

CATEGORIZACIÓN DE INDIVIDUOS EN SESIONES VIRTUALES Y PRESENCIALES CON SISTEMAS DE INFERENCIA DIFUSA Y LINGÜÍSTICA

Ángel López-Gómez* y Enrique Ramos-López
Facultad de Estudios Superiores Cuautitlán, UNAM

[*angell@unam.mx](mailto:angell@unam.mx)

Resumen

Los sistemas de inferencia difusa (FIS por sus siglas en inglés, *Fuzzy Inference System*) son reconocidos modelos lógicos basados en la lógica difusa de Zadeh que han mostrado su utilidad cuando se requiere combinar información tanto cuantitativa como cualitativa en condiciones de incertidumbre. Son sistemas basados en reglas con un esquema de inferencia para la generación de resultados. Se presenta un caso de estudio que genera agrupamientos categorizados a ser identificados para sesiones presenciales y virtuales haciendo uso de variables lingüísticas y software libre, como el lenguaje R y los paquetes FuzzyR y ClusterR.

Palabras clave: FuzzyR, lenguaje R, variable lingüística.

Introducción

La inferencia clásica o booleana se define como el proceso de razonamiento que genera una conclusión a partir de una o varias premisas.

En relación con el desarrollo de los sistemas de inferencia computacionales se considera que el primer trabajo presentado es el de Ehud Y. Shapiro (Shapiro, 1991) para realizar un proceso algorítmico de inferencia a partir de ejemplos o hechos.

La parte algorítmica resulta relevante por ser un método por medio del cual es posible resolver un problema utilizando una secuencia de reglas definidas, o instrucciones colocadas de manera lógica y ordenada.

Los modelos de sistemas de inferencia (MIS por sus siglas en inglés, *Model Inference System*) están identificados dentro de lo que conocemos como aprendizaje de máquina, y también como metodología para representación de conocimiento.

En (Lisi y Straccia, 2013) se analizan los conceptos de vaguedad e incompletud que se presentan como propiedades inherentes del conocimiento bajo el dominio de muchos casos reales. Principalmente impregnados de estas características en aquellos ambientes donde sus elementos participantes pueden ser mejor descritos aplicando lenguaje natural.

En (Zadeh, 1965) se presenta una nueva aproximación para el análisis de sistemas complejos haciendo uso de su cambio metodológico y paradigma computacional. Su propuesta incorpora el uso de las variables lingüísticas en lugar de solamente variables numéricas. Una variable lingüística se refiere en relación con sus valores como sentencias en términos de lenguaje natural o artificial.

Asimismo, se introducen las declaraciones condicionales difusas que son expresiones de la forma *If A then B*, en principio parecerían idénticas a su forma clásica pero su distinción se observa en que tanto A como B tienen un significado difuso. Por ejemplo, encontramos sentencias como: *If x is nada caluroso then y is agradable*.

Este par de elementos son acompañados por un algoritmo difuso que se conserva como una secuencia ordenada de instrucciones con la diferencia que pueden contener asignaciones difusas, así como condicionales también difusas. Por ejemplo, $x = \text{muy caluroso}$; *if x is poco caluroso then y is fresco*. En este tipo de

expresiones o reglas su ejecución es controlada por la regla de inferencia composicional, que es la manera en que es extendida al ámbito difuso en contraste con su forma clásica.

En (Guillaume y Charnomordic, 2012) se presenta una revisión sobre el desarrollo de los denominados sistemas de inferencia difusos (FIS por sus siglas en inglés, *Fuzzy Inference System*).

Los FIS han consolidado su efectividad práctica debido al comportamiento en una manera interpretable desde el punto de vista del conocimiento humano.

La base de los sistemas FIS es la lógica difusa introducida por Zadeh en 1965 (Guillaume y Charnomordic, 2012), caracterizándose por manejar simultáneamente datos numéricos como conocimiento expresado lingüísticamente.

La lógica difusa difiere de la lógica clásica en su conceptualización de la segunda en términos absolutos de verdadero o falso, negro o blanco, encendido o apagado y excluyentes entre sí. En cambio, en la primera una declaración puede asumir cualquier valor dentro del rango 0 y 1, lo cual representa el grado de pertenencia a un conjunto dado, que es identificado como un conjunto difuso.

Objetivo

Presentar una propuesta algorítmica y computacional, con la herramienta del lenguaje R, como caso de estudio, categorizar a un grupo de estudiantes para participar en clases virtuales o presenciales, haciendo uso de un sistema de inferencia difusa y aplicando reglas difusas tomando como base en variables lingüísticas de referencia.

Materiales y métodos

La pandemia declarada en el año 2020 trastornó todas las actividades sociales y económicas mundiales, los países en mayor o menor medida restringieron la movilidad. En México la labor docente se hizo por vía remota o virtual.

Se establecieron propuestas de una participación híbrida en las clases y otras actividades. En un porcentaje que rondaba alrededor del 30%. Sin embargo, en las indicaciones en ningún momento se definieron los criterios para conformar los dos grupos participantes.

Por la complejidad del problema y la falta de un esquema de representación matemática la alternativa de aplicar un Sistema de Inferencia Difusa basándose en la propuesta de variables lingüísticas para su representación y aplicando el paquete FuzzyR del lenguaje R, permite establecer un modelo y tener como salida un valor numérico que actúa como estimativo de una distancia o métrica que sirve para categorizar a los alumnos y aplicando una segunda fase de separación con el método de Kmeans del mismo lenguaje R.

La información recabada, mediante la aplicación de encuestas, quienes cursaban la primera asignatura teórica del área de la Química Analítica, impartida en la F.E.S. Cuautitlán durante el semestre 2022-II, se agrupó en dos aspectos generales: traslado y salud.

Se indagó el esquema de vacunación y comorbilidades del alumno y familiares; con quienes comparte el mismo recinto, y de esos familiares, el número de integrantes considerados de la tercera edad y/o menores de 18 años. Con respecto al traslado, se conformaron preguntas relacionadas con el tiempo y costo de traslado a la facultad y viceversa, número de transbordos y condición económica familiar.

El valor de pertenencia correspondiente al traslado (VPT), se obtuvo aplicando la siguiente expresión: $VPT = \{C/\{A-\{B \times 10\}\}} \times D$

Donde: A => Tiempo de traslado redondo (de ida y vuelta) en minutos.

B => Número de transbordos realizados en el viaje. C => Costo por viaje redondo.

D => Situación económica familiar. 10 => Factor de ajuste relacionado con el tiempo, que en promedio tarda una persona en realizar un transbordo.

La expresión anterior considera la movilidad del estudiante con un componente económico; es decir, con base en el tiempo que permanece en el transporte público ($A-\{B \times 10\}$), y el costo de traslado, se obtuvo el costo de traslado por minuto ($\{C/\{A-\{B \times 10\}\}$), para facilitar la comparación entre estudiantes); el cual, posteriormente fue afectado por el factor “D”, derivado del descriptor cualitativo que el alumno seleccionó.

Con respecto al valor de pertenencia para el aspecto de salud (VTS), se tomaron en cuenta el cumplimiento al esquema de vacunación del estudiante y familiares, y las comorbilidades, incumplimiento al esquema y familiares menores a 18 años, como una situación negativa ($\{-\{A-B\}-C\}+B\}+\{D-E\}$); después de considerar estos componentes, se obtuvo la fracción de integrantes familiares con mayor o menor aspectos favorables, con respecto a la siguiente expresión: $VPS = \{-\{A-B\}-C\}+B\}+\{D-E\} / \{1+A\}$

Donde: A => Número de familiares, quienes cohabitan con el alumno encuestado.

B => Número de familiares vacunados. C => Número de familiares con comorbilidades.

D => Esquema de vacunación del alumno. E => Número de comorbilidades que sufre el alumno encuestado.

Se construye una escala y los límites de la escala son el resultado del valor de pertenencia menor y mayor, aproximándolo a su número entero más cercano que lo incluya.

Cada escala se conforma de 3 conjuntos difusos. A cada conjunto se le asignó un descriptor cualitativo (variable lingüística): fácil, regular y complicado para traslado, y vulnerable, normal y favorable para los 3 intervalos para salud.

Resultados

La Tabla 1 presenta el código respectivo en R utilizando el paquete FuzzyR para diseñar y evaluar un sistema FIS y obtener como salida un valor numérico actuando como estimativo de la preferencia. La Tabla 2, ilustra una muestra de los valores numéricos utilizados como entrada (columnas 1 y 2) obtenidos de las encuestas aplicadas a los alumnos y los datos de salida como estimativos de afinidad a clase presencial (columna 3) o bien virtual (columna 4).

Tabla 1. Código de lenguaje R utilizando a FuzzyR para diseñar un FIS.

```
// uso del paquete FuzzyR para diseñar y evaluar modelo FIS
library(FuzzyR)
fis <- newfis('clase1')
fis <- addvar(fis, 'input', 'traslado', c(0, 7))
fis <- addvar(fis, 'input', 'salud', c(-2, 2))
fis <- addvar(fis, 'output', 'presencial', c(0, 1))

fis <- addmf(fis, 'input', 1, 'facil', 'trimf', c(0,0,3.5))
fis <- addmf(fis, 'input', 1, 'regular', 'trimf', c(1.75,3.5,5.25))
fis <- addmf(fis, 'input', 1, 'complicado', 'trimf', c(3.5,7,7))

fis <- addmf(fis, 'input', 2, 'vulnerable', 'trimf', c(-2,-2,0))
fis <- addmf(fis, 'input', 2, 'normal', 'trimf', c(-1,0,1))
fis <- addmf(fis, 'input', 2, 'favorable', 'trimf', c(0,2,2))

fis <- addmf(fis, 'output', 1, 'rojo', 'trimf', c(0,0,0.4))
fis <- addmf(fis, 'output', 1, 'anaranjado', 'trimf', c(0.2,0.4,0.6))
fis <- addmf(fis, 'output', 1, 'amarillo', 'trimf', c(0.4,0.6,0.8))
fis <- addmf(fis, 'output', 1, 'verde', 'trimf', c(0.6,1,1))

ruleList<- bind(c(1,1,1,1,1),c(1,2,3,1,1),c(1,3,4,1,1),c(2,1,2,1,1),c(2,2,3,1,1),c(2,3,4,1,1),c(3,1,1,1,1),c(3,2,2,1,1),c(3,3,3,1,1))
fis <- addrule(fis, ruleList)
```

Tabla 2. Muestra de datos de entrada (columnas 1 y 2) y salida (columnas 3 y 4) del FIS.

Estimación - Traslado	Estimación - Salud	Inferencia para Presencial	Inferencia para Virtual
3.571429	0.75	0.7342253	0.4800582
0.76	0.5	0.680241	0.319759
2.833333	0.17	0.6286765	0.5130243
0.2407407	1.67	0.8664771	0.1335229

Tabla 3. Código para desarrollar agrupamiento o categorización de nube de datos aplicando ClusterR.

```
# Loading data
# data(salon)
```

```
salon <-read.table("datos.txt", header=FALSE, sep= ",")

# Structure
str(salon)

# Loading package
library(ClusterR)
library(cluster)

# Removing initial label of
# Species from original dataset
# iris_1 <- iris[, -5]

# Fitting K-Means clustering Model
# to training dataset
set.seed(240) # Setting seed
kmeans.re <- kmeans(salon, centers = 2, nstart = 20)
kmeans.re

# Cluster identification for
# each observation
kmeans.re$cluster

# Confusion Matrix
# cm <- table(iris$Species, kmeans.re$cluster)
# cm

# Model Evaluation and visualization
plot(salon[c(1,2)])
plot(salon[c(1,2)],
      col = kmeans.re$cluster)
plot(salon[c(1,2)],
      col = kmeans.re$cluster,
      main = "K-means with 2 clusters")

## Plotting cluster centers
kmeans.re$centers
kmeans.re$centers[, c(1,2)]
# cex is font size, pch is symbol
points(kmeans.re$centers[, c(1,2)],
       col = 1:2, pch = 8, cex = 2)
```

```
## Visualizing clusters
y_kmeans <- kmeans.re$cluster
clusplot(salon[, c(1,2)],
  y_kmeans,
  lines = 0,
  shade = TRUE,
  color = TRUE,
  labels = 2,
  plotchar = FALSE,
  span = TRUE,
  main = paste("Cluster clase"),
  xlab = 'Presencial',
  ylab = 'Virtual')
```

La Figura 1 gráfica la salida del agrupamiento o categorización en dos bloques (no necesariamente sin intersección) resultante de la aplicación del método de Kmeans y el paquete ClusterR mediante el código de la Tabla 3, tomando como datos la salida del sistema FIS y dando una pertenencia relativa de cada alumno a una de las dos alternativas: clase presencial o clase virtual.

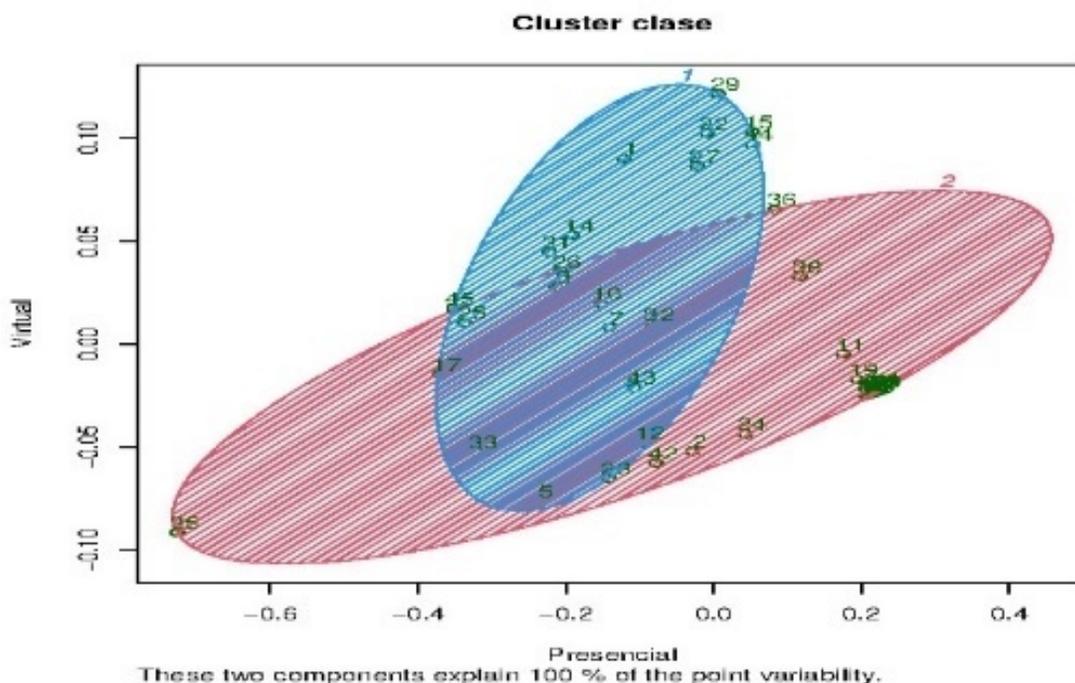


Figura 1. Salida como propuesta de categorización de clases presenciales y virtuales.

La Tabla 4 presenta la salida de la ejecución del método K-means utilizando el paquete ClusterR de R resaltando las líneas: `> kmeans.re$cluster` [1] 1 2 1 2 1 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1 1 2 1 1 1 2 1 2 1 1 2 2 2 2 [37] 2 2 2 2 2 2 1 1 1, donde se presenta la asociación del individuo evaluado por su posición de entrada y la propuesta de asignación por grado de pertenencia al grupo dominante difuso.

Tabla 4. Ejecución del código para crear agrupamientos utilizando método de Kmeans y el paquete ClusterR.

```
> # Loading data
> # data(salon)
> salon <- read.table("datos.txt", header=FALSE, sep=" ",)
>
> # Structure
> str(salon)
'data.frame': 45 obs. of 4 variables:
 $ V1: num  0.734 0.68 0.629 0.866 0.533 ...
 $ V2: num  0.48 0.32 0.513 0.134 0.467 ...
 $ V3: num  3.571 0.76 2.833 0.241 6.25 ...
 $ V4: num  0.75 0.5 0.17 1.67 0.8 1.4 0.25 1 1 0.33 ...
>
> # Loading package
> library(ClusterR)
Loading required package: gtools
> library(cluster)
Clustering vector:
[1] 1 2 1 2 1 2 2 2 2 2 2 2 1 1 2 1 2 2 2 2 1 1 2 1 1 1 2 1 2 1 1 2 2 2
[37] 2 2 2 2 2 2 1 1 1
>
> ## Visualizing clusters
> y_kmeans <- kmeans.re$cluster
> clusplot(salon[, c(1,2)],
+   y_kmeans,
+   lines = 0,
+   shade = TRUE,
+   color = TRUE,
+   labels = 2,
+   plotchar = FALSE,
+   span = TRUE,
+   main = paste("Cluster clase"),
+   xlab = 'Presencial',
+   ylab = 'Virtual')
>
```

Discusión

En el caso presentado se diseñaron con base en criterios de conjuntos y lógica difusa, las variables cuantitativas y cualitativas Traslado y Salud, en un contexto de identificación de variables significativas al contexto del problema.

Su dato salida se aplica como la distancia del sujeto a su preferencia o afinidad a tomar una clase presencial o bien una virtual. Dicho valor se utiliza como los valores a procesar con el método Kmeans para encontrar cercanías entre puntos de atracción que iterativamente producen agrupamientos en un número de clases predeterminadas.

Nuevamente los agrupamientos propuestos no necesariamente son excluyentes. En otras palabras, un alumno puede preferir o ser afín a un tipo de clase, pero eso no excluye que tenga otra valoración para el otro tipo de agrupamiento.

La separación estricta en dos agrupamientos y en el porcentaje requerido corresponde a criterios numéricos clásicos que se sobreponen a la propuesta difusa para reducir los casos que cuenten con grados de pertenencia cercanos.

Conclusión

El diseño, construcción y evaluación de un sistema de inferencia difusa aplicable a problemas complejos con ausencia de un modelo matemático estricto basado en la utilización de variables lingüísticas cualitativas y cuantitativas y desarrollado con el paquete FuzzyR del lenguaje R ha sido presentada.

La salida del sistema es introducida para su categorización en agrupamientos, no necesariamente excluyentes entre sí, aplicando el método Kmeans y el paquete ClusterR del lenguaje R, obteniendo una propuesta numérica y su gráfica para apoyar la decisión cuantitativa de clasificación clásica de los datos originales.

La metodología expuesta es aplicable, con sus adecuaciones, a problemas similares complejos y con la participación de la lógica difusa como esquema de representación y solución numérica aproximativa.

Agradecimientos

Este trabajo ha sido parcialmente soportado por el programa PAPIME clave PE107921.

Agradecemos igualmente la comunicación con la Lic. Jessica Y. Díaz S. para la consideración y elaboración de este documento.

Referencias

Džeroski, S., & Lavrač, N. (2001). An Introduction to Inductive Logic Programming. *Language Learning*, 00(c): 1–16. <http://www-ai.ijs.si/SasoDzeroski/RDMBook/TOC/rdm-03.html>

Guillaume, S., & Charnomordic, B. (2012). Fuzzy inference systems: An integrated modeling environment for collaboration between expert knowledge and data using FisPro. *Expert Systems with Applications*, 39(10). <https://doi.org/10.1016/j.eswa.2012.01.206>

Lisi, F.A., & Straccia, U. (2013). Dealing with incompleteness and vagueness in inductive logic programming. *CEUR Workshop Proceedings*, 1068.

Popper, K.R. (1963). Philosophy of Science: Conjectures and Refutations. The growth of scientific knowledge. *Science*, 140(3567). <https://doi.org/10.1126/science.140.3567.643.a>

Shapiro E.Y. (1991). Inductive Inference of Theories From Facts. *Published in Computational Logic–Essays Computational Logic-Essays in Honor of Alan Robinson. Research Report* 192. <https://cpsc.yale.edu/sites/default/files/files/tr192.pdf>

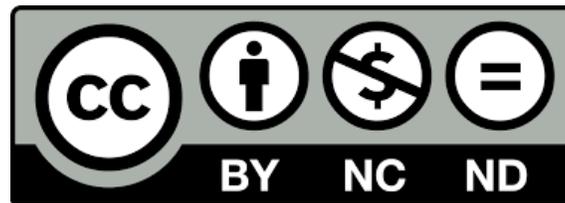
Zadeh, L.A. (1965). "Fuzzy sets", Information and Control. En *INFOR~ATIO~ AND CONTROL* Vol. 8.



D. R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO.

Excepto donde se indique lo contrario esta obra está bajo una licencia Creative Commons Atribución No comercial, No derivada, 4.0 Internacional (CC BY NC ND 4.0 INTERNACIONAL).

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



ENTIDAD EDITORA

Facultad de Estudios Superiores Cuautitlán.

Av. Universidad 3000, Universidad Nacional Autónoma de México, C.U., Delegación Coyoacán, C.P. 04510, Ciudad de México.

FORMA SUGERIDA DE CITAR:

López-Gómez, A. y Ramos-López, E. (2022). Categorización de individuos en sesiones virtuales y presenciales con sistemas de inferencia difusa y lingüística. *MEMORIAS DEL CONGRESO NACIONAL DE TECNOLOGÍA (CONATEC)*, Año 5, No. 5, septiembre 2022 - agosto 2023. Facultad de Estudios Superiores Cuautitlán. UNAM.

https://tecnicosacademicos.cuautitlan.unam.mx/CongresoTA/memorias2022/mem2022_ExtensioPaper14.html