SDRDPy: Una aplicación para la visualización gráfica del conocimiento obtenido con algoritmos de reglas descriptivas supervisadas

María Asunción Padilla Rascón[1,2] [0009−0000−4033−7044] mprascon@ujaen.es, Pedro González[1,2] [1111−2222−3333−4444] pglez@ujaen.es, and Cristóbal J. Carmona[1,2] [2222−−3333−4444−5555] ccarmona@ujaen.es

[1] Department of Computer Science, University of Jaén, E-23071 Jaen (Spain)
[2] Andalusian Research Institute in Data Science and Computational Intelligence, University of Jaén, E-23071 Jaen (Spain)

Abstract. SDRDPy es una aplicación de escritorio que proporciona a los expertos una representación gráfica y tabular intuitiva del conocimiento extraído por cualquier algoritmo de Descubrimiento de Reglas Descriptivas Supervisadas (DRDS). La aplicación facilita el análisis de los datos mediante la visualización de información relevante del conjunto de datos, así como de las relaciones entre las reglas descubiertas, los ejemplos y las medidas de calidad asociadas a cada regla, independientemente de la herramienta utilizada para ejecutar el algoritmo. Toda esta información se presenta a través de una interfaz amigable que facilita el análisis experto y la generación de informes exportables en diversos formatos. El trabajo completo ha sido publicado en SoftwareX [8].

Keywords: Reglas descriptivas supervisadas · Descubrimiento de subgrupos · Patrones emergentes · UI

1 Introducción

En la actualidad, la disciplina de la Ciencia de Datos está experimentando un gran auge debido al crecimiento continuo de la información recopilada digitalmente. Gracias a ella, es posible extraer conocimiento relevante para la toma de decisiones [5]. El proceso de Ciencia de Datos comprende diversas fases: recopilación y selección de datos, selección del modelo o estrategia, evaluación y, finalmente, difusión y uso de los resultados. Sin embargo, existe una carencia notable en esta última fase: no se ha invertido el mismo esfuerzo en el desarrollo de interfaces usables y amigables que permitan a expertos de otros dominios interpretar los resultados obtenidos sin necesidad de la asistencia de un especialista en Ciencia de Datos. Este problema se acentúa especialmente en el ámbito del Descubrimiento de Reglas Descriptivas Supervisadas (DRDS) [2], que engloba técnicas como el Descubrimiento de Subgrupos [7], la Minería de Patrones Emergentes [6], la Minería de Conjuntos de Contraste [1] y la Minería de Cambios [4]. Estas técnicas combinan inducción descriptiva con aprendizaje supervisado y se han aplicado con éxito en dominios como la medicina, la industria o el comercio

electrónico. No obstante, los expertos en estos dominios carecen de herramientas que les permitan visualizar los resultados de los algoritmos DRDS de forma intuitiva y atractiva.

Las herramientas comerciales y académicas populares presentan limitaciones severas para este fin (véase Tabla 1): plataformas como Weka o KNIME no incluyen algoritmos DRDS, mientras que KEEL y Orange, aunque los incorporan, carecen de funciones de visualización e interacción entre reglas y datos. Además, todas ellas exigen adaptar los resultados al formato propio de la plataforma.

Con el objetivo de cubrir esta laguna, se presenta SDRDPy, una aplicación de escritorio desarrollada en Python que permite visualizar gráfica y tabularmante los resultados de cualquier algoritmo DRDS, sin necesidad de reejecutar el algoritmo ni adaptar su código. La aplicación procesa directamente los ficheros de reglas generados por el algoritmo y los presenta de forma clara al usuario experto, ofreciendo tanto la visualización de las reglas como el análisis de su relación con los datos del conjunto de estudio.

Características	**SDRDPy**	**KEEL**	**Weka**	**Orange**	**Knime**
Algoritmo de reglas descriptivas SDRD	Sí	Sí	No	Sí	No
Visualización de resultados del algoritmo SDRD	Sí	No	No	Sí	No
Inclusión del algoritmo SDRD	No, solo evaluación	Sí	No	Sí	Sí
Ver relación entre reglas y datos	Sí	No	No	No	No
Adaptación del resultado del algoritmo a la plataforma	No	Sí	No	Sí	No

Table 1: Comparativa de características entre diferentes herramientas

2 Descripción del software

SDRDPy es una aplicación de escritorio implementada en Python 3.11 con dependencias mínimas (Tkinter, Matplotlib y Numpy), que permite visualizar los resultados obtenidos por cualquier algoritmo DRDS.

La arquitectura del sistema, mostrada en la Figura 1, sigue un esquema sencillo centrado en el usuario. Éste importa dos ficheros de entrada: el fichero de datos, que contiene el conjunto de datos en formato KEEL, y el fichero de reglas, que contiene las reglas generadas por el algoritmo DRDS. A partir de estos ficheros, el sistema extrae la información necesaria, evalúa las reglas internamente y genera las representaciones gráficas y tabulares correspondientes. Finalmente, el usuario puede exportar los resultados en formato PDF o LaTeX.

La aplicación ofrece las siguientes funcionalidades principales: visualización analítica general del modelo y desglose detallado por regla; identificación de la cobertura mutua entre reglas y datos; y soporte para algoritmos tanto crisp como difusos (calculando en estos últimos el grado de pertenencia). Para la evaluación interna, el sistema genera la tabla de contingencia y deriva cuatro medidas de calidad [7]: **TPr** y **FPr** (ratios de verdaderos y falsos positivos), **Confianza** (precisión de la regla sobre los ejemplos cubiertos) y **WRACCN** (equilibrio entre generalidad y precisión).

3 Ejemplo ilustrativo

Para ilustrar las capacidades de SDRDPy, se analizaron los resultados del algoritmo NMEEFSD [3] sobre el conjunto de datos Iris. El análisis en la aplicación

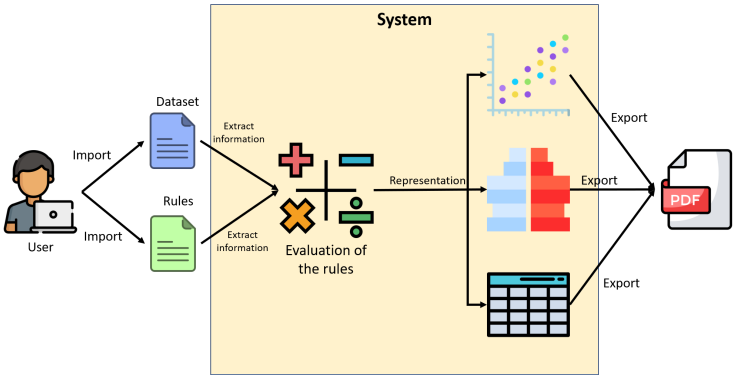


Fig. 1: Arquitectura general y flujo de trabajo de SDRDPy, desde la importación de ficheros por el usuario hasta la generación de informes.

se estructura en dos niveles interactivos (Fig. 2): una vista general de todas las reglas y un análisis detallado por regla individual.

El primer nivel, la vista general, se muestra tras importar los ficheros de datos y reglas, mostrando una interfaz organizada en tres zonas (Figura 2a): (1) el listado completo de reglas generadas, que permite interactuar con cada una de ellas; (2) un panel gráfico compuesto por un diagrama de dispersión y uno de pirámide para analizar la distribución de TPr y FPr, facilitando la identificación visual de los patrones de mayor calidad; y (3) una tabla que mapea la relación exacta entre reglas y datos mediante códigos de color (azul para cobertura correcta, naranja para incorrecta). Al posicionar el cursor sobre cualquier ejemplo, se despliegan sus atributos completos.

Por su parte, el análisis detallado se muestra al seleccionar una regla específica en el listado, ofreciendo un desglose pormenorizado en otras tres áreas clave (Figura 2b): (1) las tablas de contingencia y de medidas de calidad específicas; (2) un diagrama de puntos individual en el espacio TPr/FPr, donde se sombrea en rojo la zona de baja calidad ($TPr < FPr$); y (3) una tabla que aísla los datos cubiertos por dicha regla, manteniendo la codificación por colores.

4 Conclusiones

En este trabajo se ha presentado SDRDPy, una aplicación de escritorio para la visualización gráfica y tabular de los resultados obtenidos por algoritmos de DRDS. Su principal característica diferenciadora es la capacidad de procesar directamente el fichero de salida del algoritmo, sin necesidad de incorporar su implementación ni adaptar su formato, lo que la hace compatible con cualquier algoritmo DRDS independientemente del lenguaje o entorno en que haya sido desarrollado.

La aplicación cubre una laguna real en las herramientas disponibles, permitiendo que expertos de dominios como la medicina o el marketing interpreten los resultados de forma autónoma, sin requerir la asistencia de un especialista en Ciencia de Datos. SDRDPy se encuentra en continuo desarrollo, con planes de



(a) Vista general (Zonas 1-3) (b) Detalle de regla (Zonas 1-3)

Fig. 2: Interfaz de SDRDPy: (a) visualización general del modelo y (b) análisis detallado por regla.

incorporar nuevos algoritmos y realizar estudios periódicos de usabilidad. La descripción completa puede consultarse en el trabajo publicado en SoftwareX [8]. El código fuente está disponible en GitHub: https://github.com/SIMIDAT/SDRDPy y en Zenodo: https://zenodo.org/doi/10.5281/zenodo.12811990.

Acknowledgments. Este estudio está financiado por el programa FPU (FPU23/02620) y el proyecto TAIH (PID2023-149511OB-I00), financiado por el Ministerio de Ciencia, Innovación y Universidades (MICIU) y la Agencia Estatal de Investigación (AEI) (10.13039/501100011033), con cofinanciación del Fondo Europeo de Desarrollo Regional (FEDER) de la Unión Europea.

References

1. Bay, S.D., Pazzani, M.J.: Detecting group differences: Mining contrast sets. Data mining and knowledge discovery **5**(3), 213–246 (2001)
2. Carmona, C.J., del Jesus, M.J., Herrera, F.: A unifying analysis for the supervised descriptive rule discovery via the weighted relative accuracy. Knowledge-Based Systems **139**, 89–100 (2018)
3. Carmona, C.J., González, P., Del Jesus, M.J., Herrera, F.: Nmeef-sd: Non-dominated multiobjective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. IEEE Transactions on Fuzzy Systems **18**(5), 958–970 (2010)
4. Chen, M.C., Chiu, A.L., Chang, H.H.: Mining changes in customer behavior in retail marketing. Expert Systems with Applications **28**, 773–781 (05 2005). https://doi.org/10.1016/j.eswa.2004.12.033
5. Dhar, V.: Data science and prediction. Communications of the ACM **56**(12), 64–73 (2013)
6. García-Vico, A., Carmona, C., Martín, D., García-Borroto, M., del Jesus, M.: An overview of emerging pattern mining in supervised descriptive rule discovery: Taxonomy, empirical study, trends, and prospects., 2017, 8, p. e1231
7. Herrera, F., Carmona, C.J., González, P., Del Jesus, M.J.: An overview on subgroup discovery: foundations and applications. Knowledge and information systems **29**(3), 495–525 (2011)
8. Padilla-Rascón, M.A., González, P., Carmona, C.J.: Sdrdpy: An application to graphically visualize the knowledge obtained with supervised descriptive rule algorithms. SoftwareX **28**, 101939 (2024). https://doi.org/10.1016/j.softx.2024.101939