

Un sistema difuso evolutivo para la detección de excepciones en descubrimiento de subgrupos

C.J. Carmona, P. González, and M.J. del Jesus

Grupo de investigación SIMIDAT, Departamento de Informática, Universidad de Jaén
{ccarmona, pglez, mjjesusolzwarth}@ujaen.es

Resumen El descubrimiento de subgrupos (SD) es una técnica de minería de datos cuyo principal objetivo es buscar descripciones de subgrupos de datos que sean estadísticamente inusuales con respecto a una propiedad de interés. En esta técnica se prefiere obtener reglas generales que describan tantos ejemplos como sea posible, lo que puede provocar que algunos ejemplos se describan incorrectamente. Estos ejemplos negativos se pueden agrupar en excepciones. El trabajo presenta un sistema difuso evolutivo para la detección de excepciones asociadas a reglas previamente obtenidas mediante algoritmos de SD. El objetivo es obtener una nueva descripción con mayor precisión que el subgrupo original. El algoritmo se puede aplicar a los resultados de cualquier algoritmo de SD, y su utilidad se verifica mediante un estudio experimental.

Keywords: Descubrimiento de subgrupos, reglas de excepción, sistemas difusos evolutivos, algoritmos evolutivos multiobjetivo.

1. Introducción

El descubrimiento de subgrupos (*Subgroup Discovery*, SD) [11,16,8] es una técnica de inducción supervisada que obtiene reglas descriptivas mediante aprendizaje supervisado. El objetivo es encontrar reglas interesantes con respecto a una variable de interés, que aporten información desconocida, que confirmen información ya conocida por intuición o que ofrezcan incluso información sorprendente que contradiga el conocimiento de los expertos.

Aunque el conocimiento extraído debe ser sencillo para que sea útil a los expertos, la búsqueda de modelos sencillos en los algoritmos de SD puede provocar una reducción de la precisión si se obtienen reglas generales con ejemplos negativos. La detección de estos ejemplos negativos y su descripción mediante reglas con excepciones [9] puede mejorar el conocimiento extraído sobre la propiedad de interés, no sólo a través de una mejora de la precisión de las reglas de SD sino ofreciendo además conocimiento novedoso y sorprendente a los expertos.

La tarea de SD suele implicar la optimización de diferentes medidas de calidad relacionadas con aspectos de precisión, simplicidad e interés de las descripciones de subgrupos, que suelen representarse mediante reglas. Los algoritmos evolutivos han mostrado un buen comportamiento en procesos de aprendizaje de reglas y problemas multiobjetivo.

Este trabajo presenta un algoritmo de post-procesamiento basado en un sistema difuso evolutivo multiobjetivo para la detección de excepciones en subgrupos. La propuesta busca, a través de un enfoque multiobjetivo, excepciones en subgrupos previamente obtenidos mediante algún algoritmo de SD. El comportamiento del algoritmo se verifica mediante un estudio experimental. El trabajo se organiza de la siguiente forma: la sección 2 describe los conceptos fundamentales utilizados. El algoritmo para la detección de excepciones asociadas a reglas de SD se presenta en la sección 3. La sección 4 detalla el estudio experimental y la sección 5 las conclusiones.

2. Preliminares

2.1. Descubrimiento de subgrupos

El concepto de SD fue introducido por Kloesgen [11] y Wrobel [16]. Se define como, dada una población de individuos y una propiedad de estos individuos en la que estamos interesados, descubrir subgrupos de la población que sean estadísticamente “más interesantes”, es decir, que sean tan grandes como sea posible y que tengan una distribución estadística lo más inusual posible respecto a la propiedad de interés (o variable objetivo).

Por tanto, el objetivo del SD es extraer conocimiento descriptivo acerca de una propiedad de interés de los datos [13]. El conocimiento se representa mediante patrones que deben caracterizar los datos representados de forma que los expertos puedan entenderlos. Así, en SD no es necesario obtener relaciones completas, sino que se buscan relaciones parciales, representadas normalmente mediante reglas de tipo $Cond \rightarrow VarObjetivo$.

Uno de los aspectos más importantes en un algoritmo de SD es el relativo a las medidas de calidad utilizadas para analizar el interés de los subgrupos obtenidos. En la literatura de SD se ha utilizado una gran variedad de medidas de calidad, que pueden ser divididas en distintos grupos dependiendo de su objetivo principal[8]: complejidad, generalidad, precisión e interés. Se puede encontrar una revisión de los algoritmos de SD existentes y sus aplicaciones en [8].

2.2. Minería de reglas de excepción

La mayoría de los métodos de minería de datos se centran en la extracción de reglas generales con alto soporte y confianza. Sin embargo, reglas con bajo soporte podrían aportar conocimiento interesante y sorprendente a los expertos. La minería de reglas de excepción fue introducida por Hussain y otros en [9] como la extracción de reglas con bajo soporte y alta confianza. En este sentido, una excepción es interesante cuando se diferencia del resto [14].

Se pueden distinguir dos enfoques al buscar reglas de excepción [9]: dirigido (o subjetivo) que obtiene un conjunto de reglas de excepción, cada una de las cuales contradice una creencia especificada por el experto; y no dirigido (u objetivo) que obtiene un conjunto de pares de regla general / excepción.

Aplicado a los resultados de cualquier algoritmo de SD, la detección de excepciones podría llevar a una mejora en la precisión y en la descripción, puesto que se descubrirían dentro del subgrupo pequeñas áreas con ejemplos incorrectamente descritos o negativos.

3. Un sistema difuso evolutivo multiobjetivo para la detección de excepciones en subgrupos

Se presenta aquí un algoritmo de post-procesamiento para la detección de excepciones en SD. El objetivo es buscar, dentro de un subgrupo dado, pequeñas áreas que describan excepciones al conocimiento representado por este subgrupo, formadas por ejemplos cuyo valor de la variable objetivo sea distinto al del subgrupo inicial. El concepto se explica en la figura 1, que representa mediante un círculo sombreado en gris un subgrupo correspondiente al valor o de la variable objetivo. Este subgrupo cubre todos los ejemplos del valor o , además de algunos ejemplos del otro valor de la variable objetivo (valor x). A pesar de cubrir algunos ejemplos negativos, la representación del conocimiento utilizada es simple y puede ser muy útil para los expertos. Estos ejemplos negativos (marcados en un rombo sombreado en gris oscuro) constituyen excepciones del subgrupo. La nueva regla, formada por el subgrupo inicial junto con sus excepciones, aporta una mejora en la descripción y un incremento de la precisión.

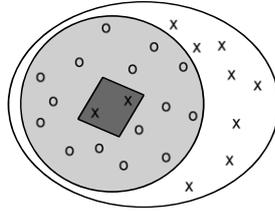


Figura 1. Detección de excepciones dentro de un subgrupo

Para lograr este objetivo, se propone un sistema difuso evolutivo (*Evolutionary Fuzzy System*, EFS) multiobjetivo [7]. Los algoritmos evolutivos son métodos de búsqueda de propósito general que se han demostrado muy eficaces en procesos de aprendizaje de reglas, y la utilización de un enfoque multiobjetivo resulta muy adecuada cuando se deben optimizar diferentes medidas de calidad, como ocurre en SD.

El sistema parte de un conjunto de i subgrupos obtenidos por cualquier algoritmo de SD ($S_R = \{R_1, R_2, \dots, R_i\}$) representados por reglas del tipo:

$$R_i : IF Cond_i THEN VarObjetivo \quad (1)$$

en el que $Cond_i$ representa el antecedente de la regla i .

A cada una de las reglas o subgrupos se les aplica el proceso evolutivo, obteniendo una nueva regla que incorpora las excepciones encontradas. El resultado final, una vez aplicado el proceso a todos los subgrupos iniciales, es un conjunto de subgrupos con excepciones, uno por cada subgrupo del conjunto inicial.

El algoritmo puede utilizarse sobre conjuntos de datos con variables continuas y/o categóricas, obteniendo subgrupos modificados que serán del mismo tipo (difusos y/o nítidos) que los originales. Las características principales de la propuesta se presentan a continuación.

3.1. Representación de los individuos

Los individuos de la población son excepciones al subgrupo inicial. Se utiliza un modelo de representación entero con tantos genes como variables tiene el conjunto de datos original, sin considerar la variable objetivo. La codificación se realiza según el enfoque “*Cromosoma = Regla*” [7], en que que el cromosoma sólo representa el antecedente.

El algoritmo puede trabajar con dominios difusos y/o nítidos (y por tanto con reglas difusas y/o nítidas) obteniendo subgrupos modificados del mismo tipo. La representación indicada se utiliza tanto para variables categóricas como continuas. Para variables continuas, el algoritmo utiliza la lógica difusa para manejar las variables continuas sin necesidad de aplicar una discretización previa, ya que esto que podría ocasionar una pérdida de calidad. Así, las variables continuas se consideran como variables lingüísticas, y los conjuntos difusos correspondientes a las variables lingüísticas son los definidos por los algoritmos de SD para obtener los subgrupos iniciales. La figura 2 muestra un ejemplo de variable con particiones difusas uniformes y funciones de pertenencia triangulares correspondientes a cinco etiquetas lingüísticas (*Muy Bajo*, *Bajo*, *Medio*, *Alto*, *Muy Alto*). En el caso de variables categóricas, cada gen estará compuesto por un valor entero que codifique el valor de la variable correspondiente.

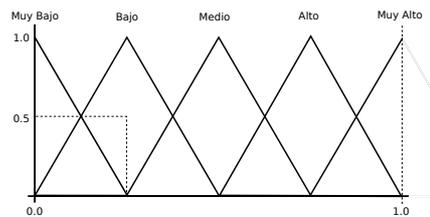


Figura 2. Uso de la lógica difusa para una variable con cinco etiquetas lingüísticas

En el individuo no se representa el consecuente de la regla puesto que está prefijado. Como los individuos de la población representan excepciones al subgrupo inicial, el valor de la variable objetivo de los individuos debe ser distinto al del subgrupo inicial (si la variable objetivo tiene dos clases, las excepciones tendrán

el valor contrario al del subgrupo inicial; si tiene más de dos clases, el valor de la variable objetivo de las excepciones será cualquiera distinto al del subgrupo inicial). Por lo tanto, si el valor de la variable objetivo del subgrupo original es $\overline{VarObjetivo}$, el valor de la variable objetivo de las excepciones es $\overline{VarObjetivo}$.

Cada regla de excepción de un subgrupo mantiene los valores de los atributos del subgrupo original, además de incorporar nuevos atributos. Consideremos el siguiente ejemplo de regla difusa de descripción de subgrupos: $IF\ x_1 = Medio\ AND\ x_3 = Bajo\ THEN\ \overline{VarObjetivo}$. Una excepción a este subgrupo, como $IF\ x_1 = Medio\ AND\ x_3 = Bajo\ AND\ x_5 = Medio\ THEN\ \overline{VarObjetivo}$, mantiene los valores del subgrupo original e incorpora nuevos atributos. Por lo tanto, las excepciones son reglas más específicas que los subgrupos originales de los que provienen.

3.2. Algoritmo evolutivo multiobjetivo

El algoritmo busca, dentro del espacio delimitado por el antecedente de la regla, pequeños grupos de ejemplos con un valor de la variable objetivo distinto al establecido en el consecuente del subgrupo; es decir, busca conjuntos de ejemplos incorrectamente descritos por el subgrupo. Esta búsqueda se realiza mediante un algoritmo evolutivo multiobjetivo que utiliza el enfoque NSGA-II [4], de acuerdo con un proceso en dos pasos:

1. *Detección de conjuntos de excepciones asociadas al subgrupo.* Se aplica el algoritmo para extraer excepciones formadas por los mismos valores de las variables presentes en el subgrupo además de otras variables adicionales, y correspondientes a un valor distinto de la variable objetivo del subgrupo.
2. *Generación de los subgrupos modificados.* Se obtiene el subgrupo modificado mediante la exclusión de las excepciones encontradas en el paso anterior:

$$R'_i : IF\ Cond_i\ AND\ \overline{Exc_i}\ THEN\ \overline{VarObjetivo} \quad (2)$$

donde Exc_i representa las condiciones de las excepciones asociadas a la regla R_i (sin las variables comunes).

El proceso evolutivo termina cuando el algoritmo alcanza un número establecido de evaluaciones, y devuelve como resultado el conjunto de individuos no repetidos del frente de Pareto que alcancen un umbral establecido de confianza. El algoritmo se ejecuta para cada subgrupo inicial (obteniendo subgrupos con excepciones) por lo que el número de subgrupos modificados obtenido es el mismo que el de subgrupos iniciales.

El algoritmo utiliza expresiones adaptadas de las medidas de sensibilidad [11] y confianza [5] tanto para evaluar las excepciones como para guiar el proceso de aprendizaje:

- **Sensibilidad.** Mide la proporción de verdaderos positivos que son correctamente identificados:

$$Sens(Exc) = \frac{TP}{TP + FN} = \frac{n(\overline{VarObjetivo} \cdot Cond)}{n(\overline{VarObjetivo})} \quad (3)$$

donde $n(\overline{VarObjetivo})$ son los ejemplos que tienen el valor contrario de la variable objetivo del subgrupo inicial, y $n(\overline{VarObjetivo} \cdot Cond)$ son los ejemplos cubiertos por la excepción con el valor contrario de la variable objetivo al del subgrupo.

- **Confianza difusa.** Mide la proporción de ejemplos correctamente identificados respecto al valor contrario de la variable objetivo:

$$FCnf(Exc) = \frac{\sum_{E^k \in E / E^k \in \overline{VarObjetivo}} APC(E^k, Exc)}{\sum_{E^k \in E} APC(E^k, Exc)} \quad (4)$$

donde $E^k = (e_1^k, e_2^k, \dots, e_v^k), VarObjetivo^k) / k = 1, \dots, N$, $VarObjetivo^k \in T$ es un ejemplo del conjunto E , v es el número de variables del ejemplo, e_v^k es el valor de la variable v en el ejemplo E^k , $VarObjetivo^k$ es el valor de la variable objetivo para el ejemplo E^k y APC es el grado de compatibilidad entre un ejemplo y el antecedente de una regla de excepción (cuando se aplica en dominios nítidos, el grado de compatibilidad entre un ejemplo y el antecedente de una regla es 0 ó 1).

3.3. Operadores genéticos

El algoritmo evolutivo de post-procesamiento incluye operadores de selección por torneo y cruce multi-punto, además de otros operadores específicos:

- *Inicialización orientada.* Se genera una población de individuos que contienen los mismos valores del subgrupo inicial, además de nuevos valores para el resto de variables. Para conseguir esto, parte de los individuos de la población se generan de forma sesgada, y el resto de forma aleatoria. Los valores de las variables que forman parte del subgrupo inicial se copian directamente en los nuevos individuos de la población, mientras que el resto de valores se obtienen de la siguiente forma: los valores del 75% de los individuos se generan considerando que sólo un máximo del 90% de las variables pueden formar parte de la regla; para el resto (25%) los valores se generan de forma aleatoria. Este operador permite comenzar el proceso evolutivo con un conjunto de reglas muy específicas puesto que los individuos generados son reglas en las que intervienen un alto número de variables.
- *Mutación orientada.* Se deriva de la mutación estándar [6], pero impidiendo la modificación de los valores que formaban parte del subgrupo inicial. Además, la mutación del valor de una variable que no formaba parte del subgrupo inicial no siempre implica su eliminación, puesto que se asigna a esta variable uno de los valores posibles que sea diferente del actual.
- *Reinicialización orientada basada en cobertura.* El algoritmo utiliza una modificación del operador definido para el algoritmo NMEEF-SD [3]. En el operador original, se realiza una comprobación antes de generar la población de la siguiente generación, para ver si el Pareto evoluciona o no. Si el Pareto no evoluciona, todos los individuos no repetidos del Pareto se introducen en

la población de la siguiente generación y el resto de individuos hasta completar la población se generan de forma que cubran ejemplos del conjunto de datos no cubiertos por los individuos que forman el Pareto. La modificación utilizada consiste en que los individuos generados deben mantener los mismos valores en las variables que intervenían en el subgrupo inicial, y por tanto serán más específicos. Los individuos generados se asocian a ejemplos no cubiertos por el Pareto, al seleccionar aleatoriamente un ejemplo no cubierto y codificar los valores del individuo con respecto a este ejemplo.

4. Estudio experimental

La experimentación se ha realizado con conjuntos de datos del repositorio KEEL¹ [2], cuyas propiedades se muestran en la tabla 1 (donde n_v es el número de variables, n_{vD} el número de variables discretas y n_{vC} el de variables continuas y N el número de ejemplos). Se ha utilizado validación cruzada con 10 particiones, y tres ejecuciones para los algoritmos no determinísticos.

Tabla 1. Conjuntos de datos utilizados del repositorio KEEL

#	Nombre	n_v	n_{vD}	n_{vC}	N
1	Australian	14	8	6	690
2	Crx	15	12	3	690
3	Heart	13	6	7	270
4	Hepatitis	19	13	6	155
5	Monk-2	6	6	0	432
6	Housevotes	16	16	0	435
7	Saheart	9	4	5	462
8	Tic-tac-toe	9	9	0	958

Para mostrar las ventajas de la propuesta se utilizan dos conocidos algoritmos de SD para obtener los subgrupos iniciales: NMEEF-SD [3] y Apriori-SD [10]. A los resultados de estos algoritmos se les aplica la propuesta para obtener subgrupos con excepciones. La evaluación de los subgrupos con excepciones se realiza mediante expresiones adaptadas para las medidas de calidad de relevancia (*Rel*) [11], atipicidad (*Ati*) [12], sensibilidad (*Sen*) [11] y confianza (*Cnf*) [1].

- Relevancia de un subgrupo con excepciones (R'):

$$Rel'(R'_i) = 2 \cdot \sum_{k=1}^{n_c} (TP_{R'_i})_k \cdot \log \frac{(TP_{R'_i})_k}{((TP + FN)_{R'_i})_k \cdot \frac{((TP + FP)_{R'_i})_k}{N}} \quad (5)$$

donde $TP_{R'_i} = TP_{R_i} - FP_{Exc_i}$, TP_{R_i} es el número de ejemplos correctamente descritos de la regla, FP_{Exc_i} es el número de ejemplos incorrectamente descritos por el conjunto de excepciones asociadas a la regla, $(TP + FN)_{R_i}$ es el número de ejemplos para los valores de la variable objetivo, $(TP + FP)_{R'_i} =$

¹ <http://www.keel.es>

$(TP + FP)_{R_i} - (TP + FP)_{Exc_i}$, $(TP + FP)_{R_i}$ es el número de ejemplos cubiertos por la regla y $(TP + FP)_{Exc_i}$ es el número de ejemplos cubiertos por el conjunto de excepciones asociadas a la regla inicial.

- Atipicidad de un subgrupo con excepciones:

$$Atip'(R'_i) = \left(\frac{TP_{R'_i}}{(TP + FP)_{R'_i}} - \frac{(TP + FN)_{R_i}}{N} \right) \cdot \frac{(TP + FP)_{R'_i}}{N} \quad (6)$$

- Sensibilidad de un subgrupo con excepciones:

$$Sens'(R'_i) = \frac{TP_{R'_i}}{(TP + FN)_{R_i}} \quad (7)$$

- Confianza difusa de un subgrupo con excepciones:

$$FCnf'(R'_i) = \frac{\sum_{E^k \in E/E^k \in VarObjetivo} APC(E^k, R'_i)}{\sum_{E^k \in E} APC(E^k, R'_i)} \quad (8)$$

donde $APC(E^k, R'_i) = APC(E^k, R_i) - APC(E^k, Exc_i)$.

La tabla 2 muestra la media de los resultados de los algoritmos de SD y la aplicación posterior del post-procesamiento propuesto, donde n_r representa la media del número de subgrupos obtenidos, n_v es la media del número de variables para cada subgrupo y los valores de las medidas de calidad indicadas (*REL*, *ATI*, *SEN* y *CNF*) son las medias de los resultados de las distintas reglas de los conjuntos de reglas. Por cuestiones de espacio, sólo se muestran aquí los resultados medios y los de los tests estadísticos realizados.

Tabla 2. Resultados obtenidos por los algoritmos

<i>Algoritmo</i>	n_r	n_v	<i>REL</i>	<i>ATI</i>	<i>SEN</i>	<i>CNF</i>
Apriori-SD	5.70	2.30	3.219	0.092	0.587	0.708
Apriori-SD+Exceptions	9.36	7.64	3.559	0.102	0.567	0.740
NMEEF-SD	4.45	2.59	3.748	0.113	0.820	0.824
NMEEF-SD+Exceptions	11.35	7.58	4.522	0.126	0.788	0.864

Para el análisis y comparación de los resultados de los distintos algoritmos, se ha utilizado el test de Wilcoxon [15]. Se ha utilizado un nivel de confianza de $\alpha = 0,10$ en todos los experimentos. La tabla 3 muestra los resultados del test de Wilcoxon para cada medida de calidad con los correspondientes *p - val*, y el resultados de la *Hipótesis*. En general, los resultados obtenidos muestran diferencias significativas en la mayoría de las medidas de calidad a favor de los resultados una vez aplicado el enfoque de post-procesamiento propuesto.

Como muestran las tablas 2 y 3, los resultados después de aplicar el post-procesamiento mejoran los obtenidos por los algoritmos de SD. Se obtienen pequeñas reducciones de los valores de sensibilidad de los algoritmos originales

Tabla 3. Test de Wilcoxon para la comparación de Apriori-SD/NMEEF-SD+Excepciones Vs. Apriori-SD/NMEEF-SD

<i>Algoritmo</i>		<i>p - val</i>	<i>Hipótesis</i>
Apriori-SD	<i>REL</i>	0.092	Rechazado por Apriori-SD+Exceptions
	<i>ATI</i>	0.030	Rechazado por Apriori-SD+Exceptions
	<i>SEN</i>	0.017	Rechazado por Apriori-SD
	<i>CNF</i>	0.025	Rechazado por Apriori-SD+Exceptions
NMEEF-SD	<i>REL</i>	0.021	Rechazado por NMEEF-SD+Exceptions
	<i>ATI</i>	0.028	Rechazado por NMEEF-SD+Exceptions
	<i>SEN</i>	0.018	Rechazado por NMEEF-SD
	<i>CNF</i>	0.012	Rechazado por NMEEF-SD+Exceptions

en comparación con los resultados de esta propuesta en todos los experimentos. Puesto que la sensibilidad cuantifica la relación de ejemplos cubiertos por la variable objetivo, el subgrupo con excepciones podrá obtener como máximo el mismo resultado que el subgrupo inicial; por tanto, no se pueden mejorar los resultados en sensibilidad ya que los subgrupos modificados sólo miden los ejemplos del valor de la variable objetivo del subgrupo inicial. Para el resto de medidas de calidad, el uso de excepciones mejora la precisión y el interés de los subgrupos iniciales.

5. Conclusiones

En este trabajo se ha presentado un algoritmo de post-procesamiento basado en un EFS multiobjetivo que permite mejorar los subgrupos obtenidos por cualquier algoritmo de SD. El objetivo es detectar excepciones dentro de los subgrupos, con dos ideas: por un lado, describir pequeños espacios en los datos con un comportamiento distinto dentro del subgrupo; y por otro, incrementar la precisión de los subgrupos al detectar ejemplos dentro de los subgrupos que pueden ser interesantes para los expertos.

El estudio experimental desarrollado, soportado por tests estadísticos, muestra que la aplicación del algoritmo de post-procesamiento a los resultados obtenidos por un algoritmo de SD permite que los valores de las medidas de calidad (confianza y sensibilidad) de los subgrupos modificados con las excepciones encontradas mejoren respecto a los obtenidos por los subgrupos iniciales. Además, no sólo se mejoran los valores de las medidas de calidad utilizadas en el proceso evolutivo, sino también otras medidas de calidad consideradas en SD. El algoritmo es plenamente aplicable a problemas reales en los que los expertos necesiten obtener información para mejorar la descripción de los datos disponibles.

Reconocimientos

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad bajo el proyecto TIN2012-33856 (Fondos FEDER), y por el Plan Andaluz de Investigación bajo el proyecto TIC-3928 (Fondos FEDER).

Referencias

1. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo, *Fast discovery of association rules*, Advances in Knowledge Discovery and data mining (U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds.), AAAI Press, 1996, pp. 307–328.
2. J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, *KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework*, Journal of Multiple-Valued Logic and Soft Computing **17** (2011), no. 2-3, 255–287.
3. C. J. Carmona, P. González, M. J. del Jesus, and F. Herrera, *NMEEF-SD: Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery*, IEEE Transactions on Fuzzy Systems **18** (2010), no. 5, 958–970.
4. K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, IEEE Transactions Evolutionary Computation **6** (2002), no. 2, 182–197.
5. M. J. del Jesus, P. González, F. Herrera, and M. Mesonero, *Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing*, IEEE Transactions on Fuzzy Systems **15** (2007), no. 4, 578–592.
6. D. E. Goldberg, *Genetic Algorithms in search, optimization and machine learning*, Addison-Wesley Longman Publishing Co., Inc., 1989.
7. F. Herrera, *Genetic fuzzy systems: taxonomy, current research trends and prospects*, Evolutionary Intelligence **1** (2008), 27–46.
8. F. Herrera, C. J. Carmona, P. González, and M. J. del Jesus, *An overview on Subgroup Discovery: Foundations and Applications*, Knowledge and Information Systems **29** (2011), no. 3, 495–525.
9. F. Hussain, H. Liu, E. Suzuki, and H. Lu, *Exception rule mining with a relative interesting measure*, Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNAI, vol. 1805, 2000, pp. 86–97.
10. B. Kavsek and N. Lavrac, *APRIORI-SD: Adapting association rule learning to subgroup discovery*, Applied Artificial Intelligence **20** (2006), 543–583.
11. W. Kloesgen, *Explora: A Multipattern and Multistrategy Discovery Assistant*, Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence, 1996, pp. 249–271.
12. N. Lavrac, P. A. Flach, and B. Zupan, *Rule Evaluation Measures: A Unifying View*, Proceedings of the 9th International Workshop on Inductive Logic Programming, LNCS, vol. 1634, Springer, 1999, pp. 174–185.
13. N. Lavrac, B. Kavsek, P. A. Flach, and L. Todorovski, *Subgroup Discovery with CN2-SD*, Journal of Machine Learning Research **5** (2004), 153–188.
14. E. Suzuki, *Data mining methods for discovering interesting exceptions from an unsupervised table*, Journal of Universal Computer Science **12** (2006), no. 6, 627–653.
15. F. Wilcoxon, *Individual comparisons by ranking methods*, Biometrics **1** (1945), 80–83.
16. S. Wrobel, *An Algorithm for Multi-relational Discovery of Subgroups*, Proceedings of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, LNAI, vol. 1263, Springer, 1997, pp. 78–87.