

Tipo de artículo: Artículo original

Temática: Inteligencia artificial, Procesamiento de imágenes, Reconocimiento de patrones

Recibido: 16/3/2012 | Aceptado: 10/3/2012

Clasificación de imágenes utilizando minería de subgrafos frecuentes aproximados

Image Classification using Frequent Approximate Subgraph Mining

Niusvel Acosta-Mendoza^{1*}, Andrés Gago-Alonso², José E. Medina-Pagola³.

1* Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV). 7ª #21812 e/ 218 y 222, Rpto. Siboney, Playa. C.P. 12200, La Habana, Cuba. Fax: (+) 537.273.0045. nacosta@cenatav.co.cu.

2 Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV). 7ª #21812 e/ 218 y 222, Rpto. Siboney, Playa. C.P. 12200, La Habana, Cuba. Fax: (+) 537.273.0045. agago@cenatav.co.cu.

3 Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV). 7ª #21812 e/ 218 y 222, Rpto. Siboney, Playa. C.P. 12200, La Habana, Cuba. Fax: (+) 537.273.0045. jmedina@cenatav.co.cu.

Resumen: El uso del cotejo aproximado de grafos para la minería de subgrafos frecuentes se ha identificado en diferentes aplicaciones como una necesidad. Con el fin de enfrentar este reto, varios algoritmos de minería han sido desarrollados y han sido utilizados en varios dominios de la ciencia. En este trabajo, se presenta un nuevo esquema para la clasificación de imágenes basado en grafos donde se utiliza un algoritmo para la minería de subgrafos frecuentes. Esta vez el enfoque del cotejo aproximado de grafos nos permite desarrollar clasificadores robustos ante distorsiones presentes en este tipo de datos. La experimentación se realiza sobre una colección real de imágenes representadas en forma de grafos. Los resultados de esta experimentación muestran la utilidad del uso de los algoritmos aproximados y del esquema propuesto.

Palabras clave: minería de grafos aproximados; cotejo aproximado de grafos; representación de imágenes; clasificación de imágenes; selección de características.

Abstract: *The use of approximate graph matching for frequent subgraph mining has been identified in different applications as a need. To meet this need, several mining algorithms have been developed and have been used in different domain of science. In this paper, a new framework for graph-based image classification is presented, where an algorithm for frequent subgraph mining is used. At this time, the graph approximate matching approaches allow us to develop robust classifiers in presence of distortions on this data types. The experimentation is performed on a real image collection where each image is represented as a graph. The results of this experimentation show the usefulness of approximate algorithms and the proposed framework.*

Keywords: *approximate graph mining; approximate graph matching; image representation; image classification; feature selection.*

1. Introducción

En los últimos años se ha incrementado la necesidad de convertir grandes volúmenes de datos en información útil. Los objetos en muchas de estas bases de datos están o pudieran estar representadas como grafos. Como respuesta a esta necesidad, varios autores han desarrollado técnicas y métodos para procesar estas bases de datos (JIMÉNEZ et al., 2010). Un ejemplo de estas técnicas es el descubrimiento de patrones frecuentes (YUN & RYU, 2011).

La minería de patrones frecuentes, especialmente la detección de subgrafos frecuentes en colecciones de grafos es un problema en tareas de minería de grafos (GAGO-ALONSO et al., 2009; NIJSSEN and KOK, 2004; YAN and HUAN, 2002). En minería de subgrafos frecuentes, existen dos enfoques para evaluar la similitud de grafos, conocido como cotejo exacto de grafos y cotejo aproximado de grafos.

El cotejo exacto de grafos ha sido satisfactoriamente utilizado en diversas aplicaciones (EICHINGER and BÖHM, 2010; GAGO-ALONSO et al., 2010; JIANG et al., 2010a); sin embargo, existen problemas concretos donde un cotejo exacto no es aplicable con resultados satisfactorios (HOLDER et al., 1992). En ocasiones, los subgrafos muestran ligeras diferencias en los datos. Un ejemplo de estas diferencias se puede ver en el procesamiento de imágenes, donde estas diferencias pueden estar debido a ruido y distorsiones, o simplemente pueden presentar ligeras diferencias espaciales entre instancias de objetos iguales. Esto significa que se debe tolerar cierto nivel de distorsiones geométricas, variaciones semánticas o desajustes entre vértices o aristas, mientras se realiza la búsqueda de los subgrafos frecuentes. Por este motivo, varios autores han expresado la necesidad del uso del cotejo aproximado entre grafos para la minería de subgrafos frecuentes en colecciones de grafos (BORGELT and BERTHOLD, 2002; HOSSAIN and ANGRYK, 2007; KETKAR et al., 2006; KOYUTÜRK et al., 2004). Estos autores defienden la idea de que se pudieran detectar subgrafos frecuentes con mayor interés para aplicaciones y usuarios.

En respuesta a esta necesidad, varios algoritmos han sido desarrollados para la minería de subgrafos frecuentes utilizando cotejo aproximado entre grafos para diferentes dominios de la ciencia como: análisis de estructuras bioquímicas (CHEN et al., 2007; JIA et al., 2011; XIAO et al., 2008; ZHANG and YANG, 2008; ZOU et al., 2009), redes genéticas regulatorias (SONG and CHEN, 2006); análisis de circuitos, redes sociales, y análisis de vínculos (HOLDER et al., 1992).

Por otro lado, la minería de subgrafos frecuentes se ha utilizado satisfactoriamente en clasificación de imágenes (BAHADIR and SELIM, 2010; JIANG and COENEN, 2008; JIANG et al., 2010b). No obstante, casi todos los enfoques han estado basados en el uso del cotejo exacto.

En este trabajo, se propone un esquema basado en grafos para la clasificación de imágenes. Este esquema utiliza los subgrafos frecuentes como características obtenidas mediante algoritmos de minería de subgrafos frecuentes en una colección de imágenes reales.

Utilizando este esquema se evalúan los algoritmos mediante la clasificación de imágenes. Este trabajo está compuesto por las siguientes secciones. La sección 2 provee algunos conceptos básicos y el estado del arte. El esquema para la clasificación de imágenes es presentado en la sección 3 como un caso de estudio para la evaluación de algoritmos para la minería de subgrafo frecuentes. Los resultados experimentales en una colección de imágenes reales son presentados en la sección 4. Finalmente, las conclusiones de la investigación y algunas ideas sobre las direcciones futuras se exponen en la sección 5.

2. Marco teórico

En esta sección se comenzará con la explicación de los conceptos básicos y las notaciones utilizadas a lo largo de este trabajo y se presentan los algoritmos más relevantes del estado del arte. Finalmente, se plantea el problema de la minería de subgrafos frecuentes aproximados.

2.1 Conceptos básicos

Este trabajo está enfocado en el uso de grafos etiquetados simples y no dirigidos. En lo adelante cuando se hable de grafo se suponen todas estas características y en otro caso se mencionará explícitamente. Antes de presentar su definición formalmente, se define el dominio de etiquetas.

Sean L_V y L_E conjuntos de etiquetas, donde L_V es un conjunto de etiquetas de vértices y L_E es un conjunto de etiquetas de aristas, el dominio de todas las posibles etiquetas es denotado por $L = L_V \cup L_E$.

Un grafo etiquetado en L es una 4-tupla, $G = (V, E, I, J)$, donde V es un conjunto en el que sus elementos son conocidos como vértices, $E \subset \{\{u, v\} | u, v \in V, u \neq v\}$ es un conjunto en el que sus elementos son conocidos como aristas (la arista $\{u, v\}$ conecta el vértice u con el vértice v), $I: V \rightarrow L_V$ es una función etiquetadora que asigna etiquetas a los vértices y $J: E \rightarrow L_E$ es una función etiquetadora que asigna etiquetas a las aristas.

Sean $G_1 = (V_1, E_1, I_1, J_1)$ y $G_2 = (V_2, E_2, I_2, J_2)$ dos grafos etiquetados en L , se dice que G_1 es un subgrafo de G_2 si $V_1 \subseteq V_2, E_1 \subseteq E_2, \forall u \in V_1, I_1(u) = I_2(u)$ y $\forall e \in E_1, J_1(e) = J_2(e)$. En este caso, se usa la notación $G_1 \subseteq G_2$ y se dice que G_2 es un supergrafo de G_1 .

Sea Ω el conjunto de todos los posibles grafos etiquetados en L , la similitud entre dos elementos $G_1, G_2 \in \Omega$ es definida como una función $sim: \Omega \times \Omega \rightarrow \mathbb{R}^+$, donde $\mathbb{R}^+ \subset \mathbb{R}$ es el conjunto de números reales no negativos. Se dice que los elementos son muy diferentes si $sim(G_1, G_2) = 0$ y mientras mayor sea el valor de $sim(G_1, G_2)$ mucho más semejantes (isomorfos) son.

Sea $D = \{G_1, \dots, G_{|D|}\}$ una colección de grafos y G un grafo etiquetado en L , el valor de soporte de G en D se obtiene mediante la siguiente ecuación:

$$sup_G(G, D) = \sum_{G_i \in D} sim(G, G_i) / |D| \quad (1)$$

Cuando $sup_G(G, D) \geq \delta$, entonces el grafo G ocurre frecuentemente en la colección D , siendo G un subgrafo frecuente aproximado en D . El valor del umbral de soporte δ está en $[0,1]$ asumiendo que la similitud está normalizada. La minería de subgrafos frecuentes consiste en encontrar todos los subgrafos conexos frecuentes aproximados en una colección de grafos D , utilizando una función de similitud sim y un umbral de soporte δ .

Existen varias funciones de similitud usadas por diferentes algoritmos en el proceso de cotejo de grafos (CONTE et al., 2004). En la siguiente sección (2.2), se presentan los algoritmos más relevantes del estado del arte los cuales usan técnicas de cotejo de grafos en la minería de subgrafos frecuentes. Todos estos algoritmos utilizan la definición anterior (soporte) implementando una función específica de similitud.

2.2 Estado del arte

Existen varios algoritmos para la minería de subgrafos frecuentes aproximados en colecciones de grafos que utilizan diferentes funciones de similitud en el cotejo de grafos. La minería de subgrafos aproximados puede dividirse en cinco según el enfoque del cotejo: *SUBDUE* (HOLDER et al., 1992) y *RNGV* (SONG and CHEN, 2006) están basados en distancia de edición de grafos, *Monkey* (ZHANG and YANG, 2008) se basa en β -arista sub-isomorfismo; *CSMiner* (XIAO et al., 2008) utiliza el sub-homeomorfismo con vértice/arista disjuntas; *MUSE* (ZOU et al., 2009) se

basa en sub-isomorfismo en grafos inciertos; *gApprox* (CHEN et al., 2007), *APGM* (JIA et al., 2011) y *VEAM* (ACOSTA-MENDOZA et al., 2011) están basados en probabilidades de sustitución. Estos últimos especifican cuáles vértices, aristas o etiquetas pueden reemplazar otras. De este modo, se defiende la idea de que no siempre una etiqueta de vértice o una etiqueta de arista puedan ser sustituidas por cualquier otra.

En el algoritmo *gApprox* se realiza la minería de subgrafos frecuentes aproximados en un solo grafo siendo de interés en este trabajo el procesamiento de colecciones de grafos. Los algoritmos *APGM* y *VEAM* usan matrices de sustitución para realizar la minería de subgrafos frecuentes aproximados en colecciones de grafos. *APGM* solamente trata las variaciones entre el conjunto de etiquetas de los vértices. Por otro lado, *VEAM* realiza el proceso de minería con los conjuntos de etiquetas de vértices y aristas.

En este trabajo, se aplican en la clasificación de imágenes las dos últimas soluciones del último de los grupos de algoritmos anteriores. Esto se debe a la necesidad de un algoritmo eficiente que permita algunas variaciones en los datos utilizando probabilidades de sustitución manteniendo la topología de los grafos.

3. Esquema para la clasificación de imágenes

En esta sección se presenta el esquema propuesto para mostrar la utilidad de los métodos aproximados en tareas de clasificación de imágenes. Los detalles de la clasificación de imágenes son introducidos con el esquema propuesto.

3.1 Esquema propuesto

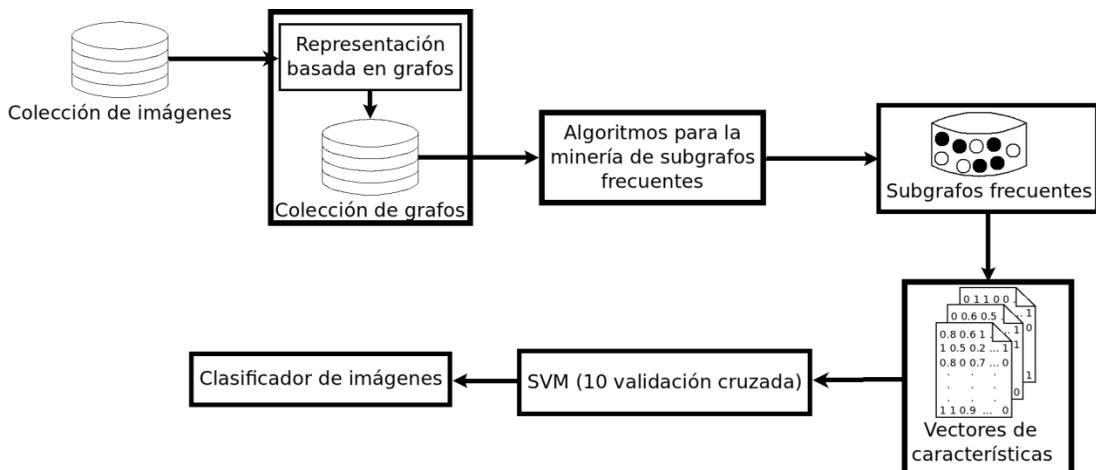


Figura 1. Esquema de clasificación basada en grafos.

Dado un conjunto de imágenes representados en forma de grafos pre-etiquetados propuesto por Riesen y Bunke (RIESEN and BUNKE, 2008) se utilizan los algoritmos para la MSFA con el objetivo de obtener todos los subgrafos frecuentes de esta colección. Luego, estos subgrafos son utilizados para construir los vectores de características de las imágenes originales. Finalmente, se emplea un clasificador usando estos vectores como datos para realizar la clasificación de dichas imágenes. En la Figura 1 se muestra el flujo completo del procedimiento de nuestro experimento.

3.1 Clasificación de imágenes

Teniendo en cuenta los subgrafos frecuentes obtenidos, se construyen los vectores de características, entonces una imagen es representada como un vector de características $V = (v_1, \dots, v_x)$, donde x es el número total de subgrafos

identificados. Por tanto, se construye una matriz donde el número de las filas ($1 \leq i \leq |D|$) corresponde al número de grafos (imágenes) en la colección y el número de las columnas ($1 \leq j \leq x$) corresponde a la cantidad de subgrafos frecuentes (características). Cada valor de característica puede ser asignado utilizando una *configuración binaria* o una *configuración de similitud*.

En la configuración binaria, una celda de la matriz $v_{i,j}$ es $v_{i,j} = 1$ si la característica j ocurre en la imagen i de la colección y en otro caso $v_{i,j} = 0$. Una celda de la matriz $v_{i,j}$ en la configuración de similitud es el mayor valor de similitud de una ocurrencia de la característica j en la imagen i . El valor de la similitud de cada característica se obtiene mediante (1) utilizando la función de similitud que le corresponda a cada algoritmo en particular.

	0	1	2	3	
0	0	1	0	0	(0) Esquina
1	1	0	0	0	(1) Intersección
2	0	1	0	0	(2) Finalización
3	0	0	1	0	(3) Circulo
	3	2	1	0	

MV

Figura 2. Matriz indizada por las etiquetas de los vértices de la colección.

Para la clasificación mediante el clasificador SVM (de sus siglas en inglés, *Support Vector Machine*) se utilizó el paquete libSVM¹. En el caso de APGM, se usa la matriz MV indizada por las etiquetas de los vértices, las cuales representan el tipo de vértice (ver Figura 2). Nótese que los resultados obtenidos mediante APGM serán similares a los de un algoritmo exacto, ya que en la matriz MV no existen posibles sustituciones entre etiquetas. En el caso de VEAM, se utiliza, además de MV , una matriz ME indizada por las etiquetas de las aristas, las cuales son líneas o arcos. Nótese que las líneas no tienen similitud con los arcos ni viceversa. La etiqueta de cada arista e es un índice obtenido mediante la función f_α en correspondencia con el ángulo α entre la arista y la horizontal. Esta función depende del valor de n , donde n es la cantidad de clases que categorizan a los posibles ángulos de modo que van a existir n^2 etiquetas de aristas. Los posibles ángulos cubren 180 grados suponiendo que estos son no dirigidos.

$$f_\alpha(e) = \begin{cases} \lfloor \frac{\alpha * n}{\pi} \rfloor & \text{si es una línea y } 0 \leq \alpha < \pi \\ 0 & \text{si es una línea y } \alpha = \pi \\ \lfloor \frac{\alpha * n}{\pi} \rfloor + n & \text{si es un arco y } 0 \leq \alpha < \pi \\ n & \text{si es un arco y } \alpha = \pi \end{cases} \quad (2)$$

4. Resultados experimentales

En esta sección se muestra, mediante los resultados experimentales, la utilidad del uso de la minería de subgrafos frecuentes aproximados en tareas de clasificación de imágenes. Los resultados de la clasificación utilizando los algoritmos VEAM (ACOSTA-MENDOZA et al., 2011) y APGM (JIA et al., 2011) son comparados. El algoritmo VEAM se escoge como representación de los algoritmos para la MSFA y APGM se comporta como un algoritmo exacto. Para los experimentos se utiliza una colección de 528 imágenes reales, donde cada imagen está representada como un grafo etiquetado.

¹ <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

Todos nuestros experimentos fueron realizados utilizando una computadora personal (Linux 64 bits) Intel(R) Core(TM)2 Duo CPU E7300 @ 2.66GHz con 2 Gb de RAM.

4.1 Colección de grafos

La colección que se utiliza en este trabajo es la “test” de las tres sub-colecciones de imágenes reales (ver Figura 3) representadas en forma de grafos pre-etiquetados propuesto por Riesen y Bunke (RIESEN and BUNKE, 2008). Estas imágenes representan símbolos de los planos arquitectónicos o electrónicos. Las aristas de cada grafo fueron reetiquetadas utilizando la función f_α con $n = 15$.

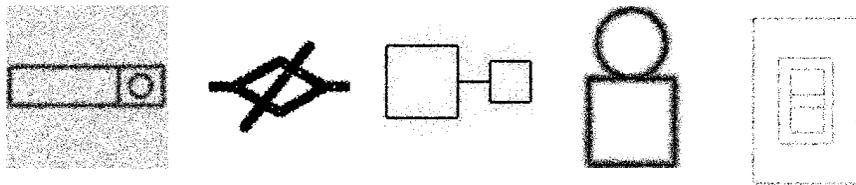


Figura 3. Ejemplo de imágenes de símbolos de los planos arquitectónicos y electrónicos.

La colección está compuesta por 528 grafos divididos en 22 clases. El tamaño promedio de estos grafos es de 11 (en términos de la cantidad de aristas) con 4 etiquetas de vértices y 30 etiquetas de aristas.

4.2 Resultados de la clasificación

Como se mencionó anteriormente en la Sección 3.2, en este trabajo se utiliza el clasificador SVM. En este se emplean dos funciones núcleo: *Lineal* y *Radial Basis Function* (RBF). Como se describe en dicha sección, la clasificación se realiza con vectores de características, donde sus valores se obtienen utilizando la configuración binaria o la configuración de similitud. La primera es usada para los resultados de APGM, mientras que la segunda se utiliza para los resultados de VEAM.

Los resultados de la clasificación son mostrados en las Figuras 4 y 5. La Figura 4 ilustra los resultados obtenidos utilizando RBF como función núcleo y la Figura 5 muestra los resultados con la función núcleo Lineal. La precisión de la clasificación es probada utilizando la validación cruzada con una ventana de tamaño 10 (en inglés, *10 cross-validation*).

Los resultados de las Figuras 4 y 5 se deben a que los patrones identificados por los algoritmos proveen información útil para la clasificación. Además, se puede observar que los patrones obtenidos por VEAM caracterizan mejor las imágenes de esta colección que los encontrados por los métodos exactos. Esto permite que VEAM obtenga mejores precisiones en todos los casos. Este hecho se observa con más claridad en el uso de la función núcleo Lineal debido a los datos utilizados en los experimentos.

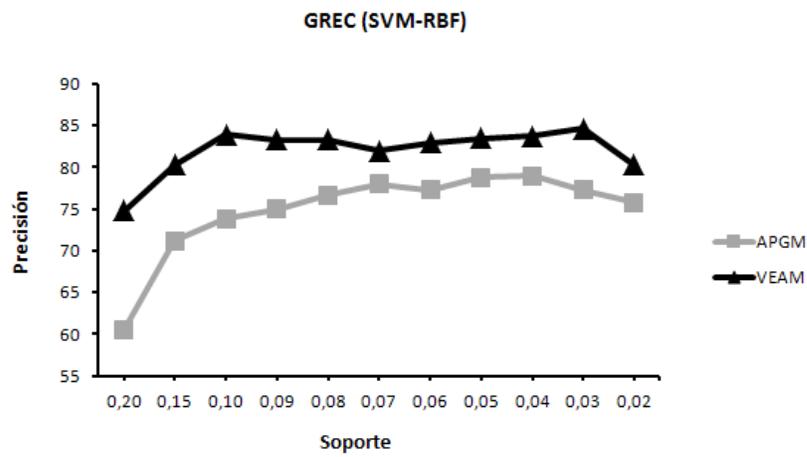


Figura 4. Precisiones obtenidas bajo la función RBF.

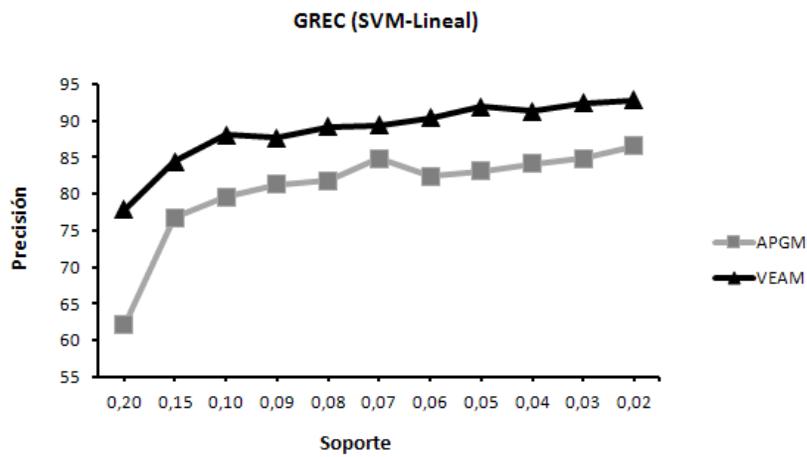


Figura 5. Precisiones obtenidas bajo la función Lineal.

Por otro lado, los resultados de los promedios de reconocimiento reportados en la Segunda Edición de Reconocimiento de Símbolos (DOSCH and VALVENY, 2005) no sobrepasan el 83,33% (ver tabla 1) sobre esta colección utilizando otro esquema de clasificación y sin el uso de la minería. Sin embargo, nuestros resultados alcanzan el 92,8% utilizando la minería en el proceso de la clasificación.

Tabla 1. Resultados generales de cada participante en la Segunda Edición de Reconocimiento de Símbolos.

Participante	Cifra de reconocimiento promedio
Andyardja Weliamto	70,28%
Mín Feng	83,33%
Jing Zhang	67,65%
Wan Zhang	82,82%

5. Conclusiones y trabajo futuro

En este trabajo, proponemos un esquema para la clasificación de colecciones de imágenes. Los resultados experimentales muestran la eficacia de los algoritmos aproximados contra los algoritmos exactos en clasificación de

imágenes. Los patrones detectados por VEAM permiten ligeras diferencias angulares entre las posiciones de los segmentos de las imágenes las cuales no son consideradas en enfoques exactos. Las variaciones semánticas en aristas pueden ser usadas para modelar muchas bases de datos reales. Un ejemplo de la aplicación de los enfoques aproximados se propone en este trabajo para colecciones de imágenes.

Las precisiones resultantes de la clasificación obtenidas con el uso de los algoritmos exactos son menores que utilizando algoritmos aproximados en todos los casos. Por lo tanto, se puede concluir que los patrones identificados por VEAM son más representativos que los exactos detectados por APMG. Finalmente, se muestra la utilidad del uso de los patrones frecuentes aproximados para mejorar la clasificación de imágenes.

Como trabajo futuro, propondremos nuevas medidas de interés para reducir la cantidad de patrones identificados manteniendo o incrementando las precisiones alcanzadas en la clasificación. También desarrollaremos nuevas estrategias de podas para obtener algoritmos más eficientes para grandes colecciones.

Referencias

- ACOSTA-MENDOZA N., GAGO-ALONSO A., MEDINA-PAGOLA J.E. *Frequent approximate subgraphs as features for graph-based image classification*. Knowledge-Based Systems, 2011, 27:381-392.
- BAHADIR O., SELIM A. *Image classification using subgraph histogram representation*. In: Proceedings of the 20th International Conference on Pattern Recognition. Washington, DC, USA: IEEE Computer Society; 2010. p. 1112–1115.
- BORGELT C., BERTHOLD M. *Mining molecular fragments: Finding relevant substructures of molecules*. In: Proceedings of 2002 International Conference on Data Mining (ICDM'02). Maebashi, Japan; 2002. p. 211–218.
- CHEN C., YAN X, ZHU F., HAN J. *gApprox: Mining frequent approximate patterns from a massive network*. In: International Conference on Data Mining. IEEE Computer Society; 2007. p. 445–450.
- CONTE D., FOGGIA P., SANSONE C., VENTO M. *Thirty years of graph matching in pattern recognition*. International Journal of Pattern Recognition and Artificial Intelligence, 2004.
- DOSCH P., VALVENY E. *Report on the second symbol recognition contest*. In: Wenyin L, Lladós J, editors. Graphics Recognition. Ten years review and future perspectives. Proc. 6th Int. Workshop on Graphics Recognition (GREC'05). Springer; LNCS 3926; 2005. p. 381–397.
- EICHINGER F., BÖHM K. *Software-Bug Localization with Graph Mining*. In: Aggarwal CC, Wang H, editors. Managing and Mining Graph Data. Springer-Verlag New York; volume 40; 2010.
- GAGO-ALONSO A., CARRASCO-OCHOA J.A., MEDINA-PAGOLA J.E., MARTÍNEZ-TRINIDAD J.F. *Duplicate candidate elimination and fast support calculation for frequent subgraph mining*. In: Proceedings of the 10th international conference on Intelligent data engineering and automated learning. Berlin Heidelberg: Springer-Verlag; 2009. p. 292–299.
- GAGO-ALONSO A., PUENTES-LUBERTA A., CARRASCO-OCHOA A., MEDINA-PAGOLA J.E., MARTÍNEZ-TRINIDAD J.F. *A new algorithm for mining frequent connected subgraphs based on adjacency matrices*. Intelligence Data Analysis 2010; 14:385–403.

- HOLDER L.B., COOK D.J., BUNKE H. *Fuzzy substructure discovery*. In: Proceedings of the ninth international workshop on Machine learning. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1992. p. 218–223.
- HOSSAIN M.S., ANGRYK R.A. *Gdclust: A graph-based document clustering technique*. In: Proceedings of the Seventh IEEE International Conference on Data Mining Workshops. Washington, DC, USA: IEEE Computer Society; 2007. p. 417–422.
- JIA Y., ZHANG J., HUAN J. *An efficient graph-mining method for complicated and noisy data with realworld applications*. Knowledge Information System. 2011; 28(2):423–447.
- JIANG C., COENEN F. *Graph-based image classification by weighting scheme*. In: Proceedings of the Artificial Intelligence. Springer, Heidelberg; 2008. p. 63–76.
- JIANG C., COENEN F., SANDERSON R., ZITO M. *Text classification using graph mining-based feature extraction*. Knowledge-Based Systems 2010a; 23(4):302–308.
- JIANG C., COENEN F., ZITO M. *Frequent sub-graph mining on edge weighted graphs*. In: Proceedings of the 12th international conference on Data warehousing and knowledge discovery. Berlin, Heidelberg: Springer-Verlag; 2010b. p. 77–88.
- JIMÉNEZ A., BERZAL F., TALAVERA J.C.C. *Frequent tree pattern mining: A survey*. Intelligence Data Analysis 2010; 14(6):603–622.
- KETKAR N., HOLDER L.B., COOK D.J. *Mining in the proximity of subgraphs*. Analysis and Group Detection KDD Workshop on Link Analysis: Dynamics and Statics of Large Networks 2006.
- KOYUTÜRK M., GRAMA A., SZPANKOWSKI W. *An efficient algorithm for detecting frequent subgraphs in biological networks*. Bioinformatics 2004: 200–207.
- NIJSEN S., KOK J.N. *A quickstart in frequent structure mining can make a difference*. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004: 647–652.
- RIESEN K., BUNKE H. *IAMGraph Database Repository for Graph Based Pattern Recognition and Machine Learning*. 2008. p. 287–297.
- SONG Y., CHEN S.S. *Item sets based graph mining algorithm and application in genetic regulatory networks*. Data Mining, IEEE International Conference on Volume, Issue 2006: 337–340.
- XIAO Y., WU W., WANG W., HE Z. *Efficient algorithms for node disjoint subgraph homeomorphism determination*. In: Proceedings of the 13th international conference on Database systems for advanced applications. Berlin, Heidelberg: Springer-Verlag; 2008. p. 452–460.
- YAN X., HUAN J. *gSpan: Graph-based substructure pattern mining*. In: Proceedings International Conference on Data Mining. Maebashi, Japan; 2002. p. 721–724.
- YUN U., RYU K.H. *Approximate weighted frequent pattern mining with/without noisy environments*. Knowledge-Based Systems 2011; 24(1):73–82.
- ZHANG S., YANG J. *RAM: Randomized approximate graph mining*. In: Proceedings of the 20th International Conference on Scientific and Statistical Database Management. ICSSDM; 2008. p. 187–203.

ZOU Z., LI J., GAO H., ZHANG S. *Frequent subgraph pattern mining on uncertain graph data*. In: Proceeding of the 18th ACM conference on Information and knowledge management. New York, NY, USA: ACM; 2009. p. 583–592.