



LABORATORIO DE INVESTIGACIONES EN GESTIÓN DE PROYECTOS

SISTEMA CLASIFICADOR BORROSO BASADO EN ALGORITMOS GENÉTICOS PARA EVALUAR EL ESTADO DE EJECUCIÓN DE PROYECTOS

Carlos Rafael Rodríguez Rodríguez

Marieta Peña Abreu

Gilberto Fernando Castro Aguilar

Pedro Yobanis Piñero Pérez

INTRODUCCIÓN

Guías, estándares y metodologías ofrecen buenas prácticas para la gestión de proyectos. Estos proponen procesos para el control, seguimiento y evaluación de proyectos.

El control de la ejecución de proyectos se caracteriza por:

- la existencia de imprecisión, vaguedad e incertidumbre
- ser basada en el conocimiento de experto
- condiciones cambiantes en el entorno de ejecución

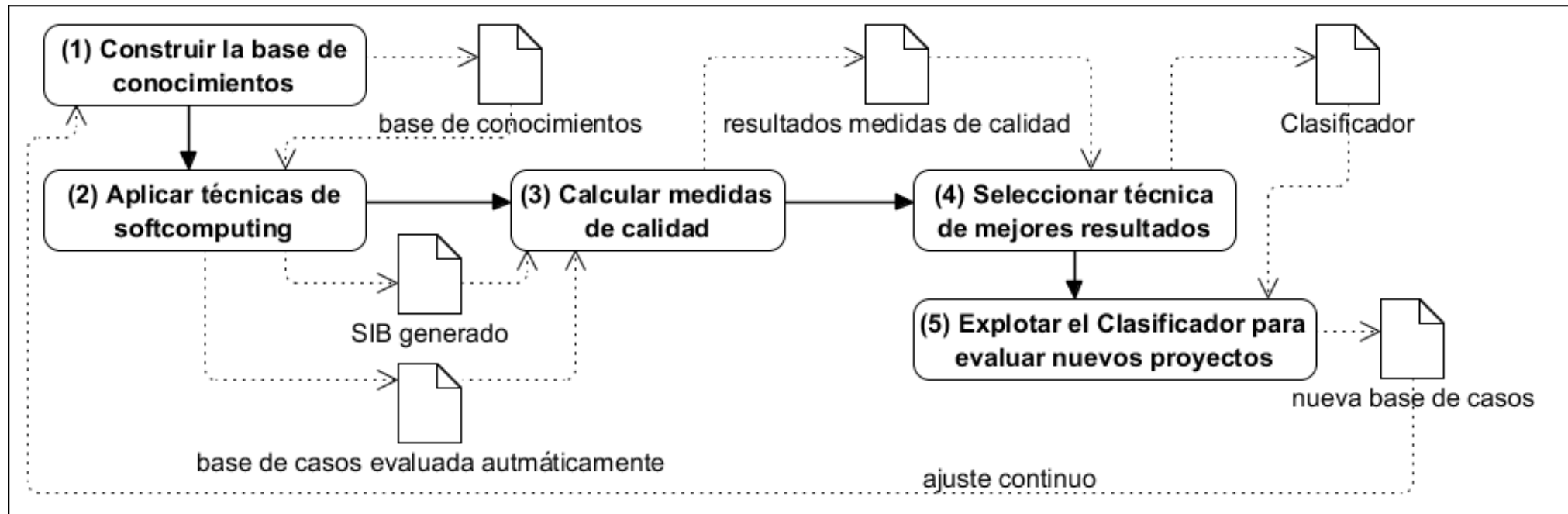
Usar un sistema borroso estático es insuficiente debido a su incapacidad para adaptarse con regularidad a los cambios

En (Bermúdez, 2015)^[1] se aborda esta problemática usando técnicas de regresión. No obstante dado el teorema **No free lunch**^[2], es necesario experimentar con otras técnicas como las de clasificación.

[1] Bermúdez, Anie. 2015. Sistema basado en técnicas de soft computing para la evaluación de la ejecución de proyectos. Universidad de las Ciencias Informáticas. La Habana : s.n., 2015. pág. 80, Tesis de Maestría.

[2] Wolpert, D. 1996. The Lack of A Priori Distinctions Between Learning Algorithms. s.l. : Massachusetts Institute of Technology, 1996. Vol. 8, 7, pp. 1341-1390. ISSN: 0899-7667.

MÉTODO



- Adopta como base la biblioteca AnalysisPro^[3].
- Utiliza como dependencias el paquete FRBS^[4] y funciones nativas de R.
- Se aplican 3 técnicas de *soft computing* basadas en AG: *gfs.gccl*^[5], *fh.gbml*^[6] y *slave*^[7].

[3] LIGP. 2015. *Biblioteca para el Análisis de Datos AnalysisPro*. Laboratorio de Investigaciones en Gestión de Proyectos, Universidad de las Ciencias Informáticas. La Habana : s.n., 2015.

[4] Riza, L.S., Bergmeir, Ch., Herrera, F. and Benítez, J.M. 2015. frbs: Fuzzy Rule-based Systems for Classification and Regression Tasks. *Journal of Statistical Software*. 2015. Vol. 65, 1, pp. 1-30.

[5] Ishibuchi, H., Yamamoto, T. and Nakashima, T. 2005. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics*. 2005. Vol. 35, 2, pp. 359-365.

[6] Ishibuchi, H., Nakashima, T. and Murata, T. 1999. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics*. 1999. Vol. 29, 5, pp. 601 - 618.

[7] González, A. and Pérez, R. 2001. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*. 2001. Vol. 31, 3, pp. 417-425.

Técnicas de *soft computing* aplicadas en la actividad 2

Sistema borroso basado en aprendizaje cooperativo - competitivo genético (*gfs.gccl*) [5]

Paso 1: Generar una población inicial de N_{pop} reglas difusas especificando al azar los conjuntos borrosos de los antecedentes de cada regla. La clase consecuente y el grado de certeza se determinan por un procedimiento heurístico.

Paso 2: Clasificar todos los casos de entrenamiento, con las reglas difusas de la población actual; luego calcular la aptitud de cada regla.

Paso 3: Generar N_{rep} nuevas reglas difusas aplicando operadores genéticos a la población actual. La clase consecuente y el grado de certeza de cada nueva regla se determinan utilizando el mismo procedimiento heurístico.

Paso 4: Reemplazar las N_{rep} reglas difusas que menor valor de aptitud tienen en la población actual con las nuevas reglas generadas.

Paso 5: Terminar si se cumple la condición de parada preestablecida, si no regresar al paso 2.

Sistema híbrido genético de los enfoques Michigan y Pittsburgh (*fh.gbml*) [6]

Paso 1: Generar una población inicial de n individuos, donde cada individuo es un conjunto de m reglas.

Paso 2: Calcular la actitud de cada individuo de la población actual.

Paso 3: Generar $n-1$ nuevos individuos mediante operadores genéticos, como en el enfoque Pittsburgh. Aplicar una iteración simple del enfoque Michigan a cada nuevo individuo con una probabilidad predeterminada.

Paso 4: Añadir el mejor individuo de la población actual a los $n-1$ nuevos individuos generados para formar la próxima población.

Paso 5: Retornar al paso 2 si la condición de parada previamente definida no se cumple.

Algoritmo de aprendizaje estructural en ambientes de vaguedad (*slave*) [7]

Paso 1: Construir una regla difusa mediante el procedimiento de un algoritmo genético.

Paso 2: Adicionar la regla al conjunto final de reglas

Paso 3: Chequear y penalizar la regla

Paso 4: Si se cumple la condición de parada, devolver el conjunto de reglas como solución, si no regresar al paso 1.

Medidas de calidad utilizadas en la actividad 3

- Cantidad de reglas generadas (CRG)
- Porcentaje de clasificaciones correctas (% CC) $\% CC = \frac{100 * \text{cant. casos bien clasif.}}{\text{total de casos}}$
- Falsos positivos (FP): Cantidad de casos clasificados de “B” cuando su estado es “M”
- Falsos negativos (FN): Cantidad de casos clasificados de “M” cuando su estado es “B”
- Error cuadrático medio (MSE) $MSE = \frac{1}{n} \sum_{i=1}^n (y_{real} - y_{pred})^2$
- Raíz del error cuadrático medio (RMSE) $RMSE = \sqrt{MSE}$
- Error Porcentual de la Media Absoluta Simétrica (SMAPE) $100 * \frac{1}{n} \sum_{i=1}^n \frac{|y_{real} - y_{pred}|}{(y_{real} + y_{pred})/2}$

Test estadísticos utilizados en la actividad 4

- Test Shapiro-Wilk: para comprobar el supuesto de normalidad de los resultados de las técnicas para cada variable.
- Test de Friedman: para comparar los resultados de todas las técnicas para cada variable.
- Test de Wilcoxon: para comparar por pares los resultados de las técnicas para cada variable.
- Con los resultados del Test de Wilcoxon se forman grupos de técnicas sin diferencias significativas entre ellas. Tal que a menor grupo mejor resultado.
- En los Test de Friedman y Wilcoxon se considera 0,05 como umbral para determinar la diferencia significativa.
- Para ambos Test se utiliza el método de Monte Carlo con 99 % en el nivel de confianza.

ANÁLISIS DE RESULTADOS

Estructura de la base de conocimientos - Población

Proyecto	-IRE-	-IRP-	-ICD-	-IRRH-	-IREF-	Evaluación
P 13	1.025	1.038	0.999	80.35	0.605	B
P 14	0.857	0.972	0.975	50.775	0.569	R
P 31	0.923	0.971	0.975	26.2	0.363	M
P 140	0	1	1	51.025	0.425	R
P 204	0.476	0.461	0.975	10	0.2	M

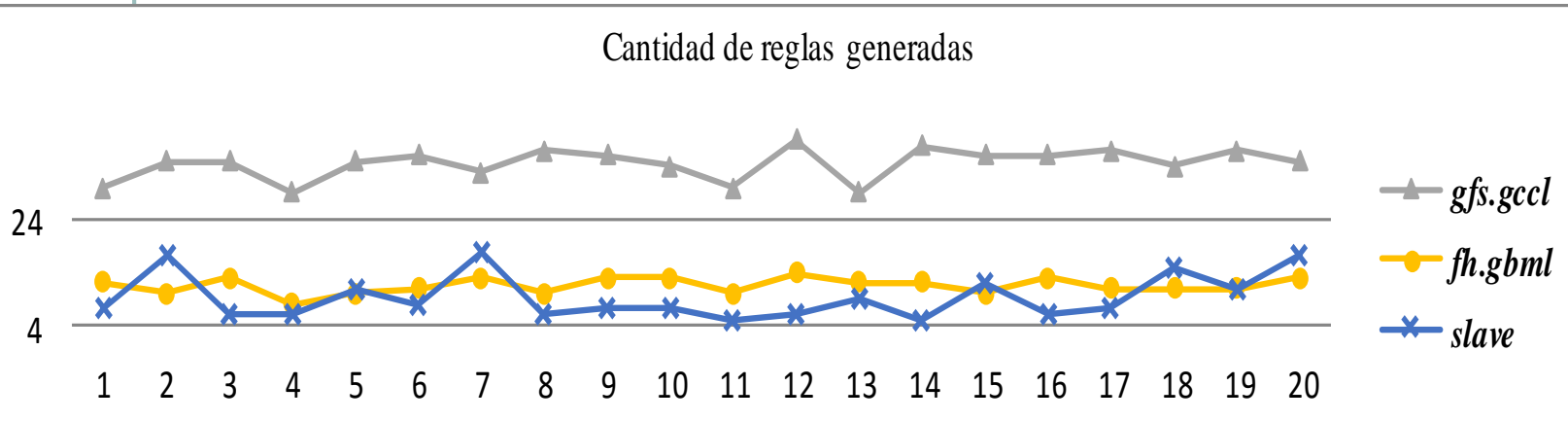
Parámetros configurados a cada técnica en el experimento

Técnica	Parámetros (nombre = valor)
gfs.gccl	<i>num_class = 3; num_labels = 3; popu_size = 20; persen_cross = 0,9; persen_mutant = 0,3; max_gen = 50</i>
fh.gbml	<i>num_class = 3; popu_size = 50; max_num_rule = 25; persen_cross = 0,9; persen_mutant = 0,3; max_gen = 100; p_dcare = 0,5; p_gccl = 1</i>
slave	<i>num_class = 3; num_labels = 3; max_iter = 20; k_lower = 0,25; k_upper = 0,75; persen_cross = 0,9; persen_mutant = 0,3; max_gen = 50; epsilon = 0,1</i>

Resultados del experimento

Variable: Cantidad de reglas generadas (CR)

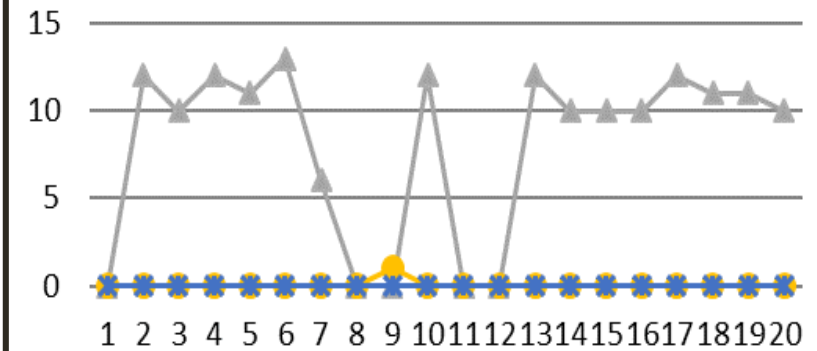
Cantidad de reglas generadas



p-valor = 1,361e-07

Variable: Cantidad de FP

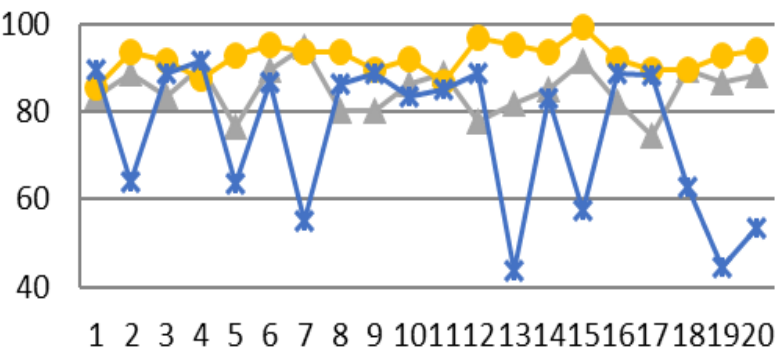
(a) Conjunto de entrenamiento



p-valor = 1,874e-06

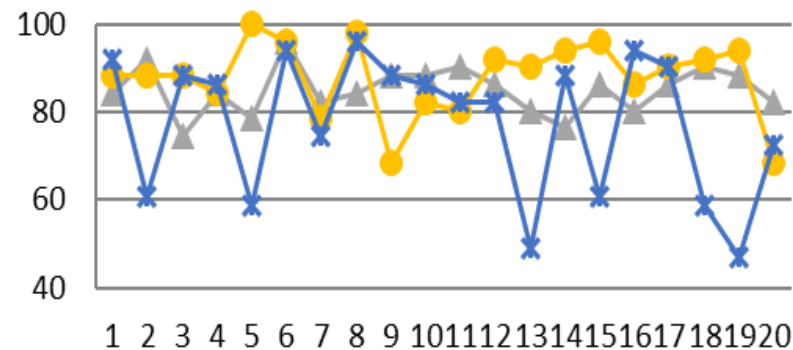
Variable: Porcentaje de clasificaciones correctas (% CC)

(a) Conjunto de entrenamiento



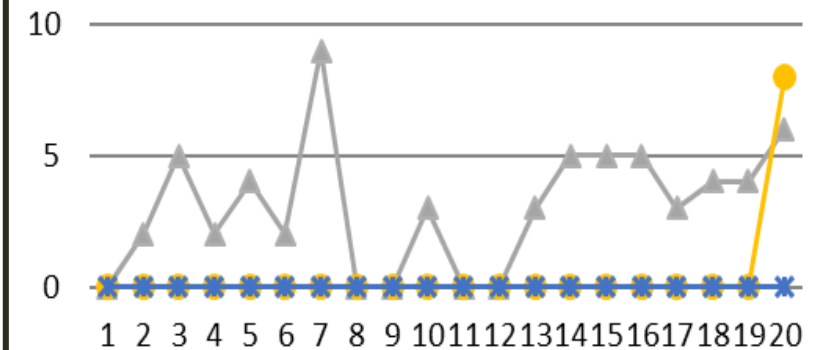
p-valor = 0,0002324

(b) Conjunto de prueba



No hay diferencias significativas

(b) Conjunto de prueba

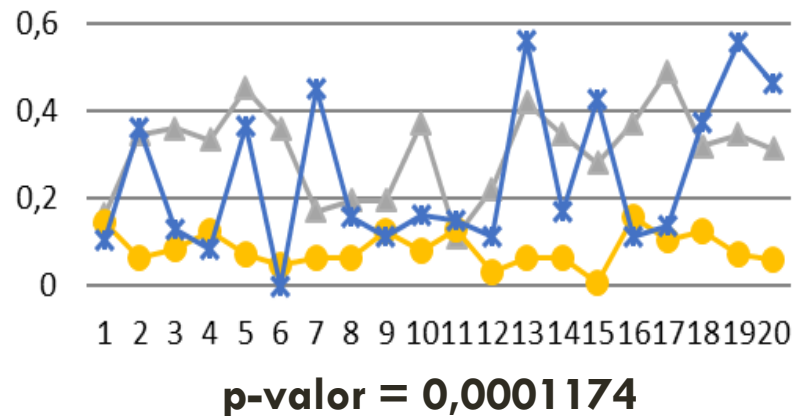


p-valor = 2,575e-06

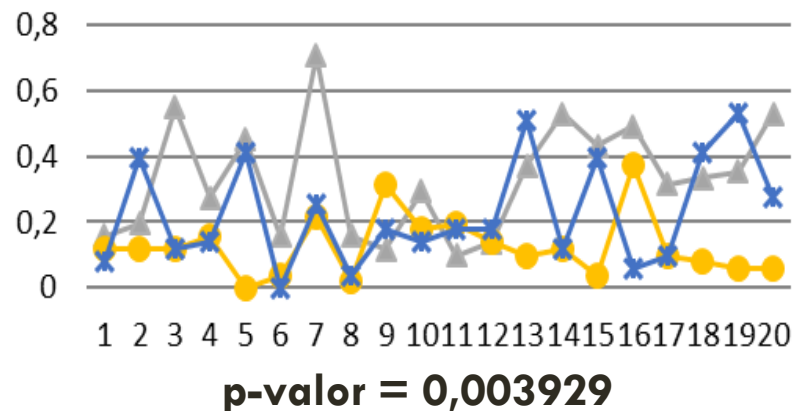
Resultados del experimento

Variable: (MSE)

(a) Conjunto de entrenamiento

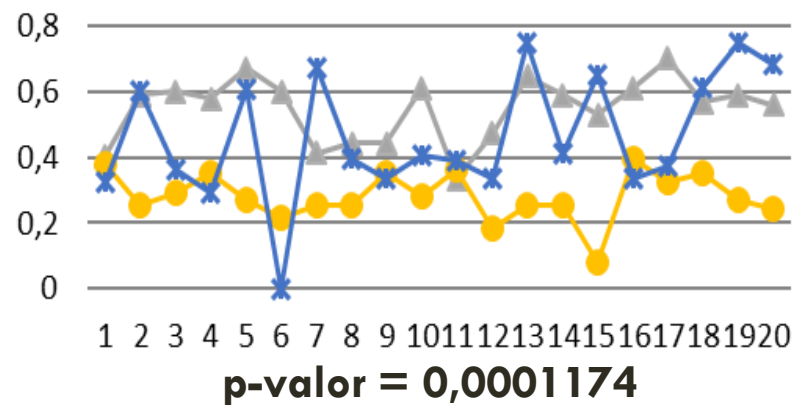


(b) Conjunto de prueba

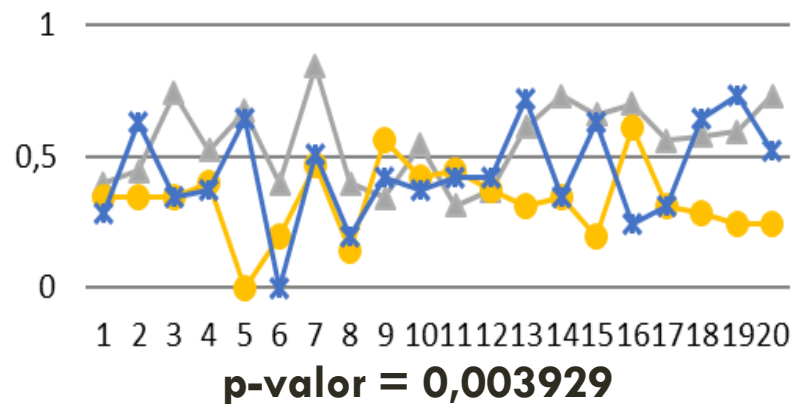


Variable: (RMSE)

(a) Conjunto de entrenamiento

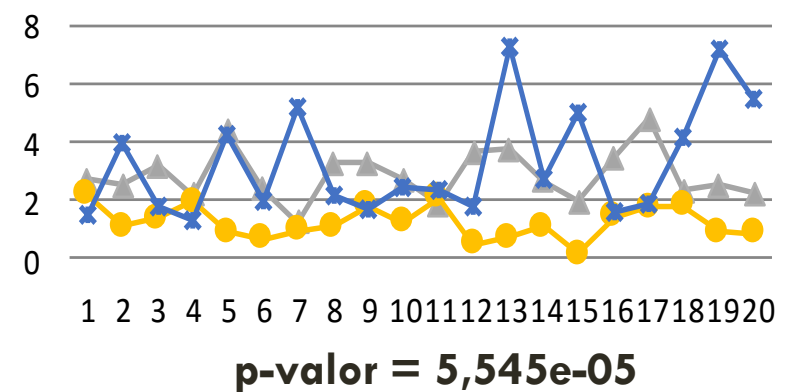


(b) Conjunto de prueba

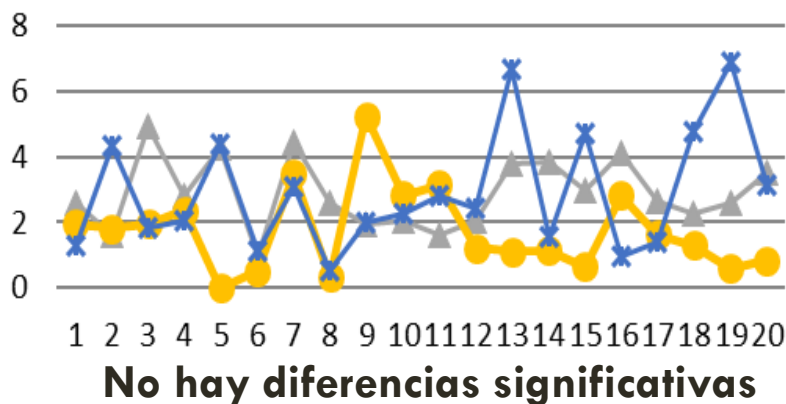


Variable: (SMAPE)

(a) Conjunto de entrenamiento



(b) Conjunto de prueba



Resumen de resultados del experimento

- En las variables % CC-test, FN-train-test y SMAPE-test: no se encontraron diferencias significativas.
- *fh.gbml* se ubica en el grupo 1 en todos los experimentos.
- *slave* se ubica en el grupo 2 solo en 4 experimentos.

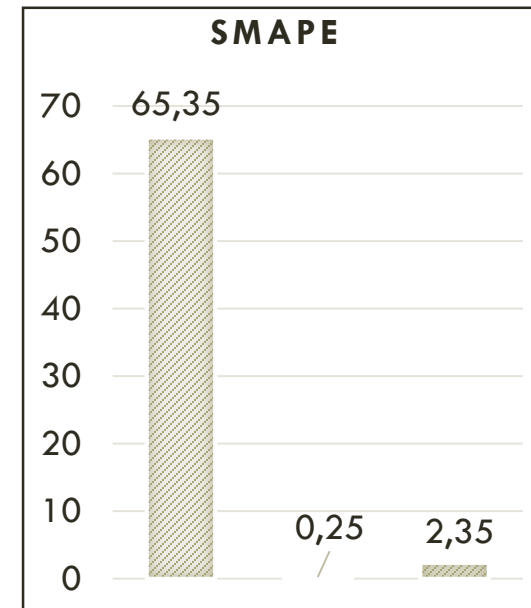
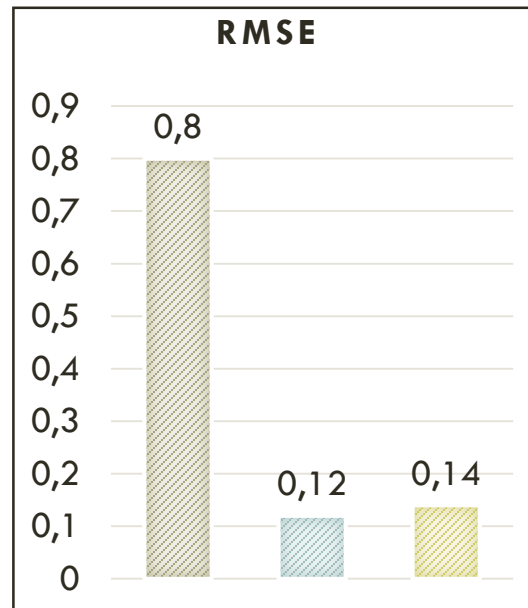
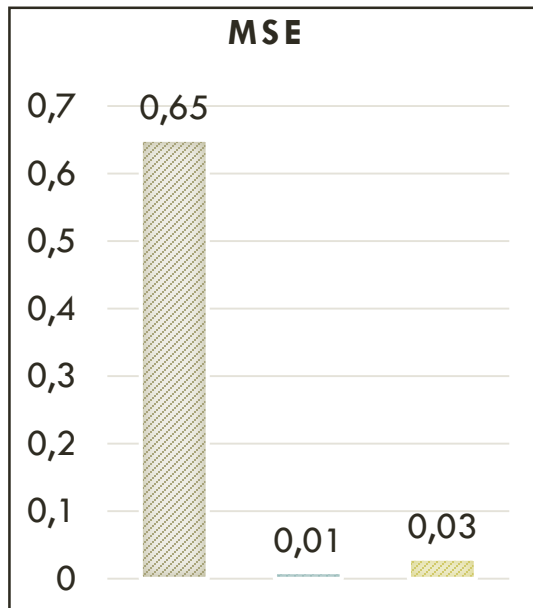
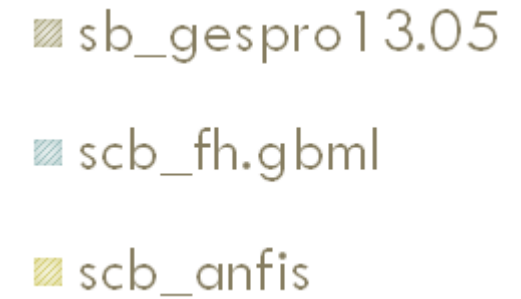
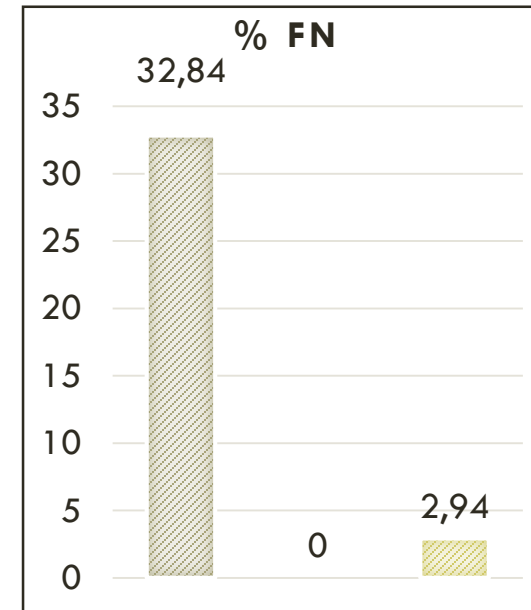
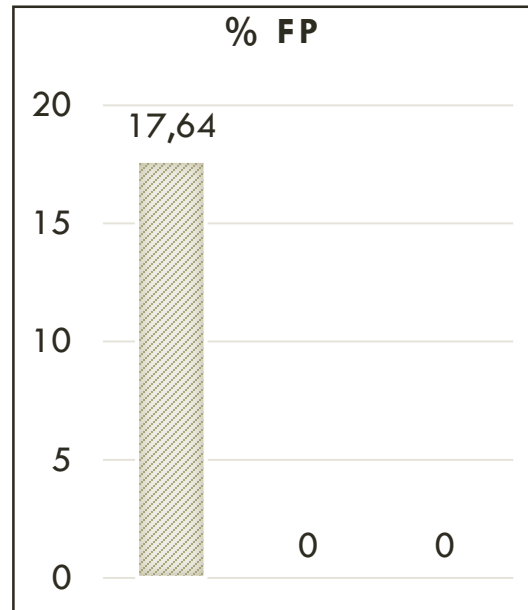
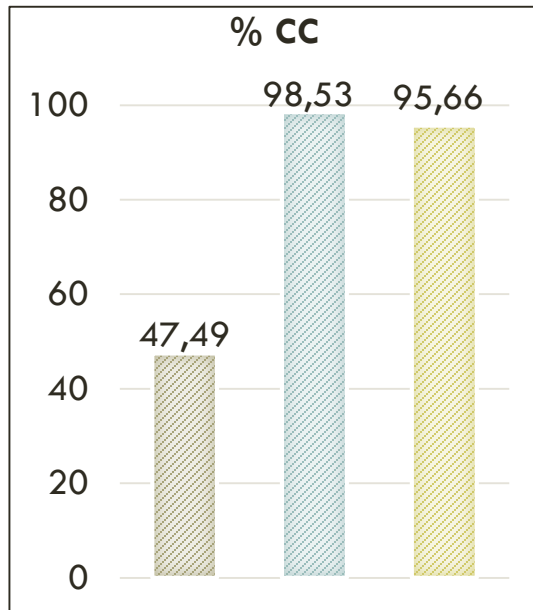
El sistema clasificador borroso que se obtuvo

- 13 reglas
- funciones de pertenencia triangulares
- t-norma = product, s-norma = max
- función de implicación = Zadeh
- 14 etiquetas lingüísticas para las 5 variables predictoras

Ejemplo de regla:

*if ire is v.1_a.12 and
irpa is dont_care and
icd is v.3_a.1 and
irrh is v.4_a.10 and
iref is v.5_a.1
then eval is 1*

Comparación con *sb_gespro13.05* y *scb_anfis*



CONCLUSIONES

- La mayoría de los aportes anteriores sobre el uso de la *soft computing* en la toma de decisiones en la Gestión de Proyectos utilizan sistemas borrosos estáticos especificados por expertos y no se integran con herramientas informáticas, lo que dificulta su uso efectivo y generalizado.
- El método propuesto facilita el tratamiento de la incertidumbre contenida en la información y la adaptación del SCB a los cambios en los estilos de gestión de la organización, en un ambiente de mejora continua.
- Las técnicas utilizadas aprenden reglas automáticamente, no especificadas por expertos. Su integración a la biblioteca *AnalysisPro* constituye una oportunidad de mejora para las herramientas de Gestión de Proyectos.
- La técnica *fh.gbml* fue la única que se ubicó en el grupo 1 para todas las métricas analizadas durante la experimentación y se considera, por tanto, la de mejor desempeño.
- El SCB que se obtuvo mejora los resultados de la evaluación del estado de ejecución de proyectos en comparación con el SB que utiliza Xedro-GESPRO 13.05.