AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

Soluciones Metodológicas para el Análisis de Datos Imprecisos: Lógica Difusa y R-Shiny

Methodological Solutions for the Analysis of Imprecise Data: Fuzzy Logic and R-Shiny

Presentación: 02/09/2024

Doctoranda:
Matilde Inés CÉSARI
Universidad Tecnológica Nacional, Facultad Regional Mendoza - Argentina
matilde.cesari@frm.utn.edu.ar

Director:
Santiago PEREZ

Resumen

El análisis de datos imprecisos es un desafío común en diversas áreas de investigación y aplicación práctica. Las observaciones imprecisas pueden surgir debido a limitaciones en las herramientas de medición y la naturaleza del fenómeno estudiado. La lógica difusa proporciona un marco flexible y realista para tratar con esta imprecisión. Este artículo explora el uso de la lógica difusa en el análisis multivariado de datos imprecisos, utilizando aplicaciones Shiny de R para mejorar la calidad y fiabilidad de los análisis. La metodología propuesta incluye la codificación y estandarización de datos, la definición de variables lingüísticas difusas y el uso de métodos multivariados para el análisis de datos borrosos

Palabras clave: Datos imprecisos, Lógica difusa , Análisis multivariado , Shiny, Entorno R, Variables lingüísticas difusas

Abstract

Imprecise data analysis is a common challenge in various research fields and practical applications. Imprecise observations can arise due to limitations in measurement tools and the inherent nature of the phenomena under study. Fuzzy logic offers a flexible and realistic framework to handle this imprecision. This paper explores the use of fuzzy logic in multivariate analysis of imprecise data, leveraging R Shiny applications to enhance the quality and reliability of the analyses. The proposed methodology includes data coding and standardization, the definition of fuzzy linguistic variables, and the use of multivariate methods for the analysis of fuzzy data.

Keywords: Imprecise data, Fuzzy logic, Multivariate analysis, Shiny, R environment, Fuzzy linguistic variables

Introducción

En un mundo impulsado por la información y los datos, la capacidad para comprender y manejar datos imprecisos es esencial para la toma de decisiones informadas en diversos campos. La imprecisión de los datos es una realidad común en muchas aplicaciones del mundo real, y su manejo adecuado es crucial para extraer conocimientos significativos. La lógica difusa es una herramienta poderosa para manejar la incertidumbre e imprecisión en datos, permitiendo una toma de decisiones más informada en diversas aplicaciones.

En el artículo de Kahraman *et al.* (2022), se exploran las extensiones recientes de la lógica difusa AHP/ANP, destacando su uso en decisiones donde la comparación directa es imprecisa, como en las evaluaciones cualitativas Kahraman *et al.,*

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

2022. Otro estudio analiza la teoría de la lógica difusa en la gestión de la incertidumbre de evaluaciones lingüísticas para estudiantes, demostrando su aplicación en la educación mediante la asignación de calificaciones lingüísticas y la toma de decisiones. Qayoom y Baig (2022) desarrollan una nueva medida de entropía difusa para tratar datos imprecisos, proporcionando una base matemática sólida para la teoría de la información difusa, y Swathi M (2023) ofrecen una visión general de la lógica difusa, incluyendo sus conceptos clave y aplicaciones, enfatizando su relevancia en inteligencia artificial y sistemas de control. Finalmente, Lian (2020) propone un nuevo sistema teórico y tecnológico para el procesamiento de información imprecisa, destacando sus ventajas sobre la tecnología difusa tradicional.

En este contexto, la lógica difusa se presenta como una herramienta poderosa para representar y manipular la incertidumbre de manera efectiva. La imprecisión en los datos puede deberse a varias razones, como valores perdidos, errores de medición, subjetividad y la complejidad del fenómeno estudiado. Estos factores pueden introducir sesgos y distorsiones en el análisis de datos, afectando la validez de los resultados y comprometiendo la calidad de las interpretaciones.

La lógica difusa se distingue de los métodos estadísticos tradicionales en que permite trabajar con datos que no se ajustan estrictamente a las categorías de verdadero o falso. A diferencia de los enfoques estadísticos convencionales, que requieren distribuciones de datos bien definidas y categorizaciones claras, la lógica difusa ofrece un enfoque más flexible y realista para el análisis de datos imprecisos. Esto se logra al permitir grados de pertenencia en lugar de categorizaciones binarias, facilitando así la modelización de problemas complejos donde la ambigüedad y la incertidumbre son inherentes.

La aplicación de la lógica difusa en el análisis de datos se explora en varios dominios, demostrando su flexibilidad y eficacia en el manejo de datos imprecisos e inciertos. Kwiatkowski *et al.* (2022) presentan un análisis multidimensional basado en lógica difusa de datos de incidentes de tráfico, destacando cómo la lógica difusa puede gestionar la incertidumbre y vaguedad inherentes en los datos de tráfico reales, mejorando así el análisis de los factores que contribuyen a los accidentes de tráfico. Bolodurina y Speshilov (2023) se centran en la gestión del transporte de carga bajo incertidumbre, proponiendo un algoritmo basado en reglas de lógica difusa para optimizar la selección de rutas y los procesos de toma de decisiones, adaptándose a las preferencias del cliente y reduciendo los riesgos. Agayan *et al.* (2023) aplican la lógica difusa al análisis de series temporales, utilizándola para identificar anomalías y características morfológicas dentro de los datos. Ulumuddin (2023) introduce un algoritmo de inferencia difusa que integra bases de datos relacionales, mejorando la eficiencia de los sistemas de inferencia difusa al minimizar el uso de memoria y el tiempo de computación. Finalmente, Djurayev y Matkurbonov (2023) proponen un modelo para mejorar la eficiencia del enrutamiento en redes de transmisión de datos utilizando lógica difusa, demostrando cómo las métricas difusas pueden mejorar las decisiones de enrutamiento adaptativo.

La accesibilidad a metodologías avanzadas, como la lógica difusa, está restringida para muchos usuarios finales debido a la necesidad de un conocimiento técnico profundo. La lógica difusa es una forma de lógica multivaluada que se utiliza para manejar la incertidumbre y la imprecisión, permitiendo una representación más flexible y matizada de la realidad. Sin embargo, implementar estas técnicas requiere habilidades técnicas avanzadas, lo que limita su adopción por parte de usuarios sin formación especializada.

La falta de herramientas intuitivas y amigables para el usuario es una barrera significativa para la implementación práctica de la lógica difusa. Esto crea un desafío para los usuarios que podrían beneficiarse de estas técnicas, pero no tienen la formación necesaria para utilizarlas eficazmente. Para mejorar la accesibilidad, se necesitan desarrollos en varias áreas. En primer lugar, es crucial crear herramientas y plataformas que simplifiquen la configuración y el uso de la lógica difusa, reduciendo la necesidad de programación compleja. Esto podría incluir interfaces gráficas de usuario (GUIs) que permitan a los usuarios construir sistemas de lógica difusa a través de interacciones visuales en lugar de código.

Este artículo propone una metodología innovadora que integra algoritmos de lógica difusa con aplicaciones Shiny de R, facilitando el análisis de datos imprecisos y promoviendo la toma de decisiones informadas.

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

El objetivo general es ampliar el alcance de la aplicación práctica y precisión de las soluciones de la lógica difusa en el tratamiento y análisis de datos imprecisos. Para lograrlo, se implementará una metodología innovadora que integra algoritmos de lógica difusa y análisis multivariado utilizando el lenguaje R y el paquete Shiny

Metodología

1. Codificación y Estandarización de Datos

La primera etapa en la metodología propuesta es la codificación y estandarización de los datos imprecisos. Los datos observados se tabulan en tablas cuantitativas y cualitativas, clasificándolos en categorías numéricas y nominales, respectivamente. Para eliminar problemas de cálculo originados por la utilización de diferentes escalas, se normalizan las puntuaciones de los atributos a un rango de 0 a 1 mediante el método de ecuación de la recta.

Código 1 *Normalización de los datos*

```
install.packages("caret")
library(caret)
datosTransf <- preProcess(datos,
                  method=c("center", "scale"))
```

2. Transformación a Datos Borrosos

El proceso de borrosificación transforma las percepciones en números difusos utilizando funciones de pertenencia. Las variables lingüísticas difusas se definen mediante descriptores lingüísticos y funciones de pertenencia como las funciones triangular, trapezoidal y gaussiana. La elección del tipo de función de pertenencia depende de la naturaleza de los datos y el objetivo del análisis.

a) Definición de Variables Lingüísticas Difusas

Para representar adecuadamente las variables difusas, se seleccionan descriptores lingüísticos apropiados y se definen sus semánticas. Las funciones de pertenencia comúnmente utilizadas incluyen la función triangular, trapezoidal y gaussiana, cada una con sus propias características y aplicaciones ideales. La definición de parámetros de las funciones de membresía puede realizarse mediante métodos como la partición óptima univariada o a partir de rangos predefinidos.

Partición Óptima Univariada: Este método, desarrollado por Fisher en 1958, se utiliza para determinar los cortes que minimizan la varianza dentro de los grupos y maximizan la varianza entre grupos. Esta técnica es útil cuando no se tiene una teoría subyacente clara que guíe la segmentación de los datos.

Código 2 *Partición Óptima Univariada*

```
install.packages("classInt")
library(classInt)
intervalos <- classIntervals(datos$variable,  n = 5,
                   style = "fisher")
```

Rangos Predefinidos: Los rangos predefinidos se utilizan cuando se tiene conocimiento previo de los límites de los datos. Estos rangos se pueden definir de manera que representen adecuadamente la variabilidad observada en los datos.

Código 3 *Definición de Rangos*

```
limites <- c(0, 1, 3, 5)
a <- limites[1]
b <- limites[3]
c <- limites[4]
triangular <- function(x, a, b, c) {
  return(pmax(pmin((x-a)/(b-a), (c-x)/(c-b)), 0))
}
```

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

b) Borrosificación en la Práctica Analítica

La borrosificación convierte las valoraciones estandarizadas de los panelistas en números difusos, utilizando funciones de pertenencia definidas para transformar los valores numéricos en grados de pertenencia a conjuntos difusos. Los datos borrosos se representan en una tabla de contingencia, permitiendo un análisis más robusto y realista.

Código 4 *Transformación a Datos Borrosos*

```
library(FuzzyR)
infer <- lapply(1:nrow(datosTransf), function(i) {
  evalmf(datosTransf[i, ], calma)
})
```

3. Análisis Multivariado

Para el análisis multivariado de los datos borrosos, se utilizan métodos factoriales de correspondencias y análisis factorial múltiple. Estas técnicas permiten identificar patrones y relaciones significativas entre las variables difusas, facilitando la interpretación de los datos.

Código 5 *Análisis Factorial*

```
install.packages("FactoMineR")
library(FactoMineR)
res.famd <- FAMD(datosTransf, graph = FALSE)
```

Tablas de Contingencia y Pruebas Estadísticas

Las tablas de contingencia se utilizan para analizar y comprender la relación entre variables categóricas. Para manejar frecuencias pequeñas o valores decimales, se pueden escalar los datos. Las pruebas estadísticas como la prueba exacta de Fisher y el valor de test se aplican para evaluar la independencia de las variables.

Código 6 *Prueba Exacta de Fisher*

```
tabla <- table(datos$variable1, datos$variable2)
fisher.test(tabla)
```

4. Aplicaciones Shiny de R

Las aplicaciones Shiny de R permiten la interacción dinámica con los usuarios, facilitando la selección de variables, el número de intervalos y la precisión de los datos. Shiny proporciona una plataforma versátil para visualizar y analizar los datos de manera efectiva.

Código 7 *Aplicación Shiny Básica*

```
# Código de R:
install.packages("shiny")
library(shiny)
ui <- fluidPage(
  titlePanel("Análisis Difuso"),
  sidebarLayout(
    sidebarPanel(
      selectInput("variable", "Variable:",
                  choices = names(datos))
    ),
    mainPanel( plotOutput("distPlot")  )
  )
)
```

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

```
server <- function(input, output) {
  output$distPlot <- renderPlot({
    hist(datos[[input$variable]],
         main = input$variable,
         xlab = "Valor", ylab = "Frecuencia")
  })
}
shinyApp(ui = ui, server = server)
```

Discusión

La lógica difusa se ha utilizado ampliamente en diversas áreas para manejar la incertidumbre y la vaguedad en los datos. Comparada con los métodos tradicionales, la lógica difusa ofrece una mayor flexibilidad y realismo en el análisis de datos imprecisos. En estudios previos, como los realizados por Bonissone y Decker (2013), y Espinilla et al. (2008), se ha demostrado que la lógica difusa mejora la interpretación de datos complejos y subjetivos.

Por ejemplo, Bonissone y Decker (2013) utilizaron lógica difusa para aplicaciones empresariales, mostrando cómo se pueden modelar situaciones con datos inciertos y vagos. Espinilla et al. (2008) compararon diversos métodos de fusión de datos heterogéneos en sistemas difusos, concluyendo que la lógica difusa proporciona resultados más precisos y significativos que los métodos convencionales. Bouhentala et al. (2019) proponen mejoras en los algoritmos de sistemas difusos que permiten una optimización más eficiente de los procesos de decisión, lo que es particularmente útil en situaciones donde la imprecisión y la incertidumbre son predominantes..

En el contexto del análisis multivariado, la lógica difusa permite manejar la complejidad de los fenómenos estudiados, como la percepción de calidad de vida o la satisfacción del cliente. La metodología propuesta en este artículo extiende estos enfoques al combinar la lógica difusa con aplicaciones Shiny de R, facilitando la visualización y el análisis interactivo de los datos imprecisos.

Las aplicaciones Shiny proporcionan una plataforma poderosa para implementar y visualizar el análisis difuso, permitiendo a los usuarios interactuar con los datos y ajustar los parámetros del análisis en tiempo real. Esto mejora significativamente la accesibilidad y la aplicabilidad de la lógica difusa en diversos contextos de investigación y práctica aplicada.

Conclusiones

El uso de la lógica difusa y las aplicaciones Shiny de R en el análisis multivariado de datos imprecisos proporciona una metodología innovadora y efectiva para manejar la incertidumbre y mejorar la calidad de los análisis. Este enfoque permite representar la imprecisión de manera más realista y obtener resultados más fiables, lo que es crucial para la investigación y la toma de decisiones informadas.

Referencias

Agayan S., Kamaev D., Bogoutdinov S., Aleksanyan A., Dzeranov B. (2023). Time Series Analysis by Fuzzy Logic Methods. Algorithms, 16(5): (p. 238-238). https://doi.org/10.3390/a16050238

Bazila, Qayoom., M., A., K., Baig. (2021). Mathematical Interpretation of Fuzzy Information Model. (p. 459-466). https://doi.org/10.1007/978-981-16-1740-9_37

Bonissone, P. P., & Decker, K. S. (2013). Fuzzy logic applications in business. Communications of the ACM, 36(3), (p. 32-43). https://www.scirp.org/journal/paperinformation?paperid=104159

Bouhentala, M., Ghanai, M., & Chafaa, K. (2019). An enhanced EDA algorithm for fuzzy systems. Expert Systems with Applications, (p. 131, 157-167). https://www.sciencedirect.com/science/article/abs/pii/S0165011418303348

AJEA (Actas de Jornadas y Eventos Académicos de UTN)
DOI: https://doi.org/10.33414/ajea.1738.2024
Publicado / Published: 10/2024

Cengiz, Kahraman., Selcuk, Cebi., Sezi, Cevik, Onar., Başar, Öztayşi. (2022). Recent Developments on Fuzzy AHP&ANP Under Vague and Imprecise Data. International Journal of the Analytic Hierarchy Process, 14(2) https://doi.org/10.13033/ijahp.v14i2.1033

Djurayev R., Matkurbonov D, Khojiakbar U. (2023). Analysis of a Model for Improving the Efficiency of Routing Control in Data Transmission Networks Based on Fuzzy Logic. Communications. https://doi.org/10.11648/j.com.20231101.11

Espinilla, M., Martínez, L., Pérez, A., & Liu, J. (2008). A comparative study of heterogeneous data fusion methods in fuzzy systems. Expert Systems with Applications, 34(3), (p. 2102-2112). https://www.worldscientific.com/doi/abs/10.1142/S0218488512400120

Hasbi, Ulumuddin. (2022). Fuzzy Inference Algorithm Using Databases. (p. 444-451). https://doi.org/10.1007/978-3-031-35314-7_39

Irina, Bolodurina., E., A., Speshilov. (2023). Application of fuzzy logic rules for data analysis and decision-making in cargo transportation management under conditions of uncertainty. Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control & Radioelectronics, 23(2): (p.52-64). https://doi.org/10.14529/ctcr230205

Kwiatkowski M., Zacharias B., Leung C., Tsu P., Thomas J., Kolisnyk M., Cuzzocrea A. (2022). A Fuzzy-Logic Based Multi-Dimensional Analysis of Traffic Incident Data. (p.1-8). https://doi.org/10.1109/FUZZ-IEEE55066.2022.9882787

Shiyou, Lian. (2020). A New Theoretical and Technological System of Imprecise-Information Processing. viXra, https://arxiv.org/pdf/1610.02751

Solo, Ashu & Gupta, Madan. (2022). Fuzzy Logic Theory and Applications in Uncertainty Management of Linguistic Evaluations for Students. (p. 243-266) https://www.igi-global.com/chapter/fuzzy-logic-theory-and-applications-in-uncertainty-management-of-linguistic-evaluations-for-students/289194

Swathi C., Ebienazer J., Swathi M., Suruthipriya S. (2023). Fuzzy Logic. International journal of innovative research in information security, 09(03): (p.147-152). https://doi.org/10.26562/ijiris.2023.v0903.19

Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), (p. 338-353). https://www.sciencedirect.com/science/article/pii/S001999586590241X