

Predicción de ataques al corazón : usando Fuzzy C Means Clustering

Jesús Emmanuel Ramos Dávila

Resumen—Enfermedades cardiovasculares son una de las principales causas de muerte global, con millones de fallecimientos año tras año. Las enfermedades cardiovasculares son un grupo de desórdenes del corazón y los vasos sanguíneos no existe causa exacta para este tipo de enfermedades. Existen algunos patrones de síntomas los cuales están muy asociados a tener este tipo de enfermedades. Actualmente no existen demasiadas investigaciones con respecto a el análisis de grupos y algoritmos de asociación para este tipo de enfermedades. En esta sección se realizara un estudio usando el algoritmo FCM (Fuzzy C Means) Clustering, para determinar el riesgo de un ataque al corazón, en este estudio se realizara un prueba usando 303 observaciones se revisara el performance y la precisión con respecto a otro algoritmo ya conocido.

I. INTRODUCCIÓN

Enfermedades cardiovasculares son una especie que se están presentando con mayor frecuencia y estas frecuentemente suceden en fallecimientos. La Organización mundial de la salud ha estimado alrededor de 12 millones de muertes alrededor del mundo anualmente, debido a este numero avances en la medicina en las últimas décadas habilito la identificación de factores de riesgo que podrían contribuir en este tipo de enfermedades cardiovasculares. La causa más común en este tipo de enfermedades es el estrechamiento o bloqueo de las arterias coronarias. Los vasos que transportan sangre al corazón mismo, Este es llamado enfermedad arteriopatía coronaria y esta sucede comúnmente con el paso del tiempo. Esta es una de las principales causas por las cuales las personas sufren ataques al corazón, Es por eso que un bloqueo que no es tratado dentro de las primeras horas causa que el musculo del corazón muera. Diagnósticos médicos son una importante, pero a la vez una tarea complicada y su automatización podría ser muy útil. Desafortunadamente no todos los doctores están especializados en esta área y no se tiene en algunos casos y no se tienen los mismos recursos médicos. Es por eso que para utilizar el conocimiento de diferentes especialistas y los datos clínicos de pacientes para facilitar el proceso de diagnóstico es considerado muy valioso ya que su integración en tomas de decisiones medicas podría reducir los errores médicos , mejorar la seguridad del paciente y reducir practicas no deseadas.

II. METODOLOGÍA

El principal objetivo de esta búsqueda es implementar el algoritmo de Fuzzy C Means Clustering usando nuestros datos de pacientes, esto para poder usarse en el soporte de toma de decisiones, por lo tanto, se desarrolla un modelo Fuzzy C Means para predecir los ataques al corazón. El algoritmo Fuzzy C-Means Clustering fue desarrollado en 1981 este es

1. Initialize the matrix of μ_{ij} , U
2. Update c_i :

$$c_i = \frac{\sum_{j=1}^N \mu_{ij}^m x_j}{\sum_{j=1}^N \mu_{ij}^m}$$

3. Update μ_{ij} :

$$\mu_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_j - c_k\|}{\|x_j - c_i\|} \right)^{\frac{2}{m-1}}}$$

4. If the change of U between two iterations is very small (predefined cutoff), then stop; otherwise, go to step #2.

Algorithm of FCM (Image by author)

Figura 1: Imagen de pasos algoritmo FCM [1]

Cuadro I: Descripción de las 13 variables dentro del dataset.

| Proposal | Condition | Identification |
|----------|--|---|
| Sex | Gender | Male/Female |
| Age | Age of patient | 20-90 |
| cp | Tipo de dolor de pecho | 0 : Angina típica 1: Angina atípica 2: non-anginal pain 3: asintomático |
| trtbps | Azucar en ayunas >120 | 290 |
| chol | Colesterol en mg dl | valor numerico |
| rest ecg | Resultado electrocardiograma en reposo | 0 : normal 1:ST-T wave abnormality 2: ventricular hypertrophy by Estes criteria |
| thalach | maximum heart rate achieved | valor numerico |
| exang | Angina de Pecho Inducida | 1:Yes 0: No |
| old peak | ST depression induced by exercise | valor numerico |
| slp | slope peak ST segment | 0 = unslowing 1 = flat 2 = downsloping |
| caa | number of major vessels | 0-3 |
| thall | thalassemia | 0 = null 1 = fixed defect 2 = normal 3 = reversable defect |

†

un extendido del algoritmo K-Means Clustering, FCM (Fuzzy C-Means) es un algoritmo no supervisado que es aplicado hacia un rango muy amplio de problemas conectados con análisis de características , clustering y clasificación. FCM también es usado en otros campos además de la medicina, tales como agricultura, ingeniería, astronomía, química, análisis de imágenes. FCM es una técnica de clustering en la cual un conjunto de datos es agrupado en n-clústeres en la cual cada punto de datos esta relacionado a un clúster el cual tendrá un alto grado de pertenencia hacia este punto siendo así que los puntos de datos que tengan un bajo grado de pertenencia hacia este clúster estarán más alejados de este.

III. ADQUISICIÓN DE DATOS

Dentro de este análisis el conjunto de datos es obtenido de UC Irvine [2] , Los datos han sido recolectados de 303

pacientes de un subconjunto conjunto de datos que contiene 13 columnas y su variable de respuesta la cual es si el si el paciente presenta enfermedad del corazón o no.

IV. RESULTADOS

Se aplico el modelo FCM usando las 303 observaciones a fin de encontrar los mejores clústeres que representen nuestros datos.

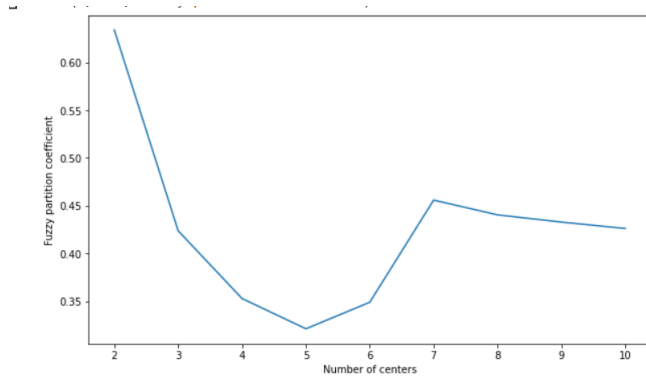


Figura 2: Grafica creada en el analisis de nuestro

Los datos obtenidos en la métrica de evaluación fueron 2 clústeres como los óptimos en una iteración de 1 hasta 9 clústeres, representado claramente en con nuestra grafica de codo. Figura 2 La métrica de evaluación para este modelo fue su **coeficiente de particiones de fuzzy (FCP)** la cual marco un notable porcentaje en 2 clusters.

V. DISCUSIÓN

Se observo un buen desempeño con respecto a el calculo del mejor número de clústeres siendo de 300ms aproximadamente sin la funcionalidad del multithreading. Los valores de las métricas que usa el modelo FCM nos permitió claramente ver el número de clústeres. Una de las comparaciones a realizar a futuro seria comparar con un algoritmo tradicional tal como K-Means o K-Medoids a fin de comparar tiempo y si es tan notable el encuentro del mejor clúster.

REFERENCIAS

- [1] Yufeng, "Fuzzy c-means clustering with python," Medium, 11 2021. [Online]. Available: <https://towardsdatascience.com/fuzzy-c-means-clustering-with-python-f4908c714081>
- [2] A. Janosi, "UCI machine learning repository," 2017. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>