5

approach and iteratively extends each FAS by adding an edge. Next, for computing the support, canonical adjacency matrix codes for each candidate isomorphism test are applied.

As we intimated above, in classification tasks, images can be represented as graphs describing their structural and topological characteristics (Morales and García *et al.*, 2011). Thus, the classification framework introduced in the works reviewed in this section is based on FAS mining, this framework is described in Figure 1. The classification process starts at the graph-based representation module, where a pre-labeled image set is represented as a graph collection, and the substitution matrices are built. Next, in the pattern extraction module, a FAS mining algorithm (VEAM) is applied for computing the frequent patterns of this graph collection. Later, in the graph embedding module, the computed FASs are used to build attribute vectors for representing the images. Finally, in the classification module, these vectors are used as input for traditional classifiers.

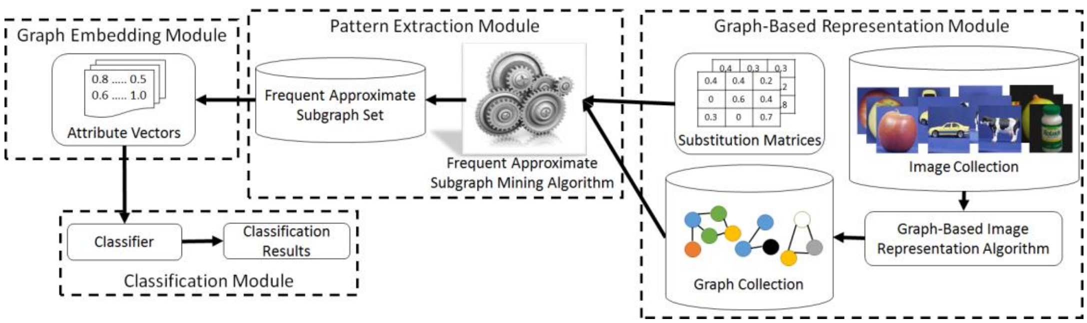

Figure 1. General graph-based image classification framework

Results and Discussion

The first advances applying FAS mining in image classification were achieved in synthetic datasets obtained from a random image generator, and the obtained results show the effectiveness of the proposal (Acosta *et al.*, 2012b) and motivated further research. These promising scores were achieved by VEAM, because its graph approximation criterion properly considers the range of variations between objects in the same class. More recently, good classification results (in terms of *Accuracy* and *F-measure* metrics) were achieved using similar graph-based image classification frameworks on several real image collections such as: *GREC* (http://www.iam.unibe.ch/fki/databases/), *COIL-100* (http://www.cs.columbia.edu/CAVE/databases/) and *ETH-80* (https://www.d2.mpi-inf.mpg.de/datasets/). These collections have a higher level of complexity than synthetic datasets; see the results reported in (Acosta *et al.*, 2012a, Acosta *et al.*, 2012c, Morales *et al.*, 2014).

Revista Cubana de Ciencias Informáticas
Vol. 9, No. 1, Enero-Marzo, 2015
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 1-5

In these works, the patterns computed by VEAM are better for classification tasks than those computed by exact graph miners. However, the number of patterns computed by VEAM is high when the support and similarity thresholds are low. In most of the cases, many of the computed FASs do not provide relevant information. Thence, a selection module was included in (Acosta *et al.*, 2013) with the aim of reducing the cardinality of the FAS set used for classification. In this work, the modified framework uses only representative FAS subsets as attributes, achieving better classification results, reducing the attribute space by the use of already known attribute selection algorithms such as: information gain, chi-square, and gain ratio feature evaluation. This fact allowed an increase in the efficiency and efficacy regarding previous contributions which use the complete set of FASs. On the other hand, in contribution (Acosta *et al.*, 2012c), a way for automatically computing substitution matrices based on image features is proposed; in this way good results are achieved. Later, this contribution is extended in (Morales *et al.*, 2014), where a criterion for selecting the similarity threshold for the mining step is also suggested. These proposals are key steps for the classification approach based on FAS mining, since these parameters are difficult to be fixed by the user. In Table 1, the main characteristics of the aforementioned contributions are summarized.

Table 1. Contributions for image classification based on FAS.

Reference	**Graph Representation**	**Similarity Threshold Calculation**	**Substitution Matrix Calculation**	**Feature Selection**	**Patterns**	**Classifier**
(Acosta *et al.* 2012b)	Quad-tree	Empirically	By expert	No	All FASs	SVM
(Acosta *et al.* 2012a)	Spatial Distribution	Empirically	By expert	No	All FASs	SVM
(Acosta *et al.* 2012c)	Irregular Pyramids	Empirically	Automatically	No	All FASs	SVM
(Acosta *et al.* 2013)	Quad-tree	Empirically	By expert	Yes	FAS subset	6 Classifiers
(Morales *et al.* 2014)	Irregular Pyramids	Automatically	Automatically	No	All FASs	SVM

It is important to highlight that the proposed frameworks can be applied in different domains where the objects under study can be represented as graphs, for example: conceptual maps, ontology, semantic and social networks, Web community analysis, and text classification.

Conclusions

In this paper, the usefulness of considering semantic distortions between graph labels, preserving the topology, is summarized by means of analyzing some recently proposed image classification approaches, based on frequent approximate subgraph (*FAS*) mining. The reported results show good behavior in some artificial and real world image collections, improving the classification accuracy regarding other state-of-the-art solutions. The accuracy in these kinds of tasks was also increased by reducing the set of FASs by applying feature selection algorithms. Thus, a considerable dimensionality reduction is achieved, which improves efficiency of the classification stage; while efficacy is not affected. On the other hand, with the aim of proposing more robust frameworks for image classification, in the most recent contributions, a strategy to automatically determine the similarity threshold and the substitution matrices have also been introduced.

Based on the results by applying feature selection, as future work, we are going to study the identification of only a subset of representative subgraphs specifically only emerging patterns. In this way, we believe that the effectiveness of FAS classifiers will be improved, reducing the runtime classifier at training stage.

References

- ACOSTA, M, N.; GAGO, A, A.; CARRASCO, O, J.A.; MARTÍNEZ, T, J.F. AND MEDINA, P, J.E. Feature Space Reduction for Graph-Based Image Classification. In Proceedings of the CIARP'13, volume Part I, LNCS 8258, pages 246-253, Havana, Cuba, 2013. Springer-Verlag Berlin Heidelberg.

- ACOSTA, M, N.; GAGO, A, A. AND MEDINA, P, J.E. Clasificaión de imágenes utilizando minería de subgrafos frecuentes aproximados. Revista Cubana de Ciencias Informáticas (RCCI), 5(4):1-10, 2012a.

- ACOSTA, M, N.; GAGO, A, A. AND MEDINA, P, J.E. Frequent approxi-mate subgraphs as features for graph-based image classification. Knowledge-Based Systems, 27:381–392, 2012b.

- ACOSTA, M, N.; MORALES, G, A.; GAGO, A, A.; GARCÍA, R, E.B. AND MEDINA, P, J.E. Classification using frequent approximate subgraphs. In Proceedings of the CIARP'12, volume LNCS 7441, pages 292–299. Buenos Aires, Argentina, Springer-Verlag Berlin Heidelberg, 2012c.

- JIANG, C.; COENEN, F. AND ZITO, M. A survey of frequent subgraph mining algorithms. Knowledge Engineering Review, 2012.

- MORALES, G, A.; ACOSTA, M, N.; GAGO, A, A.; GARCÍA, R, E.B. AND MEDINA, P, J.E. A new proposal for graph-based image classification using frequent approximate subgraphs. Pattern Recognition, 47(1):169-177, 2014.

- MORALES, G, A. AND GARCÍA, R, E.B. Simple object recognition based on spatial relations and visual features represented using irregular pyramids. Multimedia Tools and Applications, pages 1-23, 2011.