Universidad de las Ciencias Informáticas

Facultad 6

Título: Evaluación de algoritmos basados en Lógica Difusa para la predicción de actividad biológica

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas

Autor(es): Isbel Delgado García

Yuniel Márquez Bocalandro

Tutor(es): Dr. Ramón Carrasco Velar

Ing. Julio Omar Prieto Entenza

Ciudad de la Habana, Cuba.

Junio ,2009

"Año del 50 aniversario del triunfo de la Revolución."

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Yuniel Márquez Bocalandro

_____

Firma del autor

Isbel Delgado García

_____

Firma del autor

Ing. Omar Prieto Entenza

_____

Firma del tutor

Dr. Ramón Carrasco Velar

_____

Firma del tutor

DATOS DE CONTACTO

Tutores:

Dr. Ramón Carrasco Velar

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

rcarrasco@uci.cu

Ing. Julio Omar Prieto Entenza

Universidad de las Ciencias Informáticas, Ciudad de La Habana, Cuba.

jprieto@uci.cu

AGRADECIMIENTOS

Agradecemos a nuestros padres por su apoyo, confianza y ejemplo en todo momento para que hoy seamos esas personas de las cuales esperan lo mejor. A nuestros tutores Julio y Carrasco, por su paciencia, preocupación y por ser quienes dedicaron parte de su tiempo en enseñarnos que la sabiduría no tiene fronteras. A nuestros familiares que confiaron plenamente en nosotros, siempre con la esperanza que venceríamos los obstáculos para lograr nuestros sueños. A nuestras amistades por preocuparse por nosotros y compartir todos estos días juntos. A todos desde lo más profundo de nuestros corazones,

Muchas Gracias

DEDICATORIA

- A mis abuelos por siempre demostrarme su amor incondicional. Los quiero con la vida.
- A mis padres por ser lo más preciado que tengo en este mundo, por ser mis fieles amigos y las personas que más amo en la vida. Siempre estaré muy orgulloso de ustedes.
- A Liannita por quererme tanto y ser la mujer más especial de mi mundo. Te quiero más que a nada.
- A mi hermanita que mucho ha ayudado a que todo saliera como hasta hoy.

Yuniel Márquez Bocalandro

- A mis padres por guiarme en la vida y demostrarme día a día su infinito amor.
- A mis abuelos, en quienes vi un ejemplo para mi persona.
- A mi familia por darme en todo momento su confianza y su apoyo.
- A mis amistades que con su confianza contribuyeron a que este momento fuera posible.
- En fin a todos los que hicieron posible este trabajo.

Isbel Delgado García

RESUMEN

El presente trabajo surge por la necesidad de incorporar al proyecto **"alasGRATO"** modelos difusos de cuantificación para predecir la relación estructura-actividad en compuestos orgánicos. Los algoritmos MOGUL-TSK, MOGUL-IRLHC, COR-GA y FRSBM basados en regresión, fueron evaluados en bases de datos estructurales de Cefalosporinas frente a Staphylococcus Aureus y Escherichia coli. Dichos algoritmos combinan la lógica difusa con algoritmos evolutivos. Se estableció una comparación entre los algoritmos a través de pruebas estadísticas las cuales arrojaron que a pesar de no existir diferencias globales entre ellos, el FRSBM posee una mayor capacidad de predicción obteniéndose resultados con un 8% de error, lo cual se considera un valor aceptable debido a la complejidad estructural de la muestra. Se propone la implementación de dicho algoritmo a la plataforma **"alasGRATO"**.

.

PALABRAS CLAVES

Predicción, lógica difusa, algoritmo genético, actividad biológica, modelos difusos.

"La imaginación es más importante que el conocimiento. El conocimiento es limitado, la imaginación no"

Albert Einstein

Introducción

El término Bioinformática, hace referencia a campos de estudios interdisciplinarios, dígase la Informática, Inteligencia Artificial, Matemática Aplicada, Estadística, Química y Bioquímica que se han vinculado, para solucionar problemas, analizar datos, y simular sistemas o mecanismos, todos ellos de índole biológica. Su núcleo principal se encuentra en la utilización de recursos computacionales para investigar y solucionar problemas a gran escala.

En la actualidad existe un estrecho vínculo entre la Bioinformática y la industria farmacéutica, que se ha encaminado en la búsqueda de nuevas soluciones al procesamiento de un gran número de compuestos químicos, su almacenamiento y posterior acceso a los datos, de una forma más rápida y estructurada. Para ello se utilizan diferentes técnicas de procesamiento, de gran utilidad para la predicción de actividades biológicas y el diseño racional de fármacos asistidos por ordenadores.

El proceso de obtención manual de nuevos fármacos es altamente costoso que requiere investigaciones de gran complejidad. Esta dificultad viene dada, entre otras cosas, por el gran número de compuestos conocidos, de los cuales una gran parte no poseen aplicaciones farmacológicas. En busca de una solución más rápida y menos costosa se han desarrollado en el mundo diversas aplicaciones para el manejo de datos de forma computarizada.

“alasGRATO” surge como una plataforma inteligente para la predicción de actividad biológica de compuestos orgánicos, la cual necesita incorporar modelos difusos de cuantificación para predecir la relación estructura-actividad en compuestos orgánicos, lo que constituye el **problema científico** a resolver.

De lo anterior, se define como **objeto de estudio** la Lógica Difusa aplicada a la predicción de actividad biológica. El **campo de acción** se enmarca en la Lógica Difusa aplicada a la predicción de actividad biológica en compuestos orgánicos empleando regresión. Para dar solución al problema planteado se trazó como **objetivo general**: proponer algoritmos que a través de regresión en lógica difusa permita predecir cuantitativamente la actividad biológica de compuestos orgánicos.

Objetivos Específicos

1. Identificar algoritmos difusos que utilicen regresión para predecir actividad biológica en compuestos orgánicos.

2. Probar los algoritmos seleccionados.

3. Comparar los resultados obtenidos.

Tareas de Investigación

- ✓ Análisis del estado del arte respecto al uso de los algoritmos difusos que utilicen regresión en problemas de predicción de actividad biológica.
- ✓ Evaluación de algoritmos difusos en problemas de predicción de actividad biológica.
- ✓ Comparación de resultados en base de datos estructurales.

Capítulo 1: Fundamentación Teórica

En este capítulo se realiza una breve explicación de los métodos empleados en la creación de nuevos fármacos asistidos por computadora y las técnicas de predicción existentes. Además se muestran una serie de soluciones informáticas vinculadas a la predicción de la relación estructura-actividad.

1.1 Introducción a la Modelación Molecular

El modelado molecular es un término general que engloba métodos teóricos y técnicas computacionales para imitar el comportamiento de moléculas. Dichas técnicas son utilizadas en los campos de la Química, Biología y Ciencia de materiales para el estudio de sistemas moleculares que abarcan desde pequeños sistemas químicos hasta grandes moléculas biológicas y disposiciones materiales [(1)].

Para poder establecer estos modelos moleculares es necesario describir la estructura química de manera cuantitativa. Estos valores son los llamados descriptores, que pueden ser tanto de naturaleza químico-cuánticos, químico-físicos, topológicos, topográficos o híbridos.

Para construir modelos de predicción se han empleado diferentes técnicas de inteligencia artificial como son los árboles de regresión, las máquinas de aprendizaje, las redes neuronales, los algoritmos genéticos y la lógica difusa. En situaciones reales con frecuencia aparecen elementos tales como el ruido, la ambigüedad, la incertidumbre, la incompletitud, la inconsistencia y la imprecisión; elementos negativos que generalmente son evitados por los modelos clásicos de aprendizaje, clasificación y regresión que no pueden ser ignorados. Una posible solución al tratamiento de estos elementos negativos de forma simultánea aparece con la aplicación de técnicas de softcomputing.

1.2 Softcomputing

El término *softcomputing* lo introduce Zadeh para denotar una aproximación al razonamiento humano que deliberadamente hace uso de la tolerancia humana a imprecisiones y vaguedades para obtener soluciones razonables que son fáciles de manipular. Bajo este principio, los sistemas borrosos o difusos, las redes neuronales, la computación evolutiva, el razonamiento probabilístico y las combinaciones de dichas técnicas son consideradas como *softcomputing* [(2)]. En general existe un grupo de propiedades que caracterizan a las técnicas y sistemas basados en softcomputing y que pueden ayudarnos a identificar los límites de este campo [(3) (4)] tales como:

- Es lo opuesto al hardcomputing, no es prescriptivo en la solución de un problema, en este campo no hay soluciones programadas para cada posible situación.
- Sus técnicas son robustas ante entornos con entradas ruidosas y tienen una alta tolerancia a la imprecisión de los datos con los que opera.
- Es una necesidad cuando la información de que disponemos es imprecisa; y segundo cuando existe una tolerancia a la imprecisión que podría ser explotada para ganar robustez, en soluciones de bajo costo y en mayor capacidad de modelación [(5)].

1.3 Lógica Difusa

La lógica difusa *(*fuzzy logic) permite tratar información imprecisa, como estatura media, temperatura baja o mucha fuerza, en términos de conjuntos borrosos o difusos. Estos conjuntos borrosos se combinan en reglas para definir acciones, como por ejemplo, *Si* la temperatura es alta *entonces* enfría mucho. De esta manera, los sistemas de control basados en lógica difusa combinan variables de entrada (definidas en términos de conjuntos borrosos [(6)]), por medio de grupos de reglas que producen uno o varios valores de salida [(7)].

Esta técnica se utiliza en la resolución de una gran variedad de problemas, principalmente los relacionados con control de procesos complejos y sistemas de decisión encontrándose extendidos en la tecnología cotidiana. Para la obtención de dichos modelos de forma automatizada investigadores como Mamdani [(8)] aplicaron la lógica difusa a diversos procesos, desarrollando los llamados Sistemas Basados en Reglas Difusas [(9)], los cuales han logrado resultados exitosos ajustando sus modelos muy cerca de los reales.

1.4 Sistemas Basados en Reglas Difusas

Los sistemas basados en reglas difusas (FRBS, de su nombre en inglés Fuzzy Rule-Based Systems) son una extensión de los sistemas clásicos de representación del conocimiento basados en reglas [10]. Al igual que estos últimos, los FRBS se componen de reglas condicionales de la forma "IF antecedente THEN consecuente", con la particularidad de que el antecedente y el consecuente son variables de la lógica difusa y no de la lógica clásica.

Los sistemas de reglas aplicados a problemas del mundo real componen su base de conocimiento mediante dos procesos: la extracción directa de información de repositorios de datos y el aprendizaje de un experto en el dominio de la aplicación. En este último, el experto vuelca el conocimiento que ha adquirido a través de su experiencia para definir, en la medida de lo posible, las variables que conforman el entorno del sistema, las variables internas del sistema y las reglas que definen la lógica del sistema. Los FRBS facilitan este proceso ya que permiten al experto expresar su conocimiento con variables lingüísticas que son propias del lenguaje humano.

1.4.1 Conceptos Fundamentales

A continuación se describen brevemente los principales conceptos que serán utilizados durante el presente trabajo [7]:

- **Universo de discurso** Conjunto de valores numéricos que puede tomar para una variable discreta, o el rango de valores posibles para una variable continua. En el caso de la variable lingüística "altura", por ejemplo, podrían ser el conjunto de valores comprendido entre 1.5 y 2.3 m.

- **Conjunto o subconjunto difuso** Se encuentra asociado a un valor lingüístico, definido por una palabra, adjetivo o etiqueta lingüística A. Al mismo se le añade una función de pertenencia o inclusión.

- **Función de pertenencia o inclusión** La función de pertenencia o inclusión $\mu_A(x)$, *de*finida como un número entre 0 y 1 (incluyéndolos a los dos), indica el grado en que la variable x está incluida en el concepto representado por la etiqueta *A*.

- **Operador difuso:** Término designado para realizar operaciones difusas sobre los componentes de un FRBS. Existen tres tipos de operadores difusos: implicación, conjunción y agregación.
- **Variable lingüística:** Cada una de las variables de entrada o salida de un FRBS, que toma valores lingüísticos dentro del conjunto de etiquetas lingüísticas definidas en el mismo.
- **Reglas Difusas** Las reglas borrosas combinan uno o más conjuntos borrosos de entrada, llamados antecedentes o premisas, y les asocian un conjunto borroso de salida, llamado consecuente o consecuencia.
- **Base de Reglas** Conjunto de reglas que expresan las relaciones conocidas entre antecedentes y consecuentes. Permiten expresar el conocimiento que se dispone sobre la relación entre antecedentes y consecuentes. Precisándose de varias reglas para expresar este conocimiento de forma completa.

1.4.2 Tipos de FRBS

Los FRBS se diferencian, atendiendo a su estructura interna, en dos tipos: Mamdani y Takagi-Sugeno-Kang.

1.4.2.1 FRBS de tipo Mamdani

E.H. Mamdani construyó el primer FRBS que, utilizando la formulación proporcionada por Zadeh, aplicó un sistema de lógica difusa a un problema de control. Esta aplicación de la lógica difusa se conoce como FLC, del inglés Fuzzy Logic Controller o controlador mediante lógica difusa. Su estructura básica, se muestra en la Figura 1.1

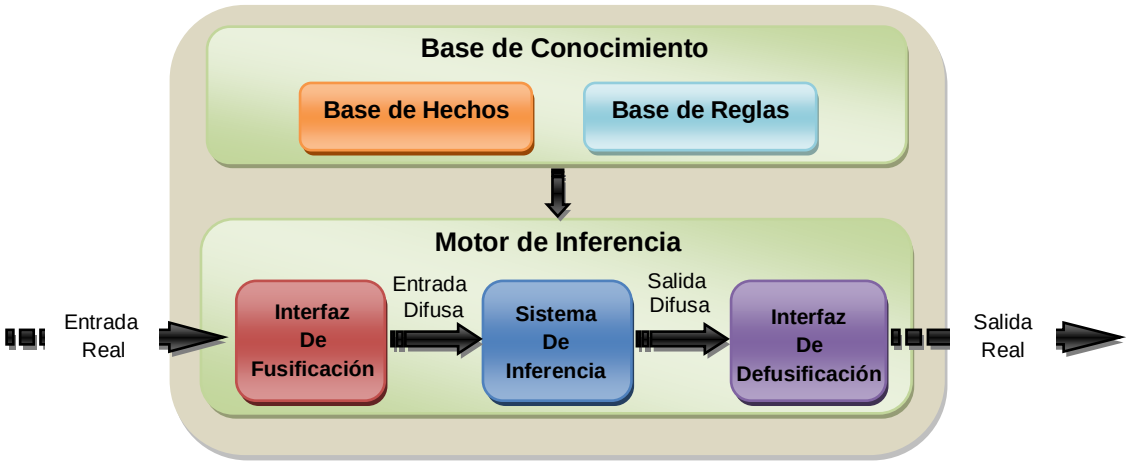

Figura 1.1 Estructura básica del FRBS tipo Mamdani o FCL.

Base de conocimiento:

Almacena todo el conocimiento disponible del problema a tratar en forma de hechos y reglas difusas, a través de los cuales opera el sistema de inferencia para generar una salida difusa para cada entrada difusa. Se divide en una base de hechos y en una base de reglas. La base de hechos contiene un conjunto de variables lingüísticas de entrada y salida del sistema, así como un conjunto etiquetas lingüísticas, cada una de las cuales define un conjunto difuso. Cada conjunto difuso consta de una función que indica el grado en que una variable lingüística pertenece a dicho conjunto, llamada función de pertenencia. La base de reglas es una colección de reglas condicionales construidas con las variables y etiquetas lingüísticas de la base de hechos, el operador lógico AND y las partículas condicionales IF y THEN.

La parte de una regla situada entre IF y THEN es el antecedente de la regla, siendo el resto el consecuente. Mostrándose de la siguiente forma:

$$R_i: IF\ X_1\ IS\ A_1\ AND\ ...\ AND\ X_n\ IS\ A_n\ THEN\ Y\ IS\ B_i$$

donde n es el número total de variables lingüísticas e i el número total de reglas difusas.

Motor de Inferencia:

El motor de inferencia consta de una interfaz de fusificación, un sistema de inferencia y una interfaz de defusificación. La interfaz de fusificación permite al FRBS traducir entradas reales a sus correspondientes valores en el universo de los conjuntos difusos, con los que opera el sistema de inferencia. El sistema de inferencia genera conjuntos difusos como salidas a partir de las entradas difusas obtenidas de la interfaz de fusificación, de acuerdo a los hechos y reglas almacenados en la base de conocimiento. La interfaz de defusificación agrega la información de dichos conjuntos difusos y los traduce en un valor no difuso correspondiente al dominio de salida del FRBS.

1.4.2.2 FRBS de tipo Takagi-Sugeno-Kang

Este tipo de FRBS, también conocido por sus siglas TSK FRBS, lleva su nombre en honor a los autores del artículo en el que fue presentado originalmente.

La diferencia principal con el modelo propuesto por Mamdani radica en que los consecuentes de las reglas difusas de las que hace uso se representan como una combinación lineal de los antecedentes de dichas reglas.

$$IF\ X_1\ IS\ A_1\ AND\ ...\ AND\ X_n\ IS\ A_n\ THEN\ Y = p_1 \cdot X_1 + ... + p_n \cdot X_n + p_0$$

donde $\vec{p} = (p_0, p_1, ..., p_n)$ es un vector de parámetros reales.

El resultado de un TSK FRBS es la media ponderada de las salidas de todas sus reglas difusas, que se calcula de la siguiente forma:

$$y_0 = \frac{\sum_{i=1}^{m} h_i \cdot y_i}{\sum_{i=1}^{m} h_i}$$

donde m es el número total de reglas, y_i es la salida de la regla $i\text{-}ésima$ y $h_i = T(A_{i1}(x_1), ..., A_{in}(x_n))$ la correspondencia entre la entrada al TSK FRBS x_0 y el antecedente de la regla $i\text{-}ésima$. T es un operador de conjunción usualmente representado por el mínimo.

Los TSK FRBS ofrecen mayor sencillez de diseño y cálculo que un FRBS de tipo Mamdani, aunque la comprensión lingüística de sus reglas difusas es menos intuitiva debido a la estructura de su consecuente, dificultando la interacción con un experto humano.

Aplicado al proceso de construcción de un modelo basado en reglas, un algoritmo genético (AG) es capaz de explorar el espacio de soluciones comprendido por la totalidad de las bases de conocimiento de un FRBS de un dominio determinado. Esta técnica codifica bases de conocimiento en forma de individuos y lo almacena dentro de la población de un AG, de tal forma que a través de los operadores de cruce, mutación y selección, se crean mejores bases de conocimiento generación tras generación hasta alcanzar una solución válida.

1.5 Algoritmos Genéticos

Los Algoritmos Genéticos son un tipo de algoritmos evolutivos usados para resolver problemas de búsqueda y optimización [(11)]. Se basan en la imitación del proceso evolutivo que se produce en la naturaleza para resolver problemas de adaptación al medio. De esta manera, son capaces de hacer evolucionar los individuos de una población hacia soluciones mundo real siempre que se encuentre una codificación adecuada del problema, siendo ésta una de las claves principales de éxito cuando se aplica este tipo de técnicas.

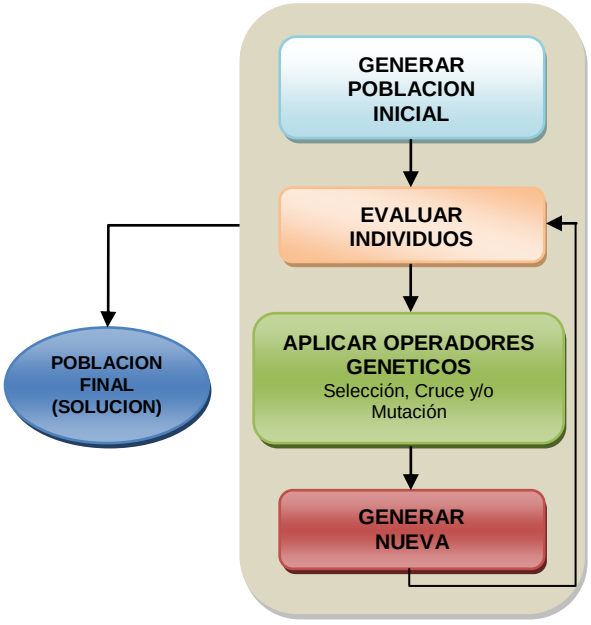

Figura 1.2 Esquema general de funcionamiento de los algoritmos genéticos

Existen dos interpretaciones para el concepto de individuo y población en los sistemas evolutivos. La primera, habitualmente llamada Pittsburg [12], que entiende por población un conjunto de individuos.

La segunda interpretación, habitualmente llamada Michigan [12], consiste en considerar como individuo las habilidades de un controlador, de forma que la población representa el conjunto de todas sus habilidades. Según esta interpretación, el genoma es la codificación de estas habilidades, una habilidad puede estar formada por un conjunto de reglas que definen su comportamiento ante cierta situación. Este tipo de interpretación se ha utilizado mayormente en el control de sistemas en tiempo real y en la robótica.

Los algoritmos genéticos comienzan con una muestra elegida de manera aleatoria en el espacio de soluciones, para ir transformándola paso a paso evaluando cada individuo mediante una función de eficiencia o idoneidad, hasta llegar a un estado estacionario, en el cual la muestra, también llamada población, está constituida por las soluciones del problema. Un AG no actúa en general directamente sobre el espacio de soluciones, sino sobre una codificación de éste, haciendo interactuar entre sí y operando sobre las cadenas resultantes de dicha codificación. Esta acción tiene como consecuencia la progresiva transformación del conjunto de elementos evaluados, hasta llegar al estado final. Para ello, los algoritmos genéticos se valen de la acción de los llamados operadores genéticos, tales como selección, cruce y mutación, a la que es sometida la población. Cada operador juega un papel diferente en la evaluación de la población. Así, la selección puede ser entendida como una competencia, mientras que los otros operadores, tales como el cruce y la mutación son capaces de crear nuevos individuos a partir de los existentes. La Figura 1.2 muestra el funcionamiento genérico de los AG.

1.5.1 Operadores genéticos

1.5.1.1 Selección universal estocástica de Baker [13]

En este método los individuos son representados como segmentos continuos de una línea, donde cada tamaño de segmento es equivalente a su valor de ajuste. Sobre la línea de forma equidistante son colocados tanta cantidad de apuntadores como individuos se desee seleccionar. Considerando P

como el número de individuos a seleccionar, $1/P$ la distancia entre los apuntadores, y la posición del primer apuntador será un número aleatorio generado en el rango $[0, 1/P]$.

Si se seleccionan 6 individuos de la Tabla 1, la distancia entre los apuntadores seria 1/6 = 0.167. La Figura 1.3 representa la selección para la muestra de número aleatorio 0.1 en el rango $[0, 0.167]$.

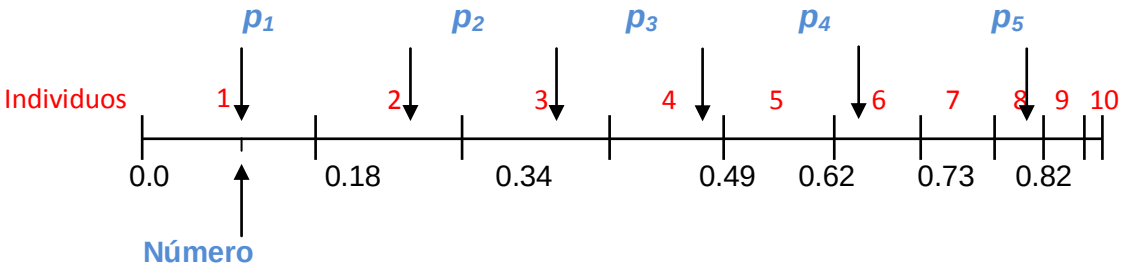

Figure 1.3 Selección universal estocástica

Individuos	1	2	3	4	5	6	7	8	9	10	11
Valor de ajuste	2.0	1.8	1.6	1.4	1.2	1.0	0.8	0.6	0.4	0.2	0.0
Probabilidad de selección	0.18	0.16	0.15	0.13	0.11	0.09	0.07	0.06	0.03	0.02	0.0

Tabla 1. Valores de ajuste y probabilidad de selección.

Como resultado de la selección se obtendrían los individuos: 1, 2, 3, 4, 6, 8.

1.5.1.2 Cruzamiento en dos puntos (two-point crossover)

Los algoritmos evolutivos utilizan el cruzamiento para crear una nueva solución, combinando algunas características de cada cromosoma padre. Esta combinación padres para este operador en particular

se produce seleccionando dos puntos de cruce en las cadenas cromosómicas padres e intercambiando las partes que quedan entre ellos, generando de esta forma dos cadenas hijas. La Tabla 2 muestra cómo trabaja operador de cruzamiento en dos puntos sobre cadenas binarias de cromosomas.

	1-punto	2-puntos
Antes del Cruzamiento	**0011** | **011010**	**001** | **101** | **1010**
	1110 | 010001	111 | 001 | 0001
Después del Cruzamiento	**0011** | 010001	**001** | 001 | **1010**
	1110 | **011010**	111 | **101** | 0001

Tabla 2. Operador de cruzamiento en cadenas binarias de cromosomas

1.5.1.3 Cruzamiento aritmético max-min (max-min-arithmetical crossover)

Este operador de cruzamiento fue propuesto por Herrera [14] y ha sido ampliamente usado en el campo de los Sistemas Difusos Evolutivos [15] [16]. Trabaja de la siguiente forma:

Si $C_v = \{c_1, \dots, c_k, \dots, c_H\}$ y $C'_w = \{c'_1, \dots, c'_k, \dots, c'_H\}$ son los cromosomas padres, los hijos resultantes del cruzamiento aritmético max-min serán los dos mejores de la siguiente descendencia:

$$O_1 = aC_w + (1-a)C_v$$
$$O_2 = aC_v + (1-a)C_w$$
$$O_3\ con\ c_{3k} = \min\{c_k, c'_k\}$$
$$O_4\ con\ c_{4k} = \max\{c_k, c'_k\}$$

siendo a una variable cuyo valor depende de la edad de la población.

Mutación uniforme [17]

Se realiza una prueba aleatoria para decidir cuál de los genes serán mutados.

Dado:

$P' = [V_1, \dots, V_k', \dots, V_m]$

Donde:

$V_k' = rnd(LB, UB)$

Se usa una distribución uniforme y [*LB, UB*] definen los rangos mínimos y máximos de la variable V_k'

Ejemplo:

$P = (5.3, -1.3, 7.8, 9.1)$

$V_k = 5.3 \;\; LB = 0.0 \;\; UB = 10.5$

$V_k' = rnd(0.0, 10.5) = 4.3$

1.5.1.4 Mutación no uniforme de Michalewicz

Michalewicz [18] propuso un operador dinámico de mutación no uniforme, diseñado de la siguiente forma:

Para cada individuo X_i^t en una población de t generaciones, crear un descendiente X_i^{t+1} a través de mutación no uniforme: si $X_i^t = \{x_1, x_2, \dots, x_m\}$ es un cromosoma (t el número de generación) y el elemento x_k es seleccionado para la mutación, el resultado es un vector $X_i^t = \{x_1', x_2', \dots, x_m'\}$ donde:

$$x_k' = \begin{cases} x_k + \Delta(t, UB - x_k), si\ el\ valor\ aleatorio\ \xi = 0 \\ x_k - \Delta(t, x_k - LB), si\ el\ valor\ aleatorio\ \xi = 1 \end{cases}$$

siendo LB y UB los límites inferiores y superiores de la variable x_k respectivamente. La función $\Delta(t, y)$ retorna un valor en el rango de $[0, y]$ tal que mientras $\Delta(t, y)$ se aproxima a cero el valor de t aumenta. Esta propiedad hace que el operador realice una búsqueda global uniforme al inicio (cuando t es pequeña), y local en los estados posteriores. Esta estrategia incrementa la probabilidad de generar un nuevo número más cercano a su sucesor que de una forma aleatoria. La siguiente función es usada:

$$\Delta(t, y) = y \cdot (1 - r^{(1-\frac{t}{T})^b})$$

donde r es un número aleatorio uniforme entre $[0,1]$, T es el número máximo de generaciones, y b es un parámetro del sistema que determina el grado de dependencia en el número de la iteración.

1.6 Software que contienen algoritmos genéticos

Herramientas como Weka y RapidMiner son líderes mundiales como aplicaciones para la extracción de conocimiento. Weka posee herramientas útiles para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación, visualización. RapidMiner por su parte, es una herramienta flexible para el aprendizaje y la exploración, sin embargo el enfoque que ambas utilizan para la generación de reglas no es difuso, por lo que no se ajustan a la investigación.

Otra herramienta a considerar es KEEL (Knowledge Extraction based on Evolutionary Learning), un software para evaluar algoritmos evolutivos en problemas de minería de datos que incluye regresión, clasificación y agrupamiento. Contiene una gran colección de algoritmos clásicos de extracción de conocimiento. Incluye además inteligencia computacional basada en algoritmos de aprendizaje evolutivos como son las reglas de aprendizaje y modelos híbridos como sistemas difusos que emplean algoritmos genéticos, las redes neuronales evolutivas, entre otros [(19)].

Dicha aplicación tiene implementados algoritmos para el aprendizaje de sistemas basados en reglas difusas para regresión tales como: Learning TSK-Fuzzy Models Based on MOGUL (MOGUL-TSK) [(20)], Iterative Rule Learning of Mamdani Rules - High Constrained Approach (MOGUL-IRLHC) [(21)], Genetic Fuzzy Rule Learning, COR Algorithm (COR-GA) [(22)], y Fuzzy and Random Sets Based Modeling (FRSBM) [(23)].

Los algoritmos antes mencionados combinan la lógica difusa y el empleo de algoritmos genéticos para la búsqueda de soluciones en un amplio espacio muestral, lo que hace factible su uso para el desarrollo de nuestra investigación.

Capítulo 2: Materiales y Métodos

En este capítulo se describen tecnologías y herramientas a emplear en la realización de la investigación. Además, se mencionan aspectos esenciales que presentan los algoritmos seleccionados en su fundamentación matemática.

2.1. Herramientas

Para el desarrollo de la investigación se utilizan varias herramientas cuya selección se realiza de acuerdo a las necesidades y problemas a resolver. A continuación se muestran algunas características de las mismas:

2.1.1. Herramienta de softcomputing (KEEL. v1.0)

KEEL [19] es una herramienta de software desarrollada para construir y utilizar diferentes modelos. Las principales características de KEEL son:

- Contiene algoritmos de pre-procesado: transformación y selección.
- Contiene una Biblioteca de Algoritmos de Extracción del Conocimiento, supervisados y no supervisados, además de algoritmos evolutivos de aprendizaje múltiple.
- Posee una biblioteca de análisis estadístico para analizar las salidas de los algoritmos.

Su entorno gráfico se distingue por tres partes esenciales:

- **Manejo de Datos:** A través de esta interfaz el usuario puede crear nuevos juegos de datos o particiones de uno existente. Asimismo, puede ver o modificar la información de su juego de datos, con la única restricción de que deben cumplir el formato del KEEL.
- **Experimentos:** Tiene el objetivo de diseñar los experimentos que se deseen usando una interfaz gráfica. El objetivo es utilizar juegos de datos y los algoritmos disponibles para generar una estructura de directorio con todos los archivos necesarios para ejecutar los experimentos diseñados en el equipo seleccionado por el usuario.

- **Docente:** Esta brinda la posibilidad de diseñar, correr los experimentos que se deseen y ver los resultados en el mismo entorno, con el inconveniente de tener un número muy reducido de algoritmos disponibles así como no contar con pruebas estadísticas. Debido a estos inconvenientes no será utilizada esta parte en la investigación.

2.1.2. Herramienta estadística (SPSS. v13.0)

Paquete Estadístico para las Ciencias Sociales (Statistical Package for the Social Sciences) es un programa estadístico muy usado en las ciencias sociales y las empresas de investigación.

Es muy popular su uso, debido a la capacidad de trabajar con bases de datos de gran tamaño. Además, de permitir la decodificación de las variables y registros según las necesidades del usuario. El programa consiste en un módulo base y módulos anexos con nuevos procedimientos estadísticos.

El sistema de módulos de SPSS, como los de otros programas (similar al de algunos lenguajes de programación) provee toda una serie de capacidades adicionales a las existentes en el sistema base. Algunos de los módulos disponibles son:

- **Modelos de Regresión**
- **Modelos Avanzados**
  - **Pruebas no paramétricas**: Permite realizar distintas pruebas estadísticas especializadas como:
    - **Levene** para comprobar la homogeneidad de varianzas con un nivel de significación de 0.05
    - **Friedman** es utilizado para comparar varias medias, cuando son del mismo grupo. Pone a prueba la hipótesis nula de que k variables proceden de la misma población. Nivel de significación de 0.05
    - **Dunnett T3** de comparación múltiple para identificar subconjuntos homogéneos, y arroja una matriz dónde los asteriscos indican las medias entre los grupos significativamente diferentes, a un nivel alfa de 0.05.

- **Tau_b de Kendall** y **Rho de Spearman** de correlación entre variables. Brinda una medida de cómo están relacionadas las variables.

2.1.3. Herramienta de ploteo (SigmaPlot 2001)

SigmaPlot está diseñado para mostrar gráficos de una forma clara y precisa. El programa está pensado para hojas de cálculo de forma que pueda elegir de entre un gran rango de opciones gráficas; escalas de ejes, múltiples ejes, múltiples intersecciones en gráficos de 3D, entre otros.

En SigmaPlot, se puede exportar gráficos como objetos dinámicos Web y colocarlos posteriormente en un sitio Web o una página Intranet en lugar de utilizar simples gráficos tipo GIF o JPEG.

Características principales:

- Software Gráfico que hace la visualización fácil de tareas.
- Personalización de cada detalle de los gráficos.
- Dibuja rápidamente sus datos desde plantillas gráficas.
- Combina las capacidades estadísticas de SPSS con SigmaPlot.

2.2. Métodos

2.2.1. Algoritmo MOGUL – TSK

Este algoritmo realiza una identificación local de modelos para inducir la competencia entre las reglas, considerando únicamente la calidad de la aproximación realizada por cada regla. El método consta de tres fases: generación inicial de reglas TSK, simplificación y afinación.

1. Proceso Local para Identificar Modelos. Este método induce competencia entre las reglas, teniendo en cuenta la integridad y la coherencia. La integridad se comprueba haciendo que cada ejemplo esté cubierto en un grado "épsilon". Para comprobar la coherencia, son considerados los conceptos de ejemplo positivo y negativo. Por lo tanto, la exactitud de una regla difusa, Rh, en el conjunto de ejemplos, E, se mide utilizando una función de idoneidad multicriterio diseñada teniendo

en cuenta tres criterios: alta frecuencia, alto valor promedio del grado a cubrir sobre los ejemplos positivos y pequeño conjunto de ejemplos negativos.

Este método puede resumirse en los siguientes pasos:

- Definir una fuerte borrosa partición para cada variable (tipo de función de pertenencia triangular distribuida uniformemente).
- Generar para cada ejemplo la regla difusa que mejor lo incluye. Luego, se evaluarán todas las reglas borrosas globales y se seleccionará la más prometedora.
- Esta regla es ajustada localmente para identificar el modelo difuso local que mejor agrupa los datos en el correspondiente sub-espacio. Este proceso es realizado considerando una función de idoneidad con una función de penalización para evitar la excesiva proximidad entre los modelos.
- Por último, el prototipo obtenido se añade al conjunto final de modelos difusos. Los datos incluidos en este conjunto, no se tendrán en cuenta para futuras iteraciones. El proceso iterativo termina cuando no hay más datos sin cubrir.

Para obtener los consecuentes, una vez que se obtiene el conjunto de modelos difusos locales y teniendo en cuenta los mismos antecedentes de las reglas, es calculada la relación lineal entrada-salida [24].

2. Proceso Genético de Selección: El proceso de simplificación está basado en un AG binario con cromosomas de longitud fija. Teniendo en cuenta las reglas contenidas en el conjunto solución derivado de la etapa anterior enumeradas de 1 hasta m, una cadena $C = (c_1,\ldots, c_m)$ representa un subconjunto de reglas candidatas a formar la base de reglas difusas finalmente obtenida como salida de este estado, de modo que, *SI* $\square_\square = 1$ *entonces* $\square_\square$ se incluye en el conjunto de reglas seleccionado.

La selección de los individuos es desarrollada utilizando Selección Universal Estocástica de Baker (sección 1.5.1.1), y los métodos de cruzamiento en dos puntos (sección 1.5.1.2) y mutación uniforme (sección 0) como operadores genéticos. La función de ajuste está basada en el error cuadrático medio

sobre el conjunto de entrenamiento (para premiar la similitud entre el modelo y el conjunto de datos) y una medida de inclusión (la cual penaliza la falta de integridad de las reglas).

3. Proceso Genético de Ajuste: Se basa en un híbrido AG-ES en el que cada individuo representa una completa base de conocimientos. En este método, el intervalo de variación es estimado para cada conjunto difuso.

2.2.2. Algoritmo MOGUL – IRLHC

Este método se basa en un proceso evolutivo de tres etapas:

Paso 1: Un método de generación, el cual permite obtener automáticamente un conjunto preliminar de reglas de tipo Mamdani para un problema concreto a partir su conjunto de datos.

Paso 2: Un proceso de simplificación que busca las mejores reglas en el conjunto de reglas borrosas obtenidos en el paso anterior, minimizando el valor del error cuadrático medio (MSE).

Este proceso de simplificación está basado en un AG binario con cromosomas de longitud fija.

Paso 3: Un proceso de ajuste que adapta las funciones de pertenencia de cada regla difusa del conjunto de reglas obtenido tras el proceso de simplificación.

Este proceso genético de ajuste se basa en un algoritmo de codificación real, donde cada parámetro estimado de la función de pertenencia para cada variable de las reglas es codificado. La función de ajuste está compuesta únicamente del criterio de la media cuadrática.

2.2.3. Algoritmo COR –GA

Esta nueva metodología para el aprendizaje rápido y la generación precisa y simple de los modelos lingüísticos difusos, llamada metodología cooperación (COR). Actúa sobre los consecuentes de las reglas difusas para encontrar las que mejor cooperan.

Dicho algoritmo en lugar de seleccionar el consecuente con más alto rendimiento en cada sub-espacio como de costumbre, considera la posibilidad de usar otro consecuente diferente del mejor, y permite

que el FRBS sea más preciso gracias a tener una base de reglas con una mejor cooperación. Con este propósito, COR realiza una búsqueda combinatoria entre las reglas candidatas buscando el conjunto de consecuentes que globalmente alcance la mejor exactitud.

2.2.4. Algoritmo FRSBM

Para construir este tipo de modelos, se divide en pequeños problemas mediante la separación del conjunto de ejemplos E en N subconjuntos E_i y luego se construyen N diferentes modelos difusos por separado.

Se establece una partición A de E para cada elemento de E_i cuyas funciones de ajuste determinarán el modelo difuso asociado a A.

Una vez relacionado cada modelo difuso con una partición, se determina la partición mejor relacionada con el modelo difuso más apropiado mediante un método de búsqueda que consiste en la generación de todas las particiones posibles de E, probando con cada modelo y selecciona por último el mejor.

Capítulo 3: Resultados y Discusión

En este capítulo se reflejan los principales resultados. Se presentan las pruebas realizadas en la creación de los modelos difusos y en el proceso de predicción, así como el estudio de regresión realizado a un conjunto de muestras.

3. Pruebas

La base de datos para las pruebas contiene dos muestras reales que describen la actividad biológica de las Cefalosporinas frente a *Escherichia coli* y *Staphylococcus aureus*, con un total de 104 entradas y 11 descriptores (ver **Anexo 3**). Fueron evaluados los algoritmos MOGUL-TSK, MOGUL-IRLHC, COR-GA y FRSBM. Se emplea K– Fold como técnica de validación cruzada donde se generan 10 ficheros de entrenamiento, con 94 vectores cada uno, y sus respectivos ficheros de pruebas, con 10 vectores para cada algoritmo. El análisis de estos ficheros arrojó los siguientes resultados:

3.1. Pruebas de predicción para la muestra Staphylococcus aureus (SA)

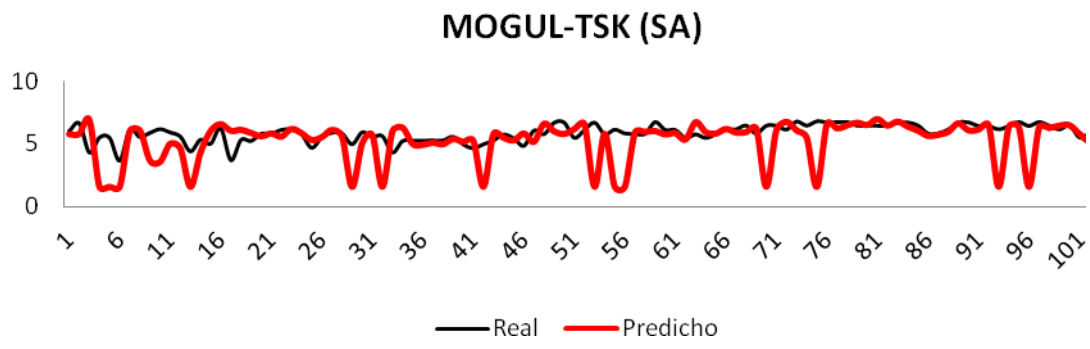

Figura 3.1 Representación de los valores predichos

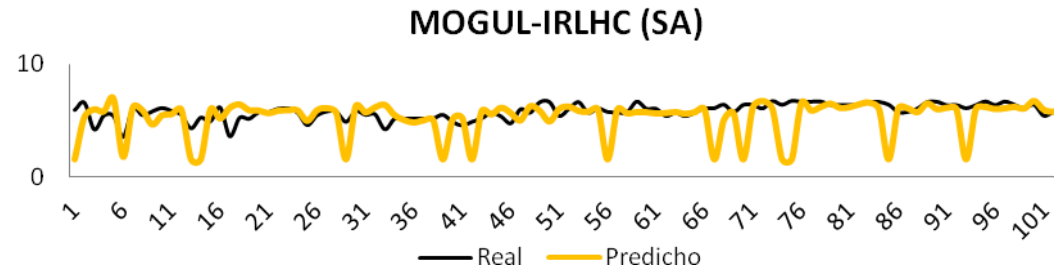

Figura 3.2 Representación de los valores predichos

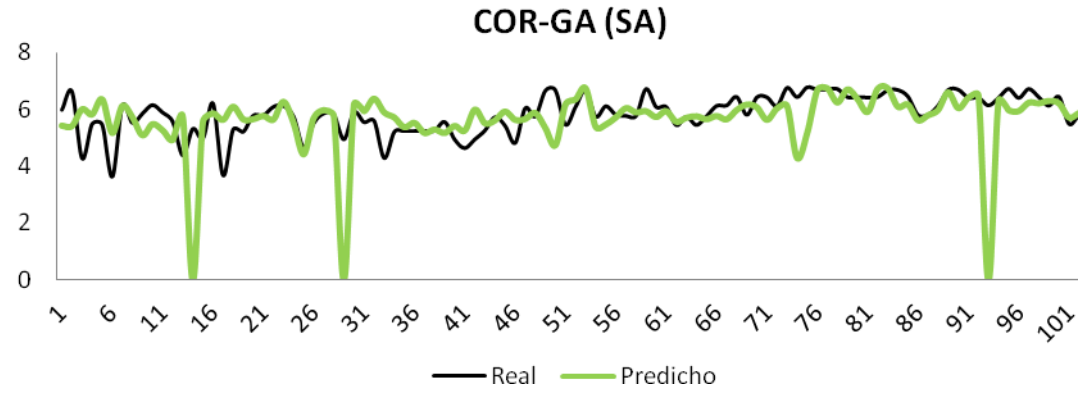

Figura 3.3 Representación de los valores predichos

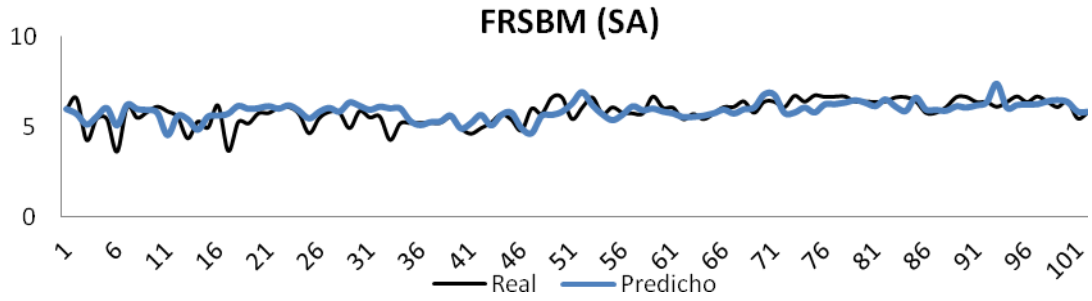

Figura 3.4 Representación de los valores predichos

Las Figuras 3.1 - 3.4 son una representación por algoritmo del comportamiento de las variables Real (valores experimentales) y sus respectivos valores predichos. Observándose un mejor ajuste entre las curvas de la Figura 3.4 correspondiente al algoritmo FRSBM.

3.1.1. Comparación entre Algoritmos

En los gráficos anteriores se mostraron los resultados de los algoritmos por separados. A continuación se establece una comparación para determinar a través de cuál se obtienen predicciones de mejor calidad.

La siguiente tabla contiene la media de los valores predichos de cada algoritmo.

Algoritmo	Valor Predicho
FRSBM	5,8492
MOGUL-IRLHC	5,3213
COR-GA	5,6736
MOGUL-TSK	5,2696
Valor Real	5,8228

Tabla 3. Media de los valores predichos (SA)

La Figura 3.5 contiene una representación más clara de la tabla anterior.

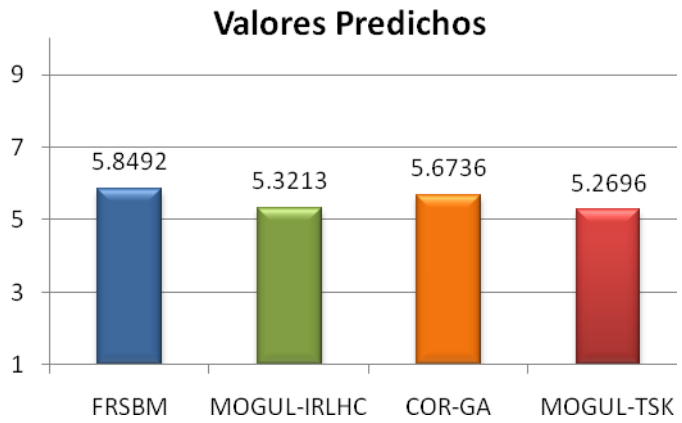

Figura 3.5 Comparación de resultados (SA)

Para constatar con mayor exactitud si existen o no diferencias entre las salidas de los algoritmos se realiza la prueba Levene para la Homogeneidad de Varianzas (Tabla 4) obteniéndose un valor de significación menor que 0,05 por lo cual no se realiza un ANOVA paramétrico sino una prueba no paramétrica equivalente, en este caso un test de Friedman para K-media relacionadas.

Error

Estadístico de Levene	gl1	gl 2	Sig.
17.203	3	412	.000

Tabla 4. Prueba de Homogeneidad de Varianzas (SA)

N	104
Chi-Cuadrado	2.681
gl	3
Sig.	.444

Tabla 5 Prueba de Friedman (SA)

La Tabla 5 muestra que la prueba de Friedman resulta no significativa y por lo tanto se acepta que no existen diferencias globales entre los errores medios de los algoritmos. A pesar de lo anterior se realiza una prueba Post Hoc de comparación múltiple para determinar cuáles difieren internamente.

Las pruebas contenidas en la Tabla 6 muestran que el valor de significación entre los algoritmos MOGUL-TSK y MOGUL-IRLHC es de 1.000 apreciándose un comportamiento similar, no siendo así el comportamiento de los anteriores respecto al FRSBM, encontrándose diferencias significativas en las medias de los errores. Finalmente el algoritmo COR-GA presenta valores por encima de 0.05 por lo que las diferencias no son estadísticamente significativas respecto a los restantes.

Variable Dependiente: Error

	(I) Algoritmo	(J) Algoritmo	Diferencia de medias (I-J)	Error Típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
Dunnett T3	MOGUL-TSK	MOGUL-IRLHC	,00400	,03284	1,000	-,0832	,0912
		COR-GA	,06408	,02941	,169	-,0141	,1422
		FRSBM	,09641(*)	,02496	,001	,0298	,1630
	MOGUL-IRLHC	MOGUL-TSK	-,00400	,03284	1,000	-,0912	,0832
		COR-GA	,06008	,02919	,220	-,0175	,1376
		FRSBM	,0924 (*)	,02470	,002	,0265	,1583
	COR-GA	MOGUL-TSK	-,06408	,02941	,169	-,1422	,0141
		MOGUL-IRLHC	-,06008	,02919	,220	-,1376	,0175
		FRSBM	,03233	,01991	,488	-,0207	,0854
	FRSBM	MOGUL-TSK	-,09641(*)	,02496	,001	-,1630	-,0298
		MOGUL-IRLHC	-,09241(*)	,02470	,002	-,1583	-,0265
		COR-GA	-,03233	,01991	,488	-,0854	,0207

* **La diferencia entre las medias es significativa al nivel 0.05.**

Tabla 6. Resultados de Pruebas Post Hoc de comparaciones múltiples (SA)

3.1.2. Estudio de regresión

Correlaciones No-paramétricas

			Real	MOGUL-TSK	MOGUL-IRLHC	COR-GA	FRSBM
Kendall's tau_b	**Real**	Coef. de Correlación	1,000	,267(**)	,183(**)	,315(**)	,387(**)
		Sig. (2-tailed)	0.000	,000	,006	,000	,000
		N	104	104	104	104	104
Spearman's rho	**Real**	Coef. de Correlación	1,000	,384(**)	,259(**)	,442(**)	,555(**)
		Sig. (2-tailed)	0.000	,000	,008	,000	,000
		N	104	104	104	104	104

**** La correlación es significativa al nivel 0.01. (2-colas).**

Tabla 7. Correlación entre los valores Real y predicho de cada algoritmo (SA) (ver Anexo 1**)**

Aunque los cuatro algoritmos presentan una correlación media y altamente significativa entre los valores experimentales y los predichos, el de mayor coeficiente de correlación tiende a ser FRSBM lo cual implica una mayor capacidad de predicción.

3.2. Pruebas de predicción para la muestra Escherichia coli (EC)

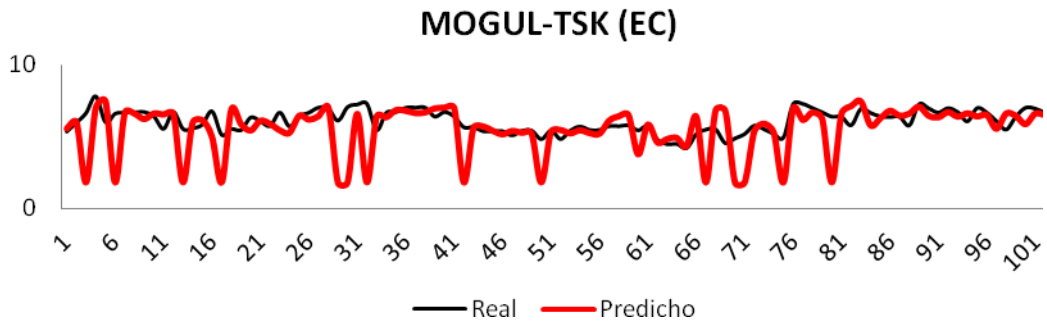

Figura 3.6 Representación de los valores predichos

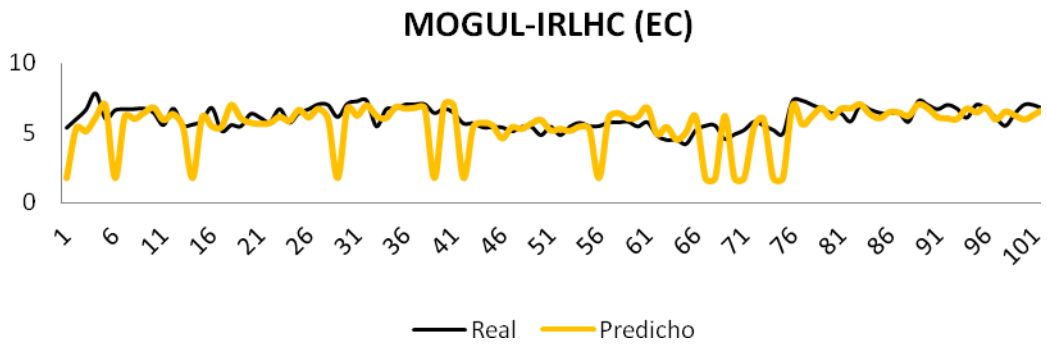

Figura 3.7 Representación de los valores predichos

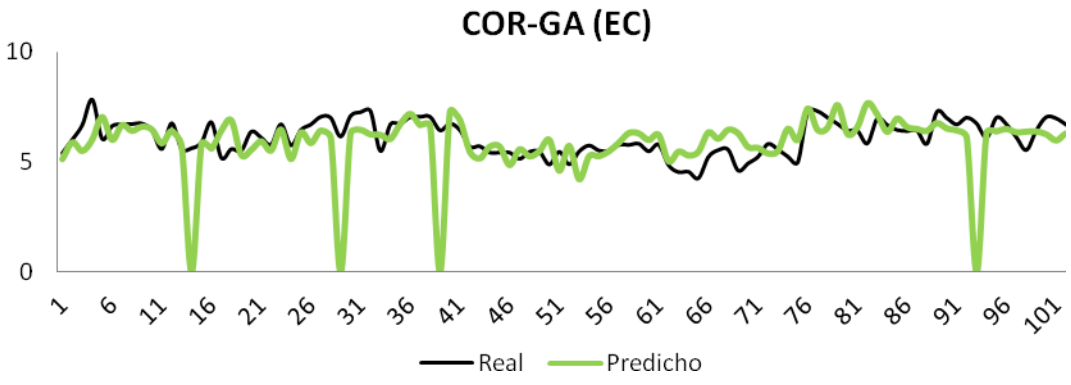

Figura 3.8 Representación de los valores predichos

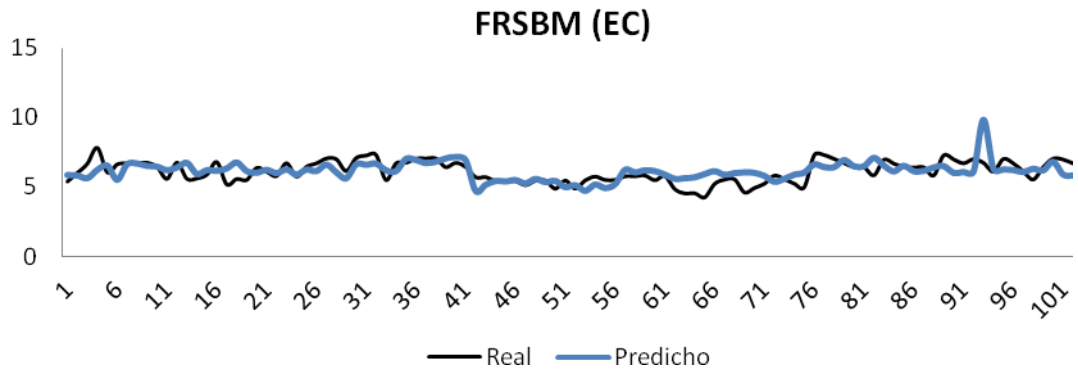

Figura 3.9 Representación de los valores predichos

Las Figuras 3.6 - 3.9 son una representación por algoritmo del comportamiento de las variables Real (valores experimentales) y sus respectivos valores predichos. Observándose un mejor ajuste entre las curvas de la Figura 3.4 correspondiente al algoritmo FRSBM.

3.2.1. Comparación entre Algoritmos

Como resultado del análisis anterior se establece una comparación para determinar a través de cual se obtienen predicciones de mejor calidad.

La siguiente tabla contiene la media de los valores predichos de cada algoritmo.

Algoritmo	Valor Predicho
FRSBM	6,1260
MOGUL-IRLHC	5,5567
COR-GA	5,8939
MOGUL-TSK	5,5307
Valor Real	6,0913

Tabla 8. Media de los valores predichos (EC)

La Figura 3.10 contiene una representación más clara de la tabla anterior.

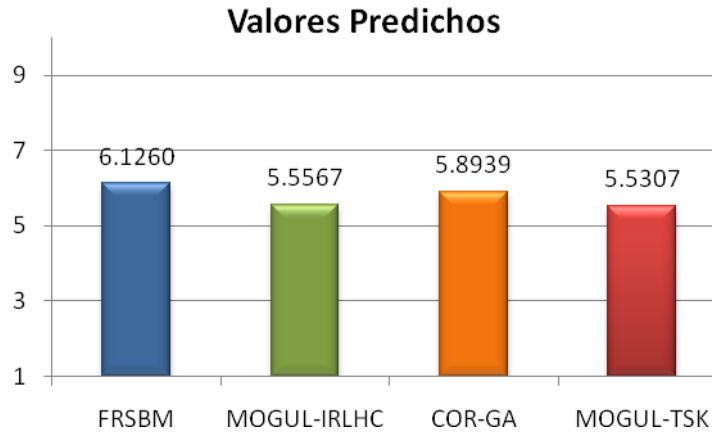

Figura 3.10 Comparación de resultados (EC)

Para comprobar si existen o no diferencias entre las salidas de los algoritmos se realiza la prueba Levene para la Homogeneidad de Varianzas (tabla 9) obteniéndose un valor de significación menor que 0,05 por lo cual no se realiza un ANOVA paramétrico sino una prueba no paramétrica equivalente, en este caso un test de Friedman para K-media relacionadas.

Error

Estadístico de Levene	gl1	gl 2	Sig.
11.876	3	412	.000

Tabla 9. Prueba de Homogeneidad de Varianzas (EC)

N	104
Chi-Cuadrado	4.835
gl	3
Sig.	.184

Tabla 10. Prueba de Friedman (EC)

La Tabla 10 muestra que la prueba resulta no significativa y por lo tanto se acepta que no existen diferencias globales entre los errores medios de los algoritmos. A pesar de lo anterior se realiza una prueba Post Hoc de comparación múltiple para determinar cuáles difieren internamente.

Las pruebas contenidas en la Tabla 11 muestran que el valor de significación entre los algoritmos MOGUL-TSK y MOGUL-IRLHC es de 1.000 apreciándose un comportamiento similar, no siendo así el comportamiento de los anteriores respecto al FRSBM encontrándose diferencias significativas en las medias de los errores. Finalmente el algoritmo COR-GA presenta valores por encima de 0.05 por lo que las diferencias no son estadísticamente significativas respecto a los restantes.

Variable Dependiente: Error

	(I) Algoritmo	(J) Algoritmo	Diferencia de medias (I-J)	Error Típico	Sig.	Intervalo de confianza al 95%	
						Límite inferior	Límite superior
Dunnett T3	MOGUL-TSK	MOGUL-IRLHC	.00327	.02973	1.000	-.0757	.0822
		COR-GA	.03415	.02873	.798	-.0422	.1104
		FRSBM	.06776 (*)	.02314	.024	.0060	.1295
	MOGUL-IRLHC	MOGUL-TSK	-.00327	.02973	1.000	-.0822	.0757
		COR-GA	.03087	.02763	.840	-.0425	.1043
		FRSBM	.06448 (*)	.02176	.021	.0064	.1226
	COR-GA	MOGUL-TSK	-.03415	.02873	.798	-.1104	.0422
		MOGUL-IRLHC	-.03087	.02763	.840	-.1043	.0425
		FRSBM	.03361	.02038	.469	-.0207	.0880
	FRSBM	MOGUL-TSK	-.06776 (*)	.02314	.024	-.1295	-.0060
		MOGUL-IRLHC	-.06448 (*)	.02176	.021	-.1226	-.0064
		COR-GA	-.03361	.02038	.469	-.0880	.0207

*** La diferencia entre las medias es significativa al nivel 0.05.**

Tabla 11. Resultados de Pruebas Post Hoc de comparaciones múltiples.

3.2.2. Estudio de regresión

Correlaciones No-paramétricas

			Real	MOGUL-TSK	MOGUL-IRLHC	COR-GA	FRSBM
Kendall's tau_b	**Real**	Coef. de Correlación	1.000	.371(**)	.267(**)	.269(**)	.429(**)
		Sig. (2-tailed)	0.000	.000	.000	.000	.000
		N	104	104	103	103	104
Spearman's rho	**Real**	Coef. de Correlación	1.000	.504(**)	.374(**)	.380(**)	.629(**)
		Sig. (2-tailed)	0.000	.000	.000	.000	.000
		N	104	104	103	103	104

**** La correlación es significativa al nivel 0.01. (2-colas).**

Tabla 12. Correlación entre los valores Real y predicho de cada algoritmo. (Ver Anexo 2**)**

Aunque los cuatro algoritmos presentan una correlación media y altamente significativa entre los valores experimentales y los predichos el de mayor coeficiente de correlación tiende a ser FRSBM lo cual implica una mayor capacidad de predicción.

3.3. Comparación General

La Tabla 13 recoge los porcientos de error de predicción por algoritmo en cada muestra.

Algoritmos	EC	SA
FRSBM	9,45%	7,91%
MOGUL-IRLHC	15,90%	17,16%
COR-GA	12,81%	11,15%
MOGUL-TSK	16,23%	17,56%

Tabla 13. Por ciento de error por algoritmo

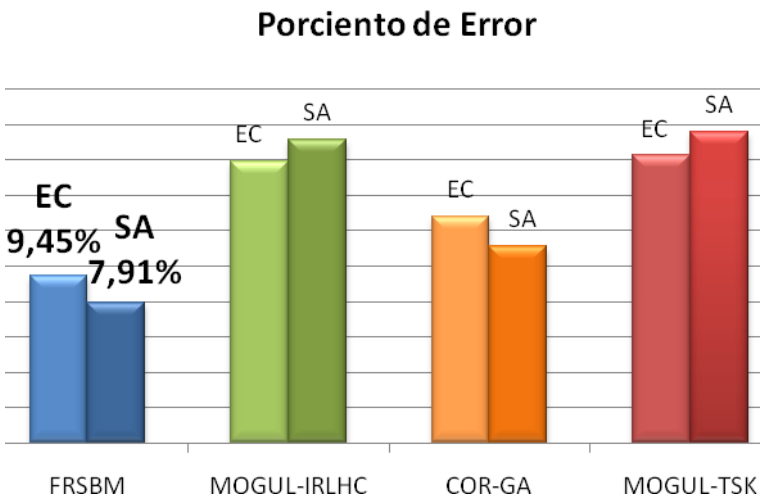

Figura 3.11 Por ciento de error por algoritmo

Como se puede observar en la Figura 3.11 el comportamiento de un mismo algoritmo frente a ambas muestras es similar, aunque se obtienen las mejores predicciones con el algoritmo FRSBM evidenciándose una mayor capacidad predictiva.

CONCLUSIONES

1. Se propone el algoritmo FRSBM por tener mayor coeficiente de correlación y menor porciento de error para las muestras de EC y SA con valores de 9.45 y 7.91% respectivamente.
2. Se realizó un estudio estructura actividad en una muestra heterogénea de cefalosporinas evaluadas en EC y SA demostrando la capacidad predictiva de la lógica difusa en problemas de regresión.

RECOMENDACIONES

1. Evaluar soluciones alternativas en grandes volúmenes de datos

2. Implementar e incorporar el algoritmo seleccionado a la plataforma “alasGRATO”

Tabla de Ilustraciones

BIBLIOGRAFÍA

1. **Leach, A. R.** *Molecular Modelling: Principles and Applications.* 2001. ISBN 0-582-38210-6.

2. **Zadeh, L.A.** *Fuzzy Logic, Neural Networks and Soft Computing.* s.l. : Communications of the ACM, 1994. pp. 77-84.

3. *Towards an Increased Role of Natural Language.* **Kacprzyk, J.** Granada : s.n., 2003, New Trends in Intelligent Systems and Soft Computing.

4. *Time-Domain Segmentation and Labelling of Speech with Fuzzy-Logic Post-Correction Rules.* **Mayora-Ibarra, O y Curatelli, F.** s.l. : Springer-Verlag Berlin Heidelberg, 2002, Vol. 2313, págs. 138-145.

5. **Piñero, Yobanis.** *Un modelo para el aprendizaje y la clasificación automática basado en técnicas de softcomputing.* s.l. : Universidad de Ciencias Informaticas, 2005. pp. 12-13.

6. *Aplicaciones de la Lógica Borrosa.* **Trillas, E.** 1992, Consejo Superior de Investigaciones Científicas.

7. **Martin del Brio, Bonifacio y Sanz Molina, Alfredo.** *Redes Neuronales y Sistemas Difusos.* Segunda Edicion. s.l. : Alfamega Ra-Ma, 2001. págs. 242-304.

8. *Application offuzzy control algothims for control of simple dinamic plant.* **Mamdani, H. E.** 12, 1974, Proc IEE, Vol. 121, págs. 1585-1588.

9. *Fuzzy Logic, A Spectrum of Theoretical & Practical Issues, .* s.l. : Springer-Verlag Berlin Heidelberg, 2007. ISSN: 1860-0808.

10. *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases.* **Cordón, O, y otros.** 2001, World Scientific.

11. **Schwefel, H. P y Männer, R.** *Parallel Problem Solving from Nature.* Berlin : Proceedings of the 1st Workshop PPSN I, 1991. págs. 307-313.

12. **Goldberg, D. E, Korb, B y Deb, K.** *Messy genetic algorithms:Motivation, analysis, and first result.Complex Systems.* 1989. págs. 493-530. Vol. 3.

13. **Baker, James, E.** *Reducing Bias and Inefficiency in the Selection Algorithm, en Proceedings of the Second International Conference on Genetic Algorithms and their Application.* Hillsdale : s.n., 1987. págs. 14-21.

14. *Tackling real-coded genetic algorithms: Operators and tools for the behaviour analysis.* **Herrera, F, Lozano, M, Verdegay J, L.** 12, 1998, Artificial Intelligence, pp. 265–319.

15. *A three-stage evolutionary process for learning descriptive and approximative fuzzy logic controller knowledge bases from examples.* **Cordon, O, Herrera, F.** Int J. Approx. Reas, Vol. 17, pp. 369–407.

16. *Applicability of the fuzzy operators in the design of fuzzy logic controllers.* **Cordon, O, Herrera, F, Peregrin, A.** 1997, Fuzzy Sets Syst., Vol. 86, pp. 15-41.

17. **Coello Coello, Carlos A.** *Introducción a la Computación Evolutiva (Notas de Curso).* 2006.

18. **Michalewicz, Z.** *Genetic Algorithms + Data Structures = Evolution Programs.* 3ra edicion. s.l. : Springer, 1996.

19. KEEL. *KEEL (Knowledge Extraction based on Evolutionary Learning).* [Online] 2004. [Cited: Octubre 25, 2008.] http://www.keel.es.

20. *Local Identification of Prototypes for Genetic Learning of Accurate TSK Fuzzy Rule-Based Systems.* **Alcalá, R, y otros.** 9, 2007, International Journal of Intelligent Systems, Vol. 22, págs. 909-941.

21. *Hybridizing Genetic Algorithms with Sharing Scheme and Evolution Strategies for Designing Approximate Fuzzy Rule-Based Systems.* **Cordón, O y Herrera, F.** 2, 2001, Fuzzy Sets and Systems, Vol. 118, págs. 235-255.

22. *COR: A methodology to improve ad hoc data-driven linguistic rule learning methods by inducing cooperation among rules.* **Casillas, J, Cordon, O y Herrera, F.** 4, 2002, IEEE Transactions on System and Man and Cybernetics and Part B: Cybernetics, Vol. 32, págs. 526-537.

23. **Sánchez, L.** *A Random Sets-Based Method for Identifying Fuzzy Models.* s.l. : Fuzzy Sets and Systems, 1998. págs. 343-354. Vol. 3.

24. *A Two-Stage Evolutionary Process for Designing TSK Fuzzy Rule-Based Systems.* **Cordon, O y Herrera, F.** 6, 1999, IEEE Transactions on Systems and Man and Cybernetics and Part B: Cybernetics, Vol. 29, págs. 703-715.

ENLACES DE INTERÉS

- http://www.softcomputing.es/es/portada.php Centro de investigación español, organiza cursos y seminarios sobre Fuzzy Logic.
- Fuzzy Logic - Lógica Difusa. Introducción a la lógica difusa y su relación con el control de procesos.
- Xfuzzy: Una herramienta de CAD gratuita para el diseño de sistemas de control basados en lógica difusa. Última versión, 3.0 en español.
- Lógica Difusa: ¿una concepción infinitesimal de la verdad?
- Morillas Raya, A. (2006): Introducción al análisis de datos difusos. Texto completo en PDF.
- Curso Introductorio de Conjuntos y Sistemas Difusos (Lógica Difusa y Aplicaciones), por el Dr. José Galindo G., Universidad de Málaga (España).

GLOSARIO DE TÉRMINOS

Actividad Biológica: Actividad que caracteriza el comportamiento biológico en compuestos químicos.

Bioinformática: Es la aplicación de métodos informáticos sobre sistemas de cómputo y tratamiento de la información para el análisis de datos experimentales (de nivel molecular, principalmente) de sistemas biológicos, así como la simulación de los mismos.

Compuestos orgánicos: Son sustancias químicas basadas en cadenas de carbono e hidrógeno. En muchos casos contienen oxígeno, y también nitrógeno, azufre, fósforo, boro y halógenos.

Descriptor: Número que caracteriza estructuralmente la molécula.

Descriptores topológico: Número que se calcula a partir de algoritmos que se aplican a la matriz de adyacencia o de conectividad del grafo molecular desprovisto de hidrógeno.

Entrenamiento: Acción que se realiza para el aprendizaje de las neuronas.

Estadística: Es una rama de la matemática que se refiere a la recolección, estudio e interpretación de los datos obtenidos en un estudio.

Metodología: Define quién hace qué, cómo y cuándo.

Modelo: Representación abstracta de la realidad.

Modelo difuso: Modelo que se crea a partir del las reglas difusas generadas.

Molécula: La partícula más pequeña de una sustancia, que mantiene las propiedades químicas específicas de esa sustancia.

Multiplataforma: Término usado para referirse a los programas, sistemas operativos, lenguajes de programación u otra clase de software, que puedan funcionar en diversas plataformas.

Predicción: Acción que el sistema realiza para emitir un resultado.

Plataforma: Es el principio, ya sea de hardware o software, sobre el cual un programa puede ejecutarse.

Procedimiento Heurístico: Procedimiento para resolver un problema de optimización bien definido mediante una aproximación intuitiva, en la que la estructura del problema se utiliza de forma inteligente para obtener una buena solución.

Software: Término genérico que designa al conjunto de programas que posibilitan realizar una tarea específica en un ordenador.

ANEXOS

Anexo 1: Tabla de valores de correlación para la muestra de SA

			Real	MOGUL-TSK	MOGUL-IRLHC	COR-GA	FRSBM
Kendall's tau_b	Real	Coeficiente Correlación	1,000	,267(**)	,183(**)	,315(**)	,387(**)
		Sig. (2-tailed)	.	,000	,006	,000	,000
		N	104	104	104	104	104
	MOGUL-TSK	Coeficiente Correlación	,267(**)	1,000	,453(**)	,442(**)	,297(**)
		Sig. (2-tailed)	,000	.	,000	,000	,000
		N	104	104	104	104	104
	MOGUL-IRLHC	Coeficiente Correlación	,183(**)	,453(**)	1,000	,481(**)	,344(**)
		Sig. (2-tailed)	,006	,000	.	,000	,000
		N	104	104	105	105	104
	COR-GA	Coeficiente Correlación	,315(**)	,442(**)	,481(**)	1,000	,461(**)
		Sig. (2-tailed)	,000	,000	,000	.	,000
		N	104	104	105	105	104
	FRSBM	Coeficiente Correlación	,387(**)	,297(**)	,344(**)	,461(**)	1,000
		Sig. (2-tailed)	,000	,000	,000	,000	.
		N	104	104	104	104	104
Spearman's rho	Real	Coeficiente Correlación	1,000	,384(**)	,259(**)	,442(**)	,555(**)
		Sig. (2-tailed)	.	,000	,008	,000	,000
		N	104	104	104	104	104
	MOGUL-TSK	Coeficiente Correlación	,384(**)	1,000	,592(**)	,572(**)	,392(**)
		Sig. (2-tailed)	,000	.	,000	,000	,000
		N	104	104	104	104	104
	MOGUL-IRLHC	Coeficiente Correlación	,259(**)	,592(**)	1,000	,654(**)	,469(**)
		Sig. (2-tailed)	,008	,000	.	,000	,000
		N	104	104	105	105	104
	COR-GA	Coeficiente Correlación	,442(**)	,572(**)	,654(**)	1,000	,622(**)
		Sig. (2-tailed)	,000	,000	,000	.	,000
		N	104	104	105	105	104
	FRSBM	Coeficiente Correlación	,555(**)	,392(**)	,469(**)	,622(**)	1,000
		Sig. (2-tailed)	,000	,000	,000	,000	.
		N	104	104	104	104	104

**** La correlación es significativa al nivel 0.01. (2-colas).**

Anexo 2: Tabla de valores de correlación para la muestra de EC

			Real	MOGUL-TSK	MOGUL-IRLHC	COR-GA	FRSBM
Kendall's tau_b	Real	Coeficiente Correlación	1.000	.371	.267	.269	.429
		Sig. (2-tailed)	.	.000	.000	.000	.000
		N	104	104	103	103	104
	MOGUL-TSK	Coeficiente Correlación	.371	1.000	.148	.210	.422
		Sig. (2-tailed)	.000	.	.029	.002	.000
		N	104	104	103	103	104
	MOGUL-IRLHC	Coeficiente Correlación	.267	.148	1.000	.579	.230
		Sig. (2-tailed)	.000	.029	.	.000	.001
		N	103	103	104	104	103
	COR-GA	Coeficiente Correlación	.269	.210	.579	1.000	.310
		Sig. (2-tailed)	.000	.002	.000	.	.000
		N	103	103	104	104	103
	FRSBM	Coeficiente Correlación	.429	.422	.230	.310	1.000
		Sig. (2-tailed)	.000	.000	.001	.000	.
		N	104	104	103	103	104
Spearman's rho	Real	Coeficiente Correlación	1.000	.504	.374	.380	.629
		Sig. (2-tailed)	.	.000	.000	.000	.000
		N	104	104	103	103	104
	MOGUL-TSK	Coeficiente Correlación	.504	1.000	.206	.308	.578
		Sig. (2-tailed)	.000	.	.037	.002	.000
		N	104	104	103	103	104
	MOGUL-IRLHC	Coeficiente Correlación	.374	.206	1.000	.762	.331
		Sig. (2-tailed)	.000	.037	.	.000	.001
		N	103	103	104	104	103
	COR-GA	Coeficiente Correlación	.380	.308	.762	1.000	.423
		Sig. (2-tailed)	.000	.002	.000	.	.000
		N	103	103	104	104	103
	FRSBM	Coeficiente Correlación	.629	.578	.331	.423	1.000
		Sig. (2-tailed)	.000	.000	.001	.000	.
		N	104	104	103	103	104

**** La correlación es significativa al nivel 0.01. (2-colas).**

Anexo 3: Conjunto de datos

Cefalosporinas	POT_EC	POT_SA	SOM1	SOM2	SOM3	OMPCL04	SN_NETA4	DIED2	S1	S2	OP1	OP2	OP3
cefaclor *	5,371302	5,96236697	0,387	0,081	0	0	0	175,11	2,931	1,865	6,6119	5,14	3,2947
CEFDIMIR	5,993876	6,59593581	0,597	0,177	0,037	0	1,6011	179,03	3,409	2,271	5,0076	3,6532	2,6186
CEFIXIME	6,655557	4,25761715	0,597	0,177	0,037	0	1,6011	177,41	4,797	3,268	6,6119	5,14	3,2947
CEFOTAX	7,81239	5,46436342	1,767	1,071	0,372	0,6781	4,3409	178,26	3,827	2,501	5,5631	3,9902	2,7509
CEFPOD	6,027799	5,43673435	1,1	0,41	0,141	0	0,5218	163,75	3,827	2,501	5,5631	3,9902	2,7509
CEFTIBUTE*	6,61216	3,61215951	0,177	0	0	0	0	164,97	4,159	2,822	7,7116	5,7213	3,8758
CY1A	6,68529	6,09422559	2,091	1,161	0,477	0,259	-1,255	173,62	3,409	2,271	5,0076	3,6532	2,6186
CY1B	6,697685	5,5045606	2,091	1,161	0,477	0,259	-1,255	174,7	3,827	2,501	5,5631	3,9902	2,7509
CY1C	6,724718	5,83262353	2,091	1,161	0,477	0,259	-1,255	174,63	4,524	2,895	6,583	4,7436	3,2631
CY1D	6,418766	6,12873139	2,091	1,161	0,477	0,259	-1,255	174,74	4,265	2,775	6,1861	4,4599	3,0706
CY1E	5,563197	5,86422651	2,091	1,161	0,477	0,259	-1,255	174,45	5,436	3,704	7,7116	5,7213	3,8758
CY1F	6,733637	5,54051236	2,091	1,161	0,477	0,259	-1,255	173,8	4,709	3,162	6,5827	5,1086	3,282
CY1G	5,562429	4,3576132	2,091	1,161	0,477	0,259	-1,255	177,77	5,473	4,232	7,4004	6,601	4,4295
CY1H	5,597598	5,29517842	2,091	1,161	0,477	0,259	-1,255	176,14	6,419	4,405	8,8078	6,809	4,765
CY1J	5,861152	4,95736666	2,091	1,161	0,477	0,259	-1,255	173,79	5,546	3,736	7,4437	5,5943	4,0294
CY1K	6,775716	6,1846514	2,091	1,161	0,477	0,259	-1,255	173,19	5,493	3,717	7,8872	6,1588	4,3198
CY1L	5,154911	3,66006062	2,091	1,161	0,477	0,259	-1,255	178,84	4,285	2,785	6,5719	4,7367	3,2593
CY2B	5,559218	5,26918354	1,884	1,068	0,412	0,335	1,5065	171,95	3,409	2,271	5,0076	3,6532	2,6186
CY2C	5,480977	5,19094194	0,862	0,286	0,089	0	-0,674	174,31	3,409	2,271	5,0076	3,6532	2,6186
CY2D	6,348758	5,75769379	1,199	0,438	0,147	0	1,1917	174,78	3,409	2,271	5,0076	3,6532	2,6186
CY2E	6,081557	5,76924568	1,444	0,552	0,203	0	2,4903	174,26	3,409	2,271	5,0076	3,6532	2,6186
CY2F	5,768313	6,06934263	1,175	0,427	0,152	0	2,3792	174,66	3,409	2,271	5,0076	3,6532	2,6186
CY2G	6,691554	6,10048896	2,181	1,161	0,602	0,2453	1,7887	176,02	3,409	2,271	5,0076	3,6532	2,6186
CY2H	5,725375	5,72537545	0,519	0,143	0,03	0	0,8559	176,13	3,409	2,271	5,0076	3,6532	2,6186
CY3A	6,432589	4,6367085	2,091	1,161	0,477	0,259	-1,255	177,33	4,797	3,268	6,6119	5,14	3,2947
HP16A	6,679914	5,48678911	1,658	0,851	0,431	0,2546	-0,072	175,59	3,827	2,501	5,5631	3,9902	2,7509
HP16B	7,030658	5,8375334	3,075	1,964	1,103	0,8386	-0,883	175,49	3,827	2,501	5,5631	3,9902	2,7509
HP16C	6,981839	5,78871484	1,75	0,912	0,478	0,2563	4,8811	175,92	3,827	2,501	5,5631	3,9902	2,7509
HP16D	6,129599	4,9247836	2,144	1,202	0,663	0,5546	23,398	175,9	3,827	2,501	5,5631	3,9902	2,7509
HP16E	7,070171	5,87704631	3,489	2,005	1,133	0,7004	-2,507	179,97	3,827	2,501	5,5631	3,9902	2,7509
HP16F	7,240183	5,52417968	2,657	1,656	0,904	0,8337	-1,411	175,76	3,827	2,501	5,5631	3,9902	2,7509
HP16G	7,289049	5,57304575	3,847	2,41	1,465	1,2412	1,4134	178,26	3,827	2,501	5,5631	3,9902	2,7509
HP16H	5,471963	4,26714731	2,058	1,229	0,694	0,2597	1,0574	175,48	3,827	2,501	5,5631	3,9902	2,7509
HP16I	6,678963	5,18341903	2,452	1,595	0,966	0,4633	2,7449	175,37	3,827	2,501	5,5631	3,9902	2,7509
HP25A	6,718514	5,21419614	2,772	2,005	1,16	2,3814	-0,845	176,84	3,827	2,501	5,5631	3,9902	2,7509
HP25E	7,021248	5,21590013	2,387	1,363	0,656	2,3827	-2,009	177,91	3,827	2,501	5,5631	3,9902	2,7509
HP25F	7,031269	5,22592085	2,98	2,085	1,284	2,5714	-1,213	179,36	3,827	2,501	5,5631	3,9902	2,7509
HP25G	7,031269	5,22592085	2,988	2,103	1,182	2,5167	-1,144	179,15	3,827	2,501	5,5631	3,9902	2,7509
HP25L	6,409546	5,53654226	3,414	2,562	1,734	2,3827	-2,305	178,51	3,827	2,501	5,5631	3,9902	2,7509
HP25N	6,721051	4,91469595	2,387	1,363	0,656	2,3827	-2,009	176,56	3,915	2,593	5,5631	3,9902	2,7509
HP25P	6,421249	4,61536752	2,988	2,103	1,182	2,5167	-1,144	179,33	3,915	2,593	5,5631	3,9902	2,7509
KI16A	5,670675	4,92639925	0,177	0	0	0	0	179,47	3,202	2,14	7,8872	6,1588	4,3198
KI16B	5,687035	5,23492139	0,177	0	0	0	0	179,43	3,417	2,239	6,5719	4,7367	3,2593
KI16C	5,40177	5,70280047	0,177	0	0	0	0	179,15	3,633	2,339	6,6119	5,14	3,2947
Cefalosporinas	**POT_EC**	**POT_SA**	**SOM1**	**SOM2**	**SOM3**	**OMPCL04**	**SN_NETA4**	**DIED2**	**S1**	**S2**	**OP1**	**OP2**	**OP3**

Continuación

Cefalosporinas	POT_EC	POT_SA	SOM1	SOM2	SOM3	OMPCL04	SN_NETA4	DIED2	S1	S2	OP1	OP2	OP3
KI16D	5,40177	5,40177047	0,177	0	0	0	0	179,3	3,684	2,456	5,0177	3,658	2,6232
KI16E	5,399539	4,79678408	0,177	0	0	0	0	178,7	3,941	2,716	5,5585	3,9891	2,7497
KI16F	5,11241	6,01688984	0,177	0	0	0	0	179,45	4,156	2,816	6,0626	4,3706	2,9837
KI16G	5,4296	5,73063012	0,177	0	0	0	0	179,04	4,371	2,916	5,942	4,81	2,8456
KI16H	5,443884	6,63700814	0,177	0	0	0	0	179,32	4,587	3,016	6,1158	4,9141	3,101
KI16I	4,854958	6,65083768	0,177	0	0	0	0	179,71	4,802	3,116	6,5985	5,2877	3,4275
KI16J	5,437783	5,43778307	0,177	0	0	0	0	179,97	4,555	2,989	7,1251	5,6318	4,1949
KI16K	4,862633	6,06744856	0,177	0	0	0	0	179,24	4,986	3,189	7,6148	6,0138	4,5273
KI16L	5,451805	6,64492972	0,177	0	0	0	0	179,12	4,77	3,089	8,1096	6,362	4,7705
KI16M	5,722263	5,72226257	0,177	0	0	0	0	179,46	3,549	2,3	6,745	5,0341	3,5421
KI16N	5,489523	6,09158309	0,177	0	0	0	0	178,65	4,78	3,054	7,7363	5,7302	4,0809
KI16O	5,489523	5,7905531	0,177	0	0	0	0	173,35	4,794	3,081	7,2273	5,3842	3,8301
KI22A	5,748849	5,74884871	1,767	1,071	0,372	0,197	0,3135	176,31	3,202	2,14	5,0177	3,658	2,6232
KI22B	5,749827	5,74982745	1,782	1,095	0,378	0,189	0,3543	179,69	3,202	2,14	5,0177	3,658	2,6232
KI22C	5,803022	6,6951164	1,966	0,995	0,516	0,2802	-1,03	176,01	3,202	2,14	5,0177	3,658	2,6232
KI23A	5,461524	6,06358449	1,767	1,071	0,372	0,197	0,3135	176,08	3,417	2,239	5,5585	3,9891	2,7497
KI23B	5,763504	6,06453382	1,782	1,095	0,378	0,189	0,3543	179,79	3,417	2,239	5,5585	3,9891	2,7497
KI23D	4,820987	5,4237425	0,387	0,081	0	0	0	175,82	3,417	2,239	5,5585	3,9891	2,7497
KI23E	4,511044	5,71585984	0,597	0,177	0,037	0	1,601	177,52	3,417	2,239	5,5585	3,9891	2,7497
KI23F	4,525815	5,42960012	0,809	0,275	0,082	0	1,7027	176,08	3,417	2,239	5,5585	3,9891	2,7497
KI23G	4,228894	5,73473933	1,1	0,41	0,141	0	0,2128	175,99	3,417	2,239	5,5585	3,9891	2,7497
KI23H	5,196919	6,10139894	2,11	1,133	0,618	0,267	-0,368	175,92	3,417	2,239	5,5585	3,9891	2,7497
KI23I	5,512379	6,11443904	2,538	1,576	0,951	0,6801	13,704	178,19	3,417	2,239	5,5585	3,9891	2,7497
KI23J	5,525906	6,41800033	2,245	1,26	0,666	0,4963	4,4707	176,36	3,417	2,239	5,5585	3,9891	2,7497
KI23K	4,580258	5,78507327	2,418	1,598	0,952	0,4521	0,3419	179,04	3,417	2,239	5,5585	3,9891	2,7497
KI24A	4,907912	6,40276247	1,767	1,071	0,372	0,197	0,3135	174,63	4,587	3,016	7,6148	6,0138	4,5273
KI24B	5,209096	6,40361033	1,782	1,095	0,378	0,189	0,3543	179,73	4,587	3,016	7,6148	6,0138	4,5273
KI25B	5,793234	6,09426413	1,782	1,095	0,378	0,189	0,3543	178,15	3,549	2,3	7,7363	5,7302	4,0809
KI25C	5,54061	6,73373478	1,966	0,995	0,516	0,2802	0,3543	177,93	3,549	2,3	7,7363	5,7302	4,0809
KI26B	5,215883	6,41039759	1,782	1,095	0,378	0,189	-1,03	179,87	4,77	3,089	6,745	5,0341	3,5421
KI26I	4,959728	6,7556081	2,538	1,576	0,951	0,68	12,615	173,88	4,77	3,089	6,745	5,0341	3,5421
SN16C	7,271796	6,66973609	2,204	1,289	0,47	0,1892	0,5153	163,71	3,409	2,271	5,0076	3,6532	2,6186
SN16D	7,284636	6,68257555	2,444	1,437	0,646	0,3329	0,5203	171,54	3,409	2,271	5,0076	3,6532	2,6186
SN16E	6,996077	6,69504716	2,66	1,547	0,713	0,2772	0,5203	173,89	3,409	2,271	5,0076	3,6532	2,6186
SN16F	6,705449	6,40441889	2,87	1,645	0,762	0,277	0,5209	163,68	3,409	2,271	5,0076	3,6532	2,6186
SN16G	6,406141	6,40614062	2,93	1,776	0,797	0,469	0,5477	176,46	3,409	2,271	5,0076	3,6532	2,6186
SN16H	6,394017	6,39401717	2,755	1,737	0,874	0,8817	0,5088	174,02	3,409	2,271	5,0076	3,6532	2,6186
SN16I	5,814199	6,41625929	3,447	2,229	1,314	0,7899	0,5062	163,65	3,409	2,271	5,0076	3,6532	2,6186
SN16IA	6,945875	6,64484512	1,782	1,095	0,378	0,1889	0,3485	163,59	3,409	2,271	5,0076	3,6532	2,6186
SN16IB	6,658431	6,65843117	2,333	1,544	0,782	0,8875	0,3205	179,01	3,409	2,271	5,0076	3,6532	2,6186
SN16J	6,429382	6,4293816	3,957	2,534	1,502	0,7912	0,5235	179,51	3,409	2,271	5,0076	3,6532	2,6186
SN16M	6,381546	5,77948556	2,204	1,289	0,47	0,1908	0,5369	179,49	3,827	2,501	5,5631	3,9902	2,7509
SN16N	6,394017	5,79195718	2,444	1,437	0,646	0,3359	0,5498	177,19	3,827	2,501	5,5631	3,9902	2,7509
SN16P	5,804081	6,10511062	2,755	1,737	0,874	0,8875	0,5076	176,7	3,827	2,501	5,5631	3,9902	2,7509
SN29D	7,260491	6,65843117	2,022	1,243	0,554	0,3356	0,3239	163,76	3,409	2,271	5,0076	3,6532	2,6186
SN29E	6,972634	6,67160414	2,239	1,353	0,621	0,2793	0,3136	179,18	3,409	2,271	5,0076	3,6532	2,6186

Cefalosporinas	POT_EC	POT_SA	SOM1	SOM2	SOM3	OMPCL04	SN_NETA4	DIED2	S1	S2	OP1	OP2	OP3

Continuación

Cefalosporinas	POT_EC	POT_SA	SOM1	SOM2	SOM3	OMPCL04	SN_NETA4	DIED2	S1	S2	OP1	OP2	OP3
SN29F	6,68439	6,3833602	2,508	1,582	0,704	0,4672	0,3421	179,53	3,409	2,271	5,0076	3,6532	2,6186
SN29G	6,988966	6,38690632	2,707	1,571	0,731	0,2783	0,3472	178,82	3,409	2,271	5,0076	3,6532	2,6186
SN29H	6,718865	6,11680526	4,052	3,879	1,047	1,088	0,4975	178,9	3,409	2,271	5,0076	3,6532	2,6186
SN29J	6,08233	6,3833602	2,553	1,678	0,906	0,9269	0,3356	175,74	3,409	2,271	5,0076	3,6532	2,6186
SN29K	6,999559	6,69852941	2,862	1,824	0,994	0,8655	0,3823	176,99	3,409	2,271	5,0076	3,6532	2,6186
SN29L	6,682576	6,38154555	2,81	1,935	1,175	0,7862	0,3175	173,55	3,409	2,271	5,0076	3,6532	2,6186
SN29M	6,092987	6,69504716	3,025	2,035	1,221	0,7897	0,3354	179,17	3,409	2,271	5,0076	3,6532	2,6186
SN29N	5,514046	6,40614062	3,24	2,135	1,268	0,789	0,3129	179,22	3,409	2,271	5,0076	3,6532	2,6186
SN29O	6,407819	6,10678904	3,535	2,34	1,41	0,791	0,3792	178,92	3,409	2,271	5,0076	3,6532	2,6186
SN29P	7,020814	6,41875358	3,589	2,434	1,423	1,2142	0,3246	165,2	3,409	2,271	5,0076	3,6532	2,6186
SN29Q	6,959461	5,46530657	1,782	1,095	0,378	0,1889	0,3485	178,39	3,827	2,501	5,5631	3,9902	2,7509
SN29R	6,671604	5,76851416	2,022	1,243	0,554	0,3355	0,3229	178,28	3,827	2,501	5,5631	3,9902	2,7509
SN29S	6,38336	5,7813002	2,333	1,544	0,782	0,8882	0,3222	179,11	3,827	2,501	5,5631	3,9902	2,7509
SN29T	6,399227	5,79716709	2,333	1,544	0,782	0,8876	0,3238	179,15	4,308	2,87	5,5631	3,9902	2,7509
Cefalosporinas	POT_EC	POT_SA	SOM1	SOM2	SOM3	OMPCL04	SN_NETA4	DIED2	S1	S2	OP1	OP2	OP3