Minería de Patrones Emergentes: Una oportunidad para la extracción evolutiva de conocimiento

Ángel M. García[1], Cristóbal J. Carmona[2], Pedro González[1], and María J. del Jesus[1]

[1] Departamento de Informática, Universidad de Jaén
agvico@ujaen.es
[2] Departamento de Ingeniería Civil, Universidad de Burgos

Resumen La minería de patrones emergentes es una tarea de minería de datos descriptiva bajo el paradigma del aprendizaje supervisado. A lo largo de la literatura se puede encontrar un amplio abanico de propuestas con buenos resultados, sin embargo no existe un amplio estudio enfocado mediante las metaheurísticas. En concreto, sólo encontramos un único modelo que demuestra la calidad de sus resultados frente a los existentes. El estudio se completa con una revisión del estado del arte de esta técnica de minería de datos, destacando las oportunidades de los algoritmos bioinspirados para su resolución.

Keywords: Mineria de patrones emergentes, metaheurísticas, algoritmos evolutivos.

1. Introducción

Tradicionalmente, la minería de datos ha sido utilizada desde dos perspectivas claramente diferenciadas: un enfoque supervisado, cuyo comportamiento es principalmente predictivo y un enfoque no supervisado, cuyo objetivo es la descripción de las relaciones existentes en los datos. Sin embargo, existen técnicas de minería de datos que se encuentran a medio camino entre estos dos enfoques, como las que se agrupan en el denominado descubrimiento de reglas descriptivas basadas en aprendizaje supervisado (SDRD) [24] cuyo objetivo es buscar conocimiento oculto o difícil de obtener por los expertos sobre el valor de una clase prefijada de antemano.

La minería de patrones emergentes (EPM) [6] es una técnica de minería de datos que se puede englobar dentro del SDRD y que permite la obtención de reglas o patrones que sean frecuentes en un valor de la clase objetivo y poco frecuentes en el resto. Por su definición, EPM puede ser utilizado tanto para predecir como para describir comportamiento emergente o diferenciador. Para ello se han utilizado diferentes tipos de heurísticas para la extracción de este tipo de patrones. Sin embargo, a día de hoy se ha desarrollado un único método

basado en algoritmos evolutivos con muy buenos resultados que abre un campo de aplicación prometedor para la aplicación de metaheurísticas en EPM.

En este trabajo se presenta una revisión del estado del arte de EPM ya que no existe un estudio que agrupe todos los conceptos básicos necesarios para introducirse en EPM de una manera sencilla. De esta forma, el artículo se organiza de la siguiente manera: En la Sección 2 se introduce el concepto de EPM y a continuación los diferentes tipos de patrones existentes. La Sección 3 introduce las medidas de calidad descriptivas más utilizadas. La Sección 4 presenta los diferentes métodos desarrollados para EPM y para finalizar se presentan las conclusiones de este trabajo.

2. Minería de Patrones Emergentes

El concepto de EPM fue introducido inicialmente por Dong y Li [6, 7] y definido como:

> "Sea un patrón X cualquiera, y sea $\rho > 1$ un valor de umbral. X se denominará como patrón emergente si y solo si su índice de crecimiento entre dos conjuntos de datos (D_1 y D_2) es mayor o igual a ρ."

Siguiendo la definición, el índice de crecimiento (GR) del patrón X de D_1 a D_2 se define como en la Ecuación 1:

$$GR(X) = \begin{cases} 0, & \text{Si } Sop_1(X) = Sop_2(X) = 0, \\ \infty, & \text{Si } Sop_1(X) = 0 \wedge Sop_2(X) \neq 0, \\ \frac{Sop_2(x)}{Sop_1(x)}, & \text{en otro caso} \end{cases} \quad (1)$$

donde $Sop_i(X)$ es el soporte de dicho patrón en el conjunto de datos i. Los principales objetivos de EPM son la detección de tendencias emergentes en conjuntos de datos marcados temporalmente, la detección de diferencias características entre clases o la detección de diferencias entre múltiples variables.

Habitualmente, la representación de estos patrones emergentes se realiza mediante reglas de la forma:

$$R : Cond \rightarrow Clase \quad (2)$$

donde $Cond$ es un conjunto de pares atributo-valor en conjunción y $Clase$ es el valor de la variable objetivo, la cual será contrastada con el resto de valores de la variable objetivo para medir su índice de crecimiento. De ahora en adelante, con esta representación, un patrón X será identificado como una regla R.

Dentro de la bibliografía especializada existe un amplio abanico de diferentes tipos de patrones emergentes. Esto se debe a que las reglas obtenidas no cumplen la propiedad Apriori [1], ya que pueden existir reglas más específicas que pueden tener mayor índice de crecimiento que reglas más generales. Este hecho hace que el problema sea NP-Duro para un gran número de variables [29]. Por eso mismo, los autores especializados han intentado obtener sólo aquellas reglas que tienen

una gran calidad diferenciadora y que permitan describir el problema de forma rápida y sencilla. De entre los existentes actualmente destacan:

- *Jumping Emerging Patterns* (JEPs). Son patrones emergentes cuyo índice de crecimiento es igual a infinito. Esto quiere decir que los patrones obtenidos definen únicamente a una clase, por lo que poseen un gran poder discriminativo entre clases, permitiendo la obtención de potentes clasificadores [8].

$$JEPs(R) = \{R | GR(R) = \infty\} \tag{3}$$

- *Strong Jumping Emerging Patterns* (SJEPs). Definen un subconjunto de los JEPs que contiene únicamente patrones minimales. Los patrones minimales son los que poseen el mayor poder discriminativo de todos los JEPs [14].

$$SJEPs(R) = \{R | GR(R) = \infty \wedge \nexists Q \subset R \text{ t.q. } GR(Q) = \infty\} \tag{4}$$

- *Maximal Emerging Patterns* (MaxEPs). Al contrario que los patrones minimales, los patrones maximales son patrones en el que ningún superconjunto de R es un EP [30].

$$MaxEPs(R) = \{R | GR(R) \geq \rho \wedge \nexists S \supset R \text{ t.q. } GR(S) \geq \rho\} \tag{5}$$

- EPs tolerantes a ruido (NEPs). Los NEPs relajan un poco la definición de JEPs, permitiendo así capturar patrones que no se podrían obtener si hubiera ruido de clases [14].

$$NEPs(R) = \{R | Sop_1(R) \leq \delta_1 \wedge Sop_2(R) \geq \delta_2\} \tag{6}$$

  En donde normalmente $\delta_2 \gg \delta_1$.
- *Chi Emerging Patterns* (Chi EPs). Son patrones similares a los NEPs, sin embargo en este tipo de patrones se incluye información sobre la capacidad discriminativa de cada par atributo-valor en el patrón mediante un test χ^2. Formalmente, una regla R será Chi EP si [13]:
  1. $Sop(R) \geq \xi$, con ξ como un umbral mínimo de soporte.
  2. $GR(R) \geq \rho$, con ρ como umbral de mínimo índice de crecimiento.
  3. $\nexists Q \subset R$ t. q. $(Sop(Q) \geq \xi) \wedge (GR(Q) \geq \rho) \wedge$ $(Strength(Q) \geq Strength(R))$ donde $Strength(R) = \frac{GR(R)}{GR(R)+1} Sop_2(R)$ [25].
  4. $|R| = 1 \vee |R| > 1 \wedge (\forall Q \subset R \text{ t.q. } |Q| = |R| - 1) \Rightarrow chi(R, Q) \geq \eta$ donde $\eta = 3{,}84$ es un umbral mínimo para chi-cuadrado y la función $chi(R, Q)$ se calcula mediante una tabla de contingencias.
- *Shared Emerging Patterns* (SEPs). Estos patrones van relacionados por pares de EPs. Se utilizan para determinar la similitud entre dos conjuntos de datos cuando no se tienen suficientes datos de entrenamiento. La capacidad de estos patrones no se encuentra en clasificar, sino en identificar conjuntos de datos similares. Un par de patrones R_1 y R_2 son SEPs de los conjuntos de datos D_1 y D_2 respectivamente si [5]:
  1. R_i es un EP para la clase C_1 en $D_i, i = 1, 2$

2. $\frac{|R_1 \cap R_2|}{|R_1 \cup R_2|} \geq \rho$, con ρ como valor umbral
3. $|SopD_1C_1(R_1) - SopD_2C_1(R_2)| < ms \wedge$
   $|SopD_1C_2(R_1) - SopD_2C_2(R_2)| < ms$
   donde $SopD_iC_i(R_i)$ indica el soporte del patrón R_i en el dataset i para la clase i y ms es un valor umbral.

- *Fuzzy Emerging Patterns* (FEPs). Este tipo de patrón utiliza lógica difusa para representar variables de tipo numérico, usando para ello selectores difusos del tipo $[Atributo \in FuzzySet]$. Usando este tipo de patrones se obtienen ventajas en interpretabilidad, ya que son más simples de entender por el experto, y en flexibilidad, ya que los patrones cubren a las instancias con un determinado grado de pertenencia [17].

3. Medidas de Calidad en Minería de Patrones Emergentes

Las medidas de calidad en EPM definen la calidad de un patrón emergente, la calidad de un algoritmo e incluso pueden servir para guiar el proceso de búsqueda. Muchas de estas medidas se derivan del análisis de las propiedades de cobertura de la regla y del consecuente de la misma considerado como clase positiva. Esta relación entre consecuente y antecedente puede ser representada fácilmente mediante una matriz de confusión similar a la de la Tabla 1.

Tabla 1. Matriz de confusión: $TP(R)$ indica los verdaderos positivos, $FN(R)$ los falsos negativos, $FP(R)$ los falsos positivos y $TN(R)$ los verdaderos negativos.

	predicción		
actual	Nº positivos	Nº negativos	
Nº positivos	$p = \lvert TP(R)\rvert$	$\bar{p} = \lvert FN(R)\rvert$	P
Nº negativos	$n = \lvert FP(R)\rvert$	$\bar{n} = \lvert TN(R)\rvert$	N
	$p + n$	$\bar{p} + \bar{n}$	$P + N$

Las medidas más importantes en EPM son:

- Índice de crecimiento (GR): Es la medida que define el concepto de patrón emergente. Esta medida indica el ratio de la diferencia de soporte entre dos datasets D_1, D_2 o entre dos subconjuntos de ejemplos con diferente clase [6]:

$$GR(X) = \begin{cases} 0, & \text{Si } Sop_1(X) = Sop_2(X) = 0, \\ \infty, & \text{Si } Sop_1(X) = 0 \wedge Sop_2(X) \neq 0, \\ \frac{Sop_2(x)}{Sop_1(x)}, & \text{en otro caso} \end{cases} \tag{7}$$

  donde $Sop_i(X)$ es el soporte para el dataset o clase i.
- Soporte (Sop). Se define como la tasa de ejemplos correctamente cubiertos por una regla entre el número total de ejemplos [20]:

$$Sop(R) = \frac{p}{P + N} \tag{8}$$

- Cobertura (Cob). Similar al soporte. Se define como la tasa de ejemplos cubiertos por una regla entre el número total de ejemplos [20]:

$$Cob(R) = \frac{p+n}{P+N} \tag{9}$$

- Confianza (Conf). Se define como el ratio de la capacidad predictiva de una regla para la clase positiva, es decir, el cociente de ejemplos correctamente cubiertos entre el total de ejemplos cubiertos por la regla [15]:

$$Conf(R) = \frac{p}{p+n} \tag{10}$$

- Atipicidad (Atipicidad). Se define como un equilibrio entre la generalidad de una regla y su ganancia de precisión o poder descriptivo [20]:

$$Atipicidad(R) = \frac{p+n}{P+N}\left(\frac{p}{p+n} - \frac{P}{P+N}\right) \tag{11}$$

- Ganacia de información (Ganancia). Se define como la ganancia de información que recibe un usuario con el conjunto de atributos que conforman la regla [31]:

$$Ganancia(R) = \frac{p}{P}\left(log\left(\frac{\frac{p}{P}}{Cov(R)}\right) - log\left(\frac{P}{P+N}\right)\right) \tag{12}$$

En [16] se realiza un estudio comparativo con un amplio conjunto de datos de las diferentes medidas de calidad para EPM. Los resultados del mismo arrojan cuatro grupos de medidas con comportamientos similares en función principalmente de su objetivo:

- Poder discriminativo. Permite obtener el potencial clasificador. En este grupo la métrica mas relevante es la confianza.
- Simplicidad. Que miden cantidad y longitud de las reglas obtenidas.
- Ganancia de información. Que miden la capacidad de una regla para aportar información nueva y/o relevante para el experto. En este grupo la medida mas importante es la atipicidad.
- Generalidad. Que miden la capacidad de abstracción de las reglas. En este grupo la medida más importante es el soporte.

4. Métodos para Minería de Patrones Emergentes

De entre los algoritmos existentes para EPM se puede hacer una clara clasificación según el tipo de estructura de datos y/o heurística utilizada para la obtención de dichos patrones. Estos se dividen en algoritmos basados en límites, basados en representación del conjunto de datos a través de árboles, basados en árboles de decisión y sistemas difusos evolutivos. La Tabla 2 muestra los algoritmos más importantes de la clasificación realizada anteriormente.

Tabla 2. Clasificación de los algoritmos desarrollados para EPM.

Algoritmos basados en límites
CAEP [9]
ConsEPMiner [33]
iCAEP [32]
JEP-C [21]
DeEPS [22]
BCEP [11]
Algoritmos basados en representación mediante árboles
Tree-based JEP-C [3]
iEPMiner [12]
StrongJEP [14]
Top-k minimal JEPs [26]
Algoritmos basados en árboles de decisión
LCMine [18]
CEPMine [19]
EP Random Forest [28]
FEPM [17]
DGCP-Tree [23]
Algoritmos basados en sistemas evolutivos difusos
EvAEP [4]

4.1. Algoritmos Basados en Límites

El concepto de límites es una heurística que permite la representación compacta de grandes cantidades de patrones emergentes introducida junto con el propio concepto de EPM en [6]. Un límite es un par $< \mathcal{L}, \mathcal{R} >$ en el que $\mathcal{L}$ es un conjunto de patrones minimales y $\mathcal{R}$ es un conjunto de patrones maximales que deben cumplir:

1. $\mathcal{L}$ y $\mathcal{R}$ son *anticadenas*. Un conjunto de patrones S es una *anticadena* si $\forall X, Y \in S, X \nsubseteq Y \wedge Y \nsubseteq X$
2. Cada elemento de $\mathcal{L}$ es un subconjunto de algún elemento de $\mathcal{R}$ y cada elemento de $\mathcal{R}$ es un superconjunto de algún elemento de $\mathcal{L}$.

Estos algoritmos emplean una búsqueda eficiente gracias al concepto de límites, cuya heurística se basa en la obtención de EPs mediante diferencias de límites, pero la complejidad computacional de estos algoritmos es exponencial en función del número de variables. Por esta razón muchos de ellos emplean JEPs para reducir el espacio de búsqueda. Estos, a pesar de ser muy discriminativos, son muy sensibles a ruido. A pesar de todos estos inconvenientes, ofrecen resultados competentes de clasificación. Sin embargo, debido al gran número de patrones obtenidos y la representación de los mismos mediante límites hacen que los resultados obtenidos no sean lo suficientemente interpretables y se pierden muchas de las capacidades descriptivas que poseen estos patrones.

El algoritmo referencia de este grupo de métodos es el algoritmo *DeEPS* [22], que utiliza un método de aprendizaje perezoso para extraer JEPs únicamente en el momento de la clasificación de una instancia de test.

4.2. Algoritmos Basados en Representación Mediante Árboles

A pesar de que el concepto de límite permite abordar de manera eficiente EPM, trabajar con límites supone una complejidad exponencial al número de variables. Los algoritmos basados en la representación mediante árboles intentan obtener mediante un recorrido exhaustivo del conjunto de datos una representación más compacta del mismo mediante una estructura de datos en árbol, permitiendo de esta forma una obtención más eficiente de EPs mediante recorridos primero en profundidad y diferentes mecanismos de poda. Esta representación no pretende encontrar todos los EPs existentes, sino un subconjunto de los mismos que contenga EPs interesantes, como por ejemplo, los JEPs. Por esta razón una búsqueda exhaustiva es asumible en este ámbito, pues el número de instancias es mucho menor que los posibles EPs existentes.

Los algoritmos que utilizan esta técnica tienen la posibilidad de discretizar automáticamente las variables de tipo numérico. Sim embargo, son muy sensibles al parámetro μ que indica el umbral de soporte mínimo de un patrón, sufriendo grandes fluctuaciones de precisión en pequeños cambios del parámetro.

La gran diferencia de tiempos respecto a algoritmos basados en límites hacen de estos métodos una gran alternativa para EPM. El algoritmo referencia de este grupo es el algoritmo *StrongJEP* [14], el cual utiliza una estructura de datos en árbol llamada *Contrast Pattern Tree* (CP-Tree), basada en FP-Tree la cual emplea una heurística exhaustiva. Con esta estructura se representan todas las instancias del conjunto de datos así como un conteo de las diferentes ocurrencias de los patrones. El proceso de minería se basa en un recorrido primero en profundidad del árbol, extrayendo únicamente SJEPs.

4.3. Algoritmos Basados en Árboles de Decisión

Los algoritmos encuadrados en esta técnica se basan en la utilización de bosques de árboles de decisión inducidos de los datos de entrenamiento y, a partir de esos árboles, obtener conjuntos de EPs para clasificar. Estos árboles de decisión son modificados para que busquen más candidatos en el espacio de búsqueda, empleando para ello todo tipo de relaciones de igualdad/desigualdad, pertenencia/no pertenencia... y un proceso de búsqueda basada en entornos variables descendentes. Esto obtiene como resultado un conjunto más reducido de EPs, así como más simples, con el fin de mejorar la interpretabilidad de los resultados obtenidos.

Los algoritmos que se encuadran en esta técnica no permiten obtener la totalidad de los EPs, como ocurre en los métodos de las secciones anteriores. Sin embargo, son capaces de obtener un conjunto de patrones de gran calidad que permite obtener resultados competitivos con menores tiempos de ejecución.

El algoritmo referencia de este grupo es el algoritmo *FEPM* [17], que permite la extracción de FEPs. Este método genera una cantidad fija de árboles de decisión generados con un enfoque voraz expandiendo solo los s nodos que particionan mejor el espacio de búsqueda. Una vez se obtienen los árboles y se extraen las reglas emergentes, se filtran para eliminar posibles redundancias.

4.4. Algoritmos Basados en Sistemas Difusos Evolutivos

Los algoritmos evolutivos [2] son metaheurísticas que simulan la evolución natural con el fin de solucionar problemas de optimización. Estas metaheurísticas permiten una gran flexibilidad en cuanto a la representación del conocimiento y la posibilidad de optimizar simultáneamente múltiples medidas de calidad mediante una búsqueda global estocástica que, aunque no aseguran la solución óptima, es capaz de obtener soluciones de gran calidad en tiempos asumibles.

El único algoritmo existente para EPM que emplea esta metaheurística es EvAEP [4], basado en un algoritmo genético mono-objetivo [10] que sigue un enfoque iterativo de aprendizaje de reglas (IRL) [27]. EvAEP obtiene en cada iteración la regla con mayor índice de crecimiento y mayor soporte de ejemplos no cubiertos por reglas anteriores de la misma clase. El algoritmo acaba ante la imposibilidad de extraer más reglas que cumplan estas condiciones. Este algoritmo demostró que supera a los algoritmos referencia de la bibliografía y abre una nueva oportunidad de investigación en aplicación de metaheurísticas a EPM, permitiendo mejoras en tiempo, en calidad e interpretabilidad de las soluciones obtenidas.

5. Conclusiones

En este trabajo se presenta una revisión bibliográfica de la EPM, detallando las distintas medidas de calidad utilizadas, los diferentes tipos de patrones extraídos y los algoritmos más importantes.

A pesar de las potenciales ventajas de la EPM, su enfoque descriptivo ha sido poco explorado en la bibliografía. Se pueden obtener grandes ventajas si se aprovechan las capacidades de estas reglas al utilizar EPM en ámbitos de conocimiento donde se hace necesario un modelo que, además de preciso, sea capaz de justificar las características discriminativas entre clases, como es el caso de la medicina u otros ámbitos relacionados.

El desarrollo de métodos para EPM basados en metaheurísticas avanzadas y/o híbridas como los algoritmos meméticos, búsqueda dispersa o con enfoque multiobjetivo es una línea de trabajo prometedora. Hasta el momento solo se han propuesto métodos evolutivos que han obtenido mejores resultados que con procesos de búsquedas exhaustivas o voraces.

Agradecimientos: Este trabajo ha sido subvencionado por el Ministerio de Economía y Competitividad bajo el proyecto TIN2015-68454-R.

Referencias

1. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I., et al.: Fast discovery of association rules. Advances in knowledge discovery and data mining 12(1), 307–328 (1996)
2. Back, T., Fogel, D.B., Michalewicz, Z.: Handbook of evolutionary computation. IOP Publishing Ltd. (1997)

3. Bailey, J., Manoukian, T., Ramamohanarao, K.: Fast algorithms for mining emerging patterns. In: Principles of Data Mining and Knowledge Discovery, pp. 39–50. Springer (2002)
4. Carmona, C.J., Pulgar-Rubio, F.J., García, A.M., González, P., del Jesus, M.J.: Análisis descriptivo mediante aprendizaje supervisado basado en patrones emergentes. In: Actas de la XVI Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA) (2015)
5. Chen, X., Zhang, W.: Similarity measure by aggregating shared emerging patterns. In: Fifth International Conference on Computational and Information Sciences (ICCIS). pp. 802–805. IEEE (2013)
6. Dong, G., Li, J.: Efficient mining of emerging patterns: Discovering trends and differences. In: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 43–52. ACM (1999)
7. Dong, G., Li, J.: Mining border descriptions of emerging patterns from dataset pairs. Knowledge and Information Systems 8(2), 178–202 (2005)
8. Dong, G., Li, J., Zhang, X.: Discovering jumping emerging patterns and experiments on real datasets. In: Proceedings of th 9th International Database Conference (IDC'99), Hong Kong. pp. 155–168 (1999)
9. Dong, G., Zhang, X., Wong, L., Li, J.: Caep: Classification by aggregating emerging patterns. In: Proceeding of Discovery Science'99. pp. 30–42. Springer (1999)
10. Eiben, A.E., Smith, J.E.: Introduction to evolutionary computing, vol. 53. Springer (2003)
11. Fan, H., Ramamohanarao, K.: A bayesian approach to use emerging patterns for classification. In: Proceedings of the 14th Australasian database conference. vol. 17, pp. 39–48. Australian Computer Society, Inc. (2003)
12. Fan, H., Ramamohanarao, K.: Efficiently mining interesting emerging patterns. In: Advances in Web-Age Information Management, pp. 189–201. Springer (2003)
13. Fan, H., Ramamohanarao, K.: Noise tolerant classification by chi emerging patterns. In: Advances in Knowledge Discovery and Data Mining, pp. 201–206. Springer (2004)
14. Fan, H., Ramamohanarao, K.: Fast discovery and the generalization of strong jumping emerging patterns for building compact and accurate classifiers. IEEE Transactions on Knowledge & Data Engineering (6), 721–737 (2006)
15. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.: Fast Discovery of association rules. the MIT Press (1996)
16. García-Borroto, M., Loyola-González, O., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: Comparing quality measures for contrast pattern classifiers. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 311–318. Springer (2013)
17. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: Fuzzy emerging patterns for classifying hard domains. Knowledge and information systems 28(2), 473–489 (2011)
18. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A., Medina-Pérez, M.A., Ruiz-Shulcloper, J.: Lcmine: An efficient algorithm for mining discriminative regularities and its application in supervised classification. Pattern Recognition 43(9), 3025–3034 (2010)
19. García-Borroto, M., Martínez-Trinidad, J.F., Carrasco-Ochoa, J.A.: A new emerging pattern mining algorithm and its application in supervised classification. In: Advances in Knowledge Discovery and Data Mining, pp. 150–157. Springer (2010)
20. Lavrač, N., Kavšek, B., Flach, P., Todorovski, L.: Subgroup discovery with cn2-sd. The Journal of Machine Learning Research 5, 153–188 (2004)

21. Li, J., Dong, G., Ramamohanarao, K.: Making use of the most expressive jumping emerging patterns for classification. Knowledge and Information systems 3(2), 131–145 (2001)
22. Li, J., Dong, G., Ramamohanarao, K., Wong, L.: Deeps: A new instance-based lazy discovery and classification system. Machine Learning 54(2), 99–124 (2001)
23. Liu, Q., Shi, P., Hu, Z., Zhang, Y.: A novel approach of mining strong jumping emerging patterns based on bsc-tree. International Journal of Systems Science 45(3), 598–615 (2014)
24. Novak, P., Lavrač, N., Webb, G.I.: Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. The Journal of Machine Learning Research 10, 377–403 (2009)
25. Ramamohanarao, K., Fan, H.: Patterns based classifiers. World Wide Web 10(1), 71–83 (2007)
26. Terlecki, P., Walczak, K.: Efficient discovery of top-k minimal jumping emerging patterns. In: International Conference on Rough Sets and Current Trends in Computing. pp. 438–447. Springer (2008)
27. Venturini, G.: Sia: a supervised inductive algorithm with genetic search for learning attributes based concepts. In: European conference on machine learning. pp. 280–296. Springer (1993)
28. Wang, L., Wang, Y., Zhao, D.: Building emerging pattern (ep) random forest for recognition. In: Image Processing (ICIP), 2010 17th IEEE International Conference on. pp. 1457–1460. IEEE (2010)
29. Wang, L., Zhao, H., Dong, G., Li, J.: On the complexity of finding emerging patterns. In: Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC). vol. 2, pp. 126–129. IEEE (2004)
30. Wang, Z., Fan, H., Ramamohanarao, K.: Exploiting maximal emerging patterns for classification. In: AI 2004: Advances in Artificial Intelligence, pp. 1062–1068. Springer (2004)
31. Yin, X., Han, J.: Cpar: Classification based on predictive association rules. In: SDM. vol. 3, pp. 331–335. SIAM (2003)
32. Zhang, X., Dong, G., et al.: Information-based classification by aggregating emerging patterns. In: Intelligent Data Engineering and Automated Learning (IDEAL). Data Mining, Financial Engineering, and Intelligent Agents. pp. 48–53. Springer (2000)
33. Zhang, X., Dong, G., Kotagiri, R.: Exploring constraints to efficiently mine emerging patterns from large high-dimensional datasets. In: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 310–314. ACM (2000)