Programación Matemática y Software (2020) 12 (3): 1-13. ISSN: 2007-3283

Modelación de sistemas de inferencia basados en la técnica de la tabla de repertorio borroso. Un estudio experimental

Modeling of inference systems based on the fuzzy repertory table technique. An experimental study

Víctor Hugo Menéndez Domínguez [1], Miguel Esteban Romero Vázquez [2], María Enriqueta CastellanosBolaños [3], Jared David Tadeo Guerrero Sosa [4]

[1,3,4] Facultad de Matemáticas, Universidad Autónoma de Yucatán
Anillo Periférico Norte, Tablaje Cat. 13615, Colonia Chuburná Hidalgo Inn. Mérida, Yucatán, México. CP 97119,
[1] mdoming@correo.uady.mx, [3] enriqueta.c@correo.uady.mx, [4] jared.guerrero@correo.uady.mx

[2] Departamento de Ciencias de la Computación y Tecnologías de la Información, Universidad del Bío-Bío
Avda. Andrés Bello 720, Casilla 447. Chillán, Provincia de Diguillín, Chile. 3800708
[2] miguel.romero@ubiobio.cl

PALABRAS CLAVE:

Reglas de inferencia, lógica borrosa, representación de conocimiento

RESUMEN

Los Sistemas de Inferencia Borrosa permiten modelar procesos complejos donde su característica principal es la incertidumbre o la imprecisión de los valores. Este tipo de sistemas emplea colecciones de reglas "Si-Entonces" que utilizan etiquetas lingüísticas para representar conceptos que no pueden tener un análisis cuantitativo preciso. En este artículo se presenta una herramienta para el desarrollo de Sistemas de Inferencia Borrosa empleando la Tabla de Repertorio Borroso, una técnica originada en el área de la Psicología para la representación del conocimiento que incorpora aspectos de la Lógica Borrosa. Dado un conjunto de ejemplos, un conjunto de dominios borrosos y sus etiquetas lingüísticas, la herramienta genera un modelo de clasificación borrosa multinivel, es decir reglas borrosas. Estos Sistemas de Inferencia Borrosa son del tipo MISO (multiple-in, simple-out), es decir, conjuntos de reglas borrosas con varias variables de entrada y una variable de salida. La herramienta permite al usuario desarrollar, evaluar y utilizar las reglas borrosas que modelan un proceso. Se presenta un caso de estudio que valida la técnica y la herramienta desarrollada. Los resultados obtenidos permiten corroborar la efectividad de la herramienta para la modelación de sistemas utilizando la técnica de la Tabla de Repertorio Borroso.

KEYWORDS:

Inference rules, fuzzy logic, knowledge representation

ABSTRACT

Fuzzy Inference Systems (FIS) allow modeling complex processes where their main characteristic is the uncertainty or imprecision of the values. This type of system employs collections of "If-Then" rules that use linguistic labels to represent concepts that cannot have a precise quantitative analysis. This article presents a tool for the development of Fuzzy Inference Systems using the Fuzzy Repertory Table, a technique originated in the area of Psychology for the representation of knowledge that incorporates aspects of Fuzzy Logic. Given a set of examples, a set of fuzzy domains and their linguistic labels, the tool generates a multilevel fuzzy classification model, that is, fuzzy rules. These Fuzzy Inference Systems are of the MISO type (multiple-in, simple-out), that is, sets of fuzzy rules with several input variables and one output variable. The tool allows the user to develop, evaluate and use the fuzzy rules that model a process. A case study is presented that validates the technique and the developed tool. The results obtained allow us to corroborate the effectiveness of the system modeling tool using the Fuzzy Repertory Table technique.

Recibido: 22 de agosto del 2020• Aceptado: 05 de septiembre de 2020 • Publicado en línea: 30 de octubre de 2020

INTRODUCCIÓN

La teoría de conjuntos borrosos es una extensión de la teoría clásica de conjuntos [1]. En la teoría clásica, un elemento pertenece a un conjunto o se encuentra completamente fuera de él. Los conjuntos borrosos permiten una pertenencia parcial. La teoría de conjuntos borrosos [2] se puede utilizar para representar expresiones lingüísticas que permiten captar la vaguedad e imprecisión lingüística, característica inherente a la inteligencia humana. Gracias a esta flexibilidad, estos conjuntos borrosos han sido empleados para desarrollar sistemas de inferencia que pueden modelar procesos complejos o sistemas incompletos o inciertos [3].
Este tipo de inferencia, denominada borrosa [4] emplea reglas borrosas "Si-Entonces" que pueden modelar los aspectos cualitativos del conocimiento humano, así como procesos de razonamiento sin emplear análisis cuantitativos precisos. Esta "modelación borrosa", tiene un impacto práctico en el control, la predicción y la deducción [5-7].
Un Sistema de Inferencia Borrosa está conformado por un conjunto de reglas lingüísticas de descripción que están basadas en un conocimiento experto [8]. Este conocimiento experto está representado en la forma de:

SI – un conjunto de condiciones antecedentes son satisfechas
ENTONCES – un conjunto de consecuencias pueden ser inferidas

Debido a que los antecedentes y consecuentes de estas reglas "Si-Entonces" están asociados a conceptos borrosos (representados como etiquetas lingüísticas), son llamados también "sentencias condicionales borrosas" [9]. Para un sistema MISO (Multiple Input-Single Output), las reglas tienen la forma:

R_1: SI (x) ES A_1 Y (y) ES B_1 ENTONCES (z) ES C_1,
R_2: SI (x) ES A_2 Y (y) ES B_2 ENTONCES (z) ES C_2,
… … … …
R_n: SI (x) ES A_n Y (y) ES B_n ENTONCES (z) ES C_n.

Donde x, y son variables lingüísticas que representan las variables de entrada sobre el estado de un proceso y z es

la variable de salida. A_i y B_i son los valores lingüísticos de las variables lingüísticas x, y en el universo de dominio U y V, respectivamente con i=1,2,…,n. Ci son los valores lingüísticos de la variable lingüística z en el universo de dominio W para el caso de un Sistema Mandami [10] y son funciones lineales de las entradas para el caso de Sistemas Takagi-Sugeno (TSK) [6].
La estructura básica de un SIB consiste en cinco elementos principales [3] (Figura 1):

1. Una serie de reglas base que contienen todas las reglas borrosas "Si-Entonces".
2. Una base de datos que establezca las funciones de pertenencia del conjunto borroso usado en las reglas borrosas.
3. Una unidad de toma de decisiones que realiza las operaciones de inferencia con base en las reglas.
4. Una interfaz de "borrosidad" que transforma las entradas no-borrosas (valores crisp) en grados de coincidencia con valores lingüísticos.
5. Una Interfaz de "desborrosidad" que transforma los resultados borrosos de la inferencia en resultados no borrosos.

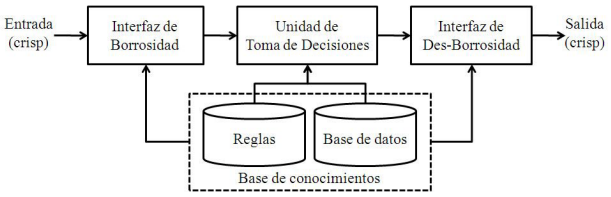

Figura 1. Componentes de un sistema de inferencia borrosa.

Una técnica empleada para identificar las características significativas de un conjunto de ejemplos, es la definida como Tabla de Repertorio Borroso (FRT, Fuzzy Repertory Table) [11].

TABLA DE REPERTORIO BORROSO

La Tabla de Repertorio Borroso está basada en la técnica de Rejilla de Repertorio (Repertory Grid) [12, 13], que es utilizada para la adquisición de conocimiento, principalmente en el área de la Psicología [14]. La FRT es una matriz rectangular con elementos (columnas) y constructos (filas). Cada intersección fila-columna contiene una razón, la cual consiste de un conjunto de funciones trapezoidales o triangulares que muestran cómo un usuario aplica un constructo dado a un elemento en particular [11].
Las funciones trapezoidales o triangulares extienden la FRT para formar restricciones, las cuales pueden

ser representadas en una rejilla permitiendo expresar un conocimiento mediante un conjunto de atributos y medidas. Las medidas implican asignar números o símbolos para medirlos. Estos números serán siempre tomados entre 0 y 1. Se usa la siguiente escala de valores [11]:

- Escala Ordinal. Un número exacto de posiciones para los elementos en una serie establecida por una escala (por ejemplo, grado de dificultad de un examen). Los valores usados para este tipo de escala son valores booleanos o valores ordenados.
- Escala Continua o Real. Un número representando una cantidad particular que define al elemento (por ejemplo, edad). Los valores usados en este tipo de escala son:
- Valores Borrosos. En este caso los usuarios asignan un término lingüístico a un elemento, el cual debe estar claramente definido dentro de un intervalo.
- Valores Crisp. Los usuarios asignan un número x, a un elemento. La función asociada con este valor es 1, si por ejemplo ciertos valores son verdaderos.
- Valores Crisp a Intervalos. El usuario asigna dos valores, x, y, a un elemento de tal forma que el intervalo entre estos dos valores tenga significado.
- Escala Nominal. Se refieren a "etiquetas" que definen los elementos (por ejemplo, sexo o código de una asignatura).

El objetivo de la FRT es construir una serie de reglas iniciales (a partir de los ejemplos de entrenamiento) y generar un conjunto de reglas definitivas o maximales que son generalizaciones de las reglas iniciales [11].

Se comienza a partir del conjunto inicial de ejemplos y de un conjunto vacío de reglas definitivas. El primer paso es convertir cada uno de los ejemplos en una regla. Para ello, el dominio de las variables de entrada se divide en regiones, a las cuales se asigna una función característica y una etiqueta lingüística. De manera que cada ejemplo establece una regla correspondiente, tomando la etiqueta que mejor se adapte para cada valor concreto de las variables de entrada.

Una vez obtenido el conjunto inicial de reglas, se construye el conjunto definitivo de reglas maximales. En este paso se genera el conjunto de reglas maximales tomando cada una de las reglas del conjunto inicial obtenido en el paso anterior y se comprueba si subsume a alguna regla del conjunto de reglas definitivas.

Se dice que una regla subsume a una segunda, si la primera regla obtiene la misma clasificación que la segunda, y además, sus conjuntos de etiquetas son subconjuntos de las etiquetas de la segunda. Si la regla subsume a otra regla definitiva se ignora, puesto que ya ha sido generalizada. Si, por el contrario, la regla no subsume a ninguna otra, sobre ella se realiza un proceso de amplificación, que consiste en tomar cada una de las variables y amplificarla, es decir, añadirle una nueva etiqueta lingüística que antes no se había considerado para ella.

La amplificación de una regla es posible si no existe alguna regla en el conjunto de reglas iniciales tal que sus etiquetas lingüísticas sean un subconjunto de las etiquetas de la nueva regla obtenida en la amplificación y cuya clase de salida sea distinta.

Tras probar todas las amplificaciones posibles en cada una de las variables de entrada para todas las etiquetas e introducir las amplificaciones posibles, obtenemos una nueva regla máxima que se incorporará al conjunto de reglas borrosas.

El algoritmo a detalle, es el siguiente [11]:

- Paso 1: Mientras haya ejemplos en el conjunto de ejemplos, tomar uno y convertirlo en una regla.
- Paso 2: Tomar una regla del conjunto de reglas iniciales (obtenido en el paso 1).
- Paso 3: Comprobar si la regla subsume a otra del conjunto de reglas definitivas. Si subsume ir al paso 2.
- Paso 4: Para cada variable en la regla escogida:
- Paso 4.1: Para cada etiqueta que no se haya considerado:
- Paso 4.1.1: Comprobar si es posible amplificar la regla utilizando esa etiqueta. Si no es posible ir al paso 1.
- Paso 4.1.2: Amplificar la regla.
- Paso 5: Si hay reglas sin amplificar en el conjunto de reglas iniciales, volver al paso 2. Si no, finalizar.

El conjunto de reglas maximales generado es validado con un conjunto de ejemplos de validación con el propósito de determinar la efectividad del sistema. Ante alguna excepción en la validación, puede tomarse la decisión de eliminar la regla maximal o realizar una amplificación de la regla considerando el ejemplo que generó la excepción. La Figura 2 presenta un resumen esquemático del proceso completo.

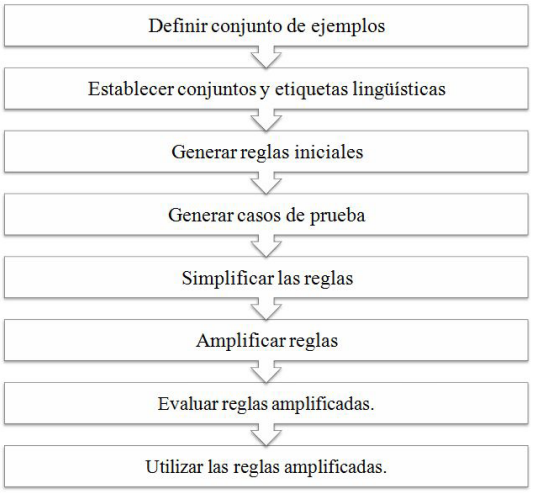

Figura 2. Proceso para la generación de una tabla de repertorio borroso.

IMPLEMENTACIÓN DE UNA HERRAMIENTA PARA LA INFERENCIA BORROSA

Con el propósito de asistir al usuario en el proceso de construcción, evaluación y utilización de reglas borrosas empleando el procedimiento indicado en la sección anterior, se ha desarrollado una aplicación de aprendizaje aproximado que automatiza dicho proceso. Dado un conjunto de ejemplos, un conjunto de dominios borrosos y sus etiquetas lingüísticas se genera un modelo de clasificación borrosa multinivel. Este Sistema de Inferencia Borrosa es del tipo MISO.

Cada variable de entrada es una variable borrosa cuyos valores se toman de una colección de etiquetas lingüísticas. La variable de salida será otra variable borrosa, a la cual se asigna una etiqueta lingüística que representará la clase de salida de esa regla. Las funciones características que representan a los conjuntos borrosos pueden ser funciones trapezoidales o triangulares.

Al final, dado un patrón de datos de entrada se decide a que clase pertenece un caso; es decir, para una serie de valores concretos de las variables de entrada se asigna una clase a la variable de salida.

Aprovechando la arquitectura y funcionalidades que ofrece la Hoja de Cálculo Microsoft Excel, se ha desarrollado un conjunto de macros en el lenguaje de programación Visual Basic for Applications (VBA). Estas macros definen una API para la generación de Sistemas de Inferencia Borrosa, la cual se ha empleado para el desarrollo de la aplicación. La herramienta se distribuye como un archivo XLS (fuzzy.xls) que contiene una serie de hojas de cálculo, en las cuales se colocan los datos necesarios para realizar algún paso.

Cada uno de los pasos del algoritmo está asociado con una hoja del archivo (Tabla 1), que incluyen botones y accesos directos que realizan los procesos. Cada hoja almacena o bien los datos de entrada o la salida de la ejecución de uno de los procesos. Todo el proceso puede realizarse de forma automática, simplificando aún más el desarrollo de un sistema de inferencia.

Hoja	Descripción
Datos	Contiene la lista de ejemplos que serán utilizados para la generación de reglas borrosas.
Parámetros	Registra la lista de etiquetas lingüísticas y el conjunto borroso para cada variable, así como la definición de la función de pertenencia.
Tabla borrosa	Contiene la lista de ejemplos ya convertidos en reglas borrosas. Utiliza los datos almacenados en las hojas “Datos” y “Parámetros”.
Casos de evaluación	Almacena los casos de ejemplo que serán utilizados para probar las reglas amplificadas. Se toma el 20% de las reglas seleccionadas de la hoja “Tabla borrosa”.
Reglas simplificadas	Almacena las reglas generadas en la hoja “Tabla borrosa” luego que han sido agrupadas. Incluye una columna que lista cuantos casos son representados por cada regla.
Reglas amplificadas	Contiene las reglas que conforman la base de conocimientos del sistema borroso. Se generan con base en las reglas almacenadas en la hoja “Reglas simplificadas”.
Evaluación	Esta hoja pretende ser utilizada para evaluar y utilizar las reglas amplificadas.

Tabla 1. Listado de hojas que conforman la aplicación.

La Figura 3 presenta el flujo de información entre las distintas hojas.

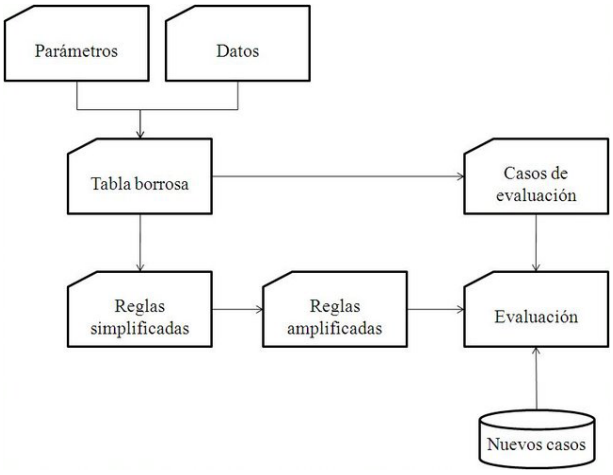

Figura 3. Diagrama de flujo.

ESTUDIO EXPERIMENTAL

Las orejas de mar, abalones o abulones (Haliotis spp.) son un género de gasterópodos, el único en la familia monotípica Haliotidae, que comprende más de un centenar de especies ampliamente distribuidas en casi todo el mundo. Su carne es un plato muy apreciado en Chile, el sudeste de Asia (China y Japón) y últimamente en algunas zonas de Estados Unidos, lo que ha llevado a problemas de conservación.

Para evitar la sobreexplotación existen intentos de crear granjas donde se cultive el abulón. La selección de los mejores individuos para el consumo está basado en su edad, pues la calidad de su carne está muy relacionada con esta.

La edad del abulón es determinada cortando la concha desde el cono superior y contando el número de anillos utilizando un microscopio, una tarea costosa que requiere de personal calificado. Sin embargo, existen otras medidas que son más fáciles de obtener como peso y longitud y que se podrían usar para determinar la edad de un abulón.

Hipótesis y objetivos

Un Sistema de Inferencia Borrosa puede determinar la edad de los abulones a partir de sus características físicas con un grado de efectividad aceptable.

El objetivo es desarrollar un sistema basado en lógica borrosa para determinar el número de anillos de un abulón. Específicamente se pretende:

- Identificar un conjunto de variables que puede ser utilizado para identificar a la edad del abulón.
- Establecer un conjunto de variables lingüísticas y sus conjuntos borrosos.
- Identificar un conjunto de reglas de inferencia.
- Diseñar un Sistema de Inferencia Borrosa para definir la edad del abulón.

Origen de los datos

Los ejemplos utilizados para generar el SIB están constituidospordosbasesdedatosdisponiblesenInternet (ftp://ftp.ics.uci.edu/pub/machine-learning-databases/abalone/):

1. Marine Resources Division, Marine Research Laboratories – Taroona, Department of Primary Industry and Fisheries, Tasmania, GPO Box 619F, Hobart, Tasmania 7001, Australia.

2. Sam Waugh (Sam.Waugh@cs.utas.edu.au), Department of Computer Science, University of Tasmania, GPO Box 252C, Hobart, Tasmania 7001, Australia.

El conjunto de ejemplos está formado por 4.177 registros. Cada uno de ellos corresponde a un abulón y se describe mediante nueve atributos donde el último, Anillos, corresponde al valor a predecir y los restantes a las variables del sistema. La Tabla 2 describe los atributos de cada individuo.

Nombre	Tipo	Medida	Descripción
Sexo			M=masculino, F=Femenino, e I=infante
Largo	continuo	mm.	La medida más larga de la concha
Diámetro	continuo	mm.	Perpendicular al largo
Altura	continuo	mm.	Altura con la carne en la concha
P_completo	continuo	gramos	Peso del abulón entero
P_neto	continuo	gramos	Peso de la carne
P_viscera	continuo	gramos	Peso de la víscera despúes de ser desangrado
P_concha	continuo	gramos	Peso de la concha después de ser secada
Anillos	entero		Si se le suma 1,5 se obtiene la edad en años. Este es el valor a predecir.

Tabla 2. Atributos de la tabla de ejemplos.

Experimentación

Para cada variable de entrada del sistema a generar se han definido etiquetas lingüísticas (Tabla 3), con la excepción de la variable Sex pues ya estaban definidas en los datos originales.

Extra pequeño	Pequeño	Mediano	Grande	Extra grande
XS	S	M	L	XL

Tabla 3. Etiquetas lingüísticas.

Los conjuntos borrosos que describen cada una de las etiquetas corresponden a funciones trapezoidales para XS y XL y triangulares para el resto, formando un dominio borroso representado en la Figura 4.

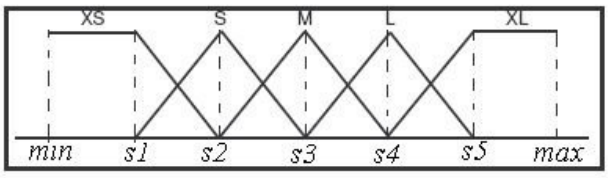

Figura 4. Dominio borroso para las etiquetas lingüísticas definidas.

Para la definición de los dominios borrosos se consideraron los valores mínimos y máximos de cada variable del sistema. A partir de estos conjuntos se determinó el valor para los cortes s1, s2, s3, s4 y s5, de tal manera que los intervalos entre los cortes fueran equidistantes. Dicha distancia d se calcula de la siguiente manera:

$$d=((Max-Min))/(cortes+1) \quad (1)$$

Donde cortes tiene valor de 5 y los valores mínimos (Min) y máximos (Max) de cada variable se presentan en la Tabla 4.

	Largo	Diámetro	Altura	P_completo	P_neto	P_viscera	P_concha	Anillos
Min	0.0750	0.0550	0.0000	0.0020	0.0010	0.0010	0.0020	1
Max	0.8150	0.6500	1.1300	2.8260	1.4880	0.7600	1.0050	29

Tabla 4. Valores mínimos y máximos para cada variable.

Cada variable tiene su propio dominio borroso pues los valores máximos y mínimos son distintos. La Figura 5 presenta un ejemplo del dominio borroso de la variable Largo donde se muestran los valores de los cortes que crean los intervalos borrosos.

Variable: Largo					
	XS	S	M	L	XL
a	0.075	0.198	0.322	0.445	0.568
b	0.075	0.322	0.445	0.568	0.692
c	0.198	0.322	0.445	0.568	0.815
d	0.322	0.445	0.568	0.692	0.815

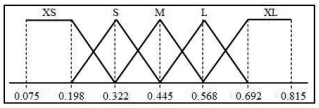

Figura 5. Dominio borroso de la variable Largo.

Para determinar la etiqueta lingüística de cada ejemplo, se evalúa el grado de pertenencia del valor a cada conjunto borroso para determinar a qué conjunto tiene mayor pertenencia, devolviendo la etiqueta asociada. La Tabla 5 presenta el resultado de este proceso de transformación para dos ejemplos. La columna Anillos se mantuvo con los valores originales porque son estos valores los que se trataban de predecir. Su finalidad es poder comparar la salida del sistema con la salida real.

	Sexo	Largo	Diámetro	Altura	P_total	P_neto	P_viscera	P_concha	Anillos
Original	F	0.5300	0.4200	0.1350	0.6770	0.2565	0.1415	0.2100	9
	M	0.4400	0.3650	0.1250	0.5160	0.2155	0.1140	0.1550	10
Borroso	F	L	L	XS	XS	XS	XS	XS	9
	M	M	M	XS	XS	XS	XS	XS	10

Tabla 5. Ejemplo de dos registros originales y su versión borrosa.

Como resultado de la ejecución de la herramienta presentada en la sección anterior se generó una tabla de reglas que constituyen al Sistema de Inferencia Borrosa. La Tabla 6 presenta algunas de estas reglas.

Sexo	Largo	Diámetro	Altura	P_total	P_neto	P_viscera	P_concha	Anillos
I	S	S	XS	XS	XS	XS	XS	6
F, I	XL	XL, XS, S	XS, S, M, L, XL	L	M, XS	L, XL	M	11
M	L	L	XS	M	S	S	S	12
F, I	XL, XS	XL, XS	XS, S	M, XS	S, XS	M, XL	L	12
M	S, M, L	S, M	M, L, XL	S, XL	M			

Tabla 6. Algunas reglas obtenidas.

A partir de las reglas obtenidas, se realizó una evaluación con los casos que no participaron en el proceso de generación de reglas. Luego se comparó el resultado esperado con el resultado del sistema con el objetivo de determinar el grado de eficiencia del sistema y si este cometía o no errores al entregar una salida. La estadística de resultados obtenidos se muestra en la Tabla 7.

Medida	Valor	Descripción
Casos evaluados	1.504	Total de casos que fueron evaluados.
Casos acertados	1.318	Total de casos en donde la salida del sistema corresponde con la salida esperada.
Casos no determinados	186	Total de casos para los cuales el sistema no pudo determinar la salida retornando como valor 0.
Total de errores	0	Total de casos en donde la salida del sistema es distinta de cero y cuyo valor esperado es diferente.
Eficacia del sistema	87.6%	Porcentaje de casos acertados con respecto al total de casos evaluados.

Tabla 7. Estadística obtenida como resultado de la experimentación.

Del análisis de los resultados podemos resaltar la utilidad de emplear un Sistema de Inferencia Borrosa para simplificar el proceso de determinación de la edad de un grupo de abulones. Empleando la técnica implementada se ha obtenido una precisión del 87.6% en la valoración de los casos de prueba.

Por otra parte, la herramienta desarrollada ha resultado un valioso auxiliar en el proceso de experimentación, facilitando en gran medida la generación y aplicación del sistema de inferencia.

DISCUSIÓN

Para analizar los resultados presentados correctamente es importante tener algún punto de comparación o referencia. En este sentido, durante el planteamiento del estado del arte encontramos varios trabajos que han abordado el problema de predecir la edad de los

abulones a partir de sus características, realizando experimentos similares y utilizando el mismo dataset.
En particular el trabajo de Mehta [15] hace una revisión completa de la literatura la cuál se ha resumido en la siguiente tabla:

Autor	Técnica	Basada en	Eficiencia del Sistema (accuracy)
Saina, Purmania [16]	SMOTE, TOMEK y SVM	SVM	99.26%
Mayukh [17]	K-means	Algoritmo de clustering	61,78 %
Wang [18]	CasPer	CasCor	30,78%
Alsabti, Ranka, Singh [19]	CLOUD	Árboles de decisión	26,4 %
Alsabti, Ranka, Singh [19]	SSE	Árboles de decisión	26,3 %
Wang [18]	Red neuronal de tres capas	Redes neuronales	25,99%
Fahlman, S, Lebiere [20]	Cascade Correltaion (CasCor)	Redes neuronales sin backpropagation	24,43%
Wang [18]	C4.5	Árboles de decisión	21,5 %
Mirza [21]	CGAN	Redes neuronales	16,29 %

Tabla 8. Resumen de métodos utilizados en el estado del arte para predecir la edad de los abulones basados en sus atributos.

Es importante destacar que la técnica que ocupa el primer lugar, manipula los datos originales utilizando dos técnicas: SMOTE que permite aumentar la cantidad de muestras de manera sintética compensando así el desequilibrio que ocurre en las clases que son minoría (las clases son las edades de salida); y TOMEK que elimina aquellas muestras que están clasificadas en dos clases diferentes pero son vecinos más cercanos entre sí. Con ambos pasos se facilita la clasificación. Luego con el nuevo conjunto de datos de entrada generado del proceso anterior se aplica el clasificador SVM.

En este sentido la comparación no es justa por que las demás técnicas trabajan con los datos originales sin ninguna alteración.
Al comparar el método FRT con las técnicas que no manipulan los datos de entrada, FRT supera a todas las presentadas en la tabla 8. Pero queda por debajo de SMOTE, TOMEK y SVM.
Por lo anterior, un trabajo interesante sería evaluar el desempeño de los distintos algoritmos del estado del arte utilizando el dataset derivado de aplicar las técnicas de SMOTE y TOMEK.

CONCLUSIONES Y TRABAJO FUTURO

En el desarrollo de este trabajo se han tomado ideas y técnicas de diversas ramas de las ciencias de la computación, tales como la Inteligencia Artificial, Ingeniería del Software, etc. Se ha obtenido un producto que incorpora una serie de características base que puede ser fácilmente mantenible y escalable.

La base de conocimiento de los modelos borrosos viene dada en la forma de una serie de reglas en las que los conjuntos borrosos que forman a las variables están determinados mediante funciones características trapezoidales o triangulares. El resultado es una hoja de cálculo que acepta una serie de valores para las variables de entrada, deduciendo una clase de salida para el caso determinado por esos valores de entrada. El producto desarrollado brinda una herramienta simple y útil para los Ingenieros del Conocimiento en su labor de construir Sistemas de Inferencia Borrosa. La aplicación desarrollada satisface los requerimientos iniciales y es la base para incorporar nuevas funcionalidades. Al utilizar VBA esta actividad resulta sencilla. En este sentido, es posible mejorar la aplicación en muchos sentidos:

- Desarrollar una mejor interfaz de usuario para la captura de datos.
- Implementar el algoritmo de amplificación de reglas iniciales como un componente NET que pueda ser compartido con otras aplicaciones, incluso en red.
- Desarrollar una variante del algoritmo utilizando técnicas de paralelismo.
- Establecer un esquema de grafication de distribución de casos para identificar de forma sencilla las excepciones.

Por otra parte es importante analizar qué mejoras pueden aplicarse al algoritmo con el propósito de incrementar su eficacia, que para este caso fue del 87.6%. En este sentido, algunas técnicas de minería de datos como algoritmos de agrupamiento, clasificación y asociación pudieran resultar útiles para flexibilizar los intervalos difusos y con ello generar patrones de comportamiento que representen mejor los distintos casos modelados, tomando en cuenta que en la técnica de la tabla de repertorio

borroso son requeridos un conjunto de datos para entrenamiento.

Otra línea futura de investigación y desarrollo es migrar la herramienta a la nube, de tal forma que sea posible, mediante una colección de servicios Web y componentes que incorporen el algoritmo presentado, realizar el proceso de analizar datos y generar reglas con el propósito de tener un simulador para el modelo de datos generado, lo que permitiría explorar otras temáticas relacionadas, como es el diseño de interfaces dinámicas, optimización de servicios, almacenamiento en la nube, desarrollo de aplicaciones Web o apps para dispositivos móviles, etcétera.

REFERENCIAS

1. Dubois, D. J., Prade, H. M. Fuzzy Sets and Systems: Theory and Applications. New York: Academic Press, 1994.
2. Babuška, R., Verbruggen, H. B. Fuzzy Set Methods for Local Modelling and Identification. In: Murray-Smith, R., Johansen T.A. (Ed.). Multiple Model Approaches to Modelling and Control. Londres: Taylor & Francis, 1997, 75-100.
3. Hwang, G. J. Knowledge acquisition for fuzzy expert systems. International Journal of Intelligent Systems. 1995, 10(6), 541- 560.
4. Delgado, M., González, A. An inductive learning procedure to identify fuzzy systems. Fuzzy Sets and Systems. 1993, 55(2), 121-132.
5. Castro-Schez, J. J., Jennings, N. R., Luo, X., Shadbolt, N. R. Acquiring domain knowledge for negotiating agents: a case of study. International Journal of Human-Computer Studies. 2004, 61(1), 3-31.
6. Takagi, T. Sugeno, M. Fuzzy Identification of Systems and Its Applications to Modeling and Control. IEEE Transactions on Systems, Man and Cybernetics. 1985, 15(1), 116-132.
7. Vinothini, S., Chandra Segar Thirumalai, Vijayaragavan, R. An efficient AFRA using an Intelligent Fuzzy Repertory Table technique. International Research Journal of Engineering and Technology. 2015, 2(2), 365-372.
8. Castro, J. L., Castro-Schez, J. J., Zurita, J. M. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. Fuzzy Sets and Systems. 1999, 101(3), 331-342.
9. Castro, J. L., Zurita, J. M. An inductive learning algorithm in fuzzy systems. Fuzzy Sets and Systems. 1997, 89(2), 193-203.
10. Mamdani, E. H. Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of the Institution of Electrical Engineers on Control & Science. 1974, 121(12), 1585-1588.
11. Castro-Schez, J. J., Castro, J. L., Zurita, J. M. Fuzzy Repertory Table: A Method for Acquiring Knowledge About Input Variables to Machine Learning Algorithm. IEEE Transactions on Fuzzy Systems. 2004, 12(1), 123-139.
12. Edwards, H. M., McDonald, S., Young, S. M. The repertory grid technique: Its place in empirical software engineering research. Information and Software Technology. 2009, 51(4), 785-798.
13. Bradshaw, J. M., Ford, K. F., Adams-Webber, J. R., Boose, J. H. Beyond the Repertory Grid: New Approaches to Constructivist Knowledge Acquisition Tool Development. International Journal of Intelligent Systems. 1993, 8, 287–333.
14. Easterby-Smith, M., Thorpe, R., Holman, D. Using repertory grids in management. Journal of European Industrial Training. 1996, 20(3), 3-30.
15. Mehta, K. Abalone Age Prediction Problem: A Review. International Journal of Computer Applications: 2019, 178(50),43-49. Doi: 10.5120/ijca2019919425.
16. Sain, H, Purnami, S.W. Combine Sampling Support Vector Machine for Imbalanced Data Classification. Procedia Computer Science: 2015, 72, 59–66.
17. Mayukh, H. Age of Abalones using Physical Characteristics:A Classification Problem. Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 2010.
18. Wang, Z Abalone Age Prediction Employing A Cascade Network Algorithm and Conditional Generative Adversarial Networks. Research School of Computer Science, Australian National University, 2018.
19. Ranka, S., Singh, V. CLOUDS: A decision tree classifier for large datasets. Proceedings of the 4th Knowledge Discovery and Data Mining Conference: 1988, 2(8).
20. Fahlman, S. E., Lebiere, C. The cascade-correlation learning architecture. Advances in neural information processing systems: 1990, p. 524-532.
21. Mirza, M., Osindero S. Conditional Generative Adversarial Nets. Department of Computer Science and Operations Research, University of Montreal, Canada: 2014. arXiv preprint arXiv:1411.1784

Acerca de los autores

Víctor Hugo Menéndez Domínguez es Licenciado en Ciencias de la Computación (1994) y Especialista en Docencia (2002) por la Universidad Autónoma de Yucatán, México. Tiene un Máster en Tecnologías Informáticas Avanzadas (2008) y es Doctor en Tecnologías Informáticas Avanzadas en 2012 por la Universidad de Castilla-La Mancha, España. Desde el año 2000 es Profesor Titular en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con la gestión y representación del conocimiento y el aprendizaje, la enseñanza virtual y la ingeniería web. Ha sido director de tesis de Licenciatura y Posgrado. Es autor de numerosas publicaciones y ponencias presentas en eventos nacionales e internacionales.

Miguel Romero es profesor asistente A en el Departamento de Ciencias de la Computación y Tecnologías de la Información, Facultad de Ciencias Empresariales de la Universidad del Bío-Bío, Chile. Desde 2007 es Magíster en Ciencias de la Computación por la Universidad de Concepción, Chile, desde 2008 es Máster en tecnologías informáticas avanzadas por la universidad de Castilla-La Mancha, España y desde 2017 es Doctor en Computación por Universidade da Coruña. Sus principales áreas de investigación son las estructuras de datos compactas y algoritmos y la indexación espacial y espacio-temporal. Ha sido director de tesis de Licenciatura y Posgrado. Es autor de varias publicaciones y ponencias presentadas en eventos nacionales e internacionales.

María Enriqueta Castellanos Bolaños es Licenciada en Ciencias de la Computación (2003) y Especialista en Docencia (2011) por la Universidad Autónoma de Yucatán, México. Tiene una Maestría en Gestión de Tecnología de Información (2007) por la Universidad Anáhuac Mayab, México. Desde el año 2006 es Profesora Titular en la Facultad de Matemáticas de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con la gestión y representación del conocimiento y el aprendizaje, la enseñanza virtual y la ingeniería web. Entre sus áreas de interés se encuentran también las métricas de software y los métodos formales. Es autora de publicaciones y ponencias presentas en eventos nacionales e internacionales.

Jared David Tadeo Guerrero Sosa es Ingeniero en Tecnologías de la Información y Comunicaciones (2017) por el Instituto Tecnológico de Chetumal, México. Es Maestro en Ciencias de la Computación (2019) por la Universidad Autónoma de Yucatán, México. Actualmente trabaja en el área de Investigación y Desarrollo de Software de la Universidad Autónoma de Yucatán, México. Su trabajo de investigación se centra en temas relacionados con los repositorios digitales de acceso abierto y su interoperabilidad, la gestión y representación del conocimiento, la Informática Educativa y la evaluación de la producción científica. Ha publicado sus trabajos de investigación en diversas revistas y en congresos internacionales.