

Cuantificación de Riesgo Transmisión de COVID-19 por Sistemas de Inferencia de Lógica Difusa

1st Rafael Santos Arnedo Mendoza
Departamento de Ingeniería
Universidad Tecnológica De Bolívar
Cartagena, Colombia
rs.arnedomendoza@gmail.com

2nd Arnovis Buendia De Avila
Departamento de Ingeniería
Universidad Tecnológica De Bolívar
Cartagena, Colombia
arnovisdee@hotmail.com

Abstract—In the COVID-19 scenario, a tool is proposed to estimate the level of risk of contagion of areas and concentration points through a fuzzy logic inference system. Two LIS were designed, one for areas and another for concentration points. The first was designed taking into account population density (PD), degree of confinement (CG) and percentage of infected (PC), the second the concurrence (C) and CG of the corresponding zone. Besides, it has a visual tool that allows us to see in a map the zones divided by risk. A dataset of Cartagena was used, with the Unidades Comuneras de Gobierno (UCG) to evade the SILDs obtaining a satisfactory response to the design.

Index Terms—COVID-19, Risk, Fuzzy, Map

I. INTRODUCCIÓN

En marco de la pandemia por enfermedad de coronavirus(COVID-19), múltiples herramientas han aparecido con el fin de disminuir la transmisión y facilitar la detección de casos sospechosos. Aplicaciones como CoronApp creada por el Gobierno Nacional de Colombia [1] es utilizada para dar recomendaciones y/o determinar si una persona es sospechosa de tener COVID-19. El gran potencia del uso de la inteligencia artificial para esto tipo de herramientas, el uso de machine learning para detectar de una forma más rápidas los casos [2] y redes neuronales convolucionales para detectar COVID-19 y descartar si es otra enfermedad respiratoria [3] dan evidencia de este hecho. Otros como [4] han implementado sistemas de inferencia de lógica difusa(SILD) para la identificación de COVID-19 en donde se mide en de la severidad de la infección del paciente a través de factores como la temperatura ambiente, temperatura corporal, nivel de etanol y medición de otros aspectos provocados por la enfermedad. Por otro lado, Kihore propone un sistema tiene como variables de entrada la edad, el grado confinamiento, la políticas de aislamiento y la disponibilidad de elementos de protección [5]. En este trabajo, se presentará un SILD que cuantifique el riesgo basados en metricas de movilidad, población y contagios el cual sera mostrado por medio de un mapa de color.

II. PROBLEMA

El SARS-CoV-2 es un virus que surge en Wuhan, China que genera la enfermedad COVID-19. El virus se transmite de persona a persona a través de las gotículas que se expulsan al hablar, toser y estornudar por lo cual se sugiere el aislamiento,

distanciamiento y el uso de tapabocas [6], sin embargo las personas se ven obligadas a salir para realizar compras o actividades esenciales con riesgo de contraer la COVID-19. El número de casos nuevos crece constantemente identificar las zonas riesgosas ayuda a la prevención de contraer el virus. La clasificación de las zonas de riesgo de contagio permite que las personas pueden observar el nivel de exposición que se encuentra al concurrir en un lugar lo que podría ayudar a las personas a escoger por ejemplo, a que supermercado ir o porque lugares transitar.

III. MÉTODOS

La figura 1 muestra el procedimiento de un algoritmo de lógica difusa. Se reciben valores nítidos como entradas, los cuales se mapean a través de un fuzzificador, usando un modulo de inferencia,el cual contiene una reglas que son usadas para hacer tomar la decisión. El modulo de inferencia da como resultado unos valores difusos que se convierten a nítidos a través de un defuzzificador.

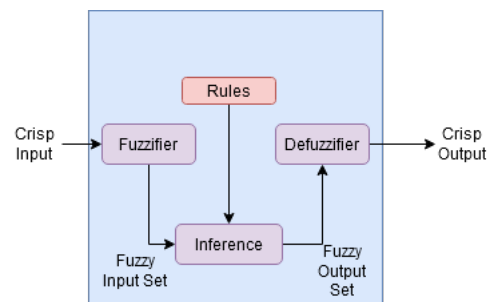


Fig. 1: Representación del algoritmo lógica de difusa

A. Algoritmo

El algoritmo de lógica difusa se basa en:

- Definir inicialmente las variables y/o términos lingüísticos
- Se construyen las funciones de pertenencia
- Se deben construir la base de las reglas que funcionaran para hacer la inferencia
- Se transforman los datos nítidos en conjuntos de datos difusos utilizando funciones de membresía (fuzzificación)

- Se evalúan las reglas en base de las reglas previamente hechas (inferencia)
- Se combinan los resultados de cada regla (inferencia)
- Se convierten los datos de salida en va no difusos (defuzzificación)

B. Fuzzificación

La fuzzificación consiste en determinar la posibilidad de pertenencia al conjunto difuso. Podemos definir el conjunto difuso como el conjunto de observaciones sobre las que se aplica la misma función de pertenencia, el conjunto lleva asociada una lógica matemática que nos permite definir ciertas operaciones entre conjuntos difusos.

C. Funciones de Pertenencia

La función de pertenencia de un conjunto X indica el grado en que cada elemento de un universo dado, pertenece a dicho conjunto X. Estas funciones son usadas para representar de manera gráfica un conjunto borroso perteneciente a un universo.

D. Valor difuso

Las operaciones de conjuntos difusos realizan una evaluación de las reglas. Se combinan todos los resultados de la evaluación para formar un resultado final. Este resultado es un valor difuso.

E. Defuzzificación

La defuzzificación se realiza de acuerdo con la función de pertenencia para la variable de salida, y de acuerdo a los métodos matemáticos de desdifusión se obtiene un valor concreto para esta variable.

IV. DISEÑO DE SILD

En el diseño del SILD, se consideran 2 riesgos. El primero perteneciente a un área(e.g. un barrio, una localidad) y el segundo correspondiente a un punto de concentración o aglomeración(e.g. Supermercados, Centros Comerciales). El área posee propiedades, las cuales son explicadas posteriormente, de densidad de población(DP), porcentaje de contagiados(PC) y grado confinamiento(GC) a diferencia del punto de concentración, en donde se tendrá en cuenta el GC en el área en el cual se encuentra este y el nivel de concurrencia(C) a este punto. Así, se establece que deben haber 2 sistemas uno diseñado para el área y otro para los puntos.

A. Entradas

1) *DP*: medirá que tan espaciado puede estar una persona de la otra, es decir, distanciamiento social en caso de que salieran todas las personas de sus casas en habitantes por km^2 .

Valores lingüísticos y Función de membresía:

- Alta: Sigmoide(30000,0.001)
- Media: Gauss(20000,5000)
- Baja: Sigmoide(10000,-0.001)

2) *PC*: medirá la cantidad de contagiados por cada habitante en el área.

Valores lingüísticos y Función de membresía:

- Alta: Sigmoide(60,0.5)
- Media: Gauss(40,10)
- Baja: Sigmoide(20,-0.5)

3) *GC*: Permitirá medir el nivel de aislamiento de las personas. Estará dado de 0 a 100 donde 0 representa un nivel de aislamiento nulo y 100 una población totalmente confinada.

Valores lingüísticos y Función de membresía:

- Alta: Sigmoide(80,0.5)
- Media: Gauss(60,10)
- Baja: Sigmoide(20,-0.5)

4) *C*: Los puntos de concentración pueden variar en riesgo si este no están visitado, es decir, si no es un lugar concurrido. Por esto, se propone utilizar este parámetro para medir el grado de concurrencia del punto. Donde su dominio será de 0 a 100 en el cual 0 representa un lugar no visitado y 100 un lugar altamente concurrido.

Valores lingüísticos y Función de membresía:

- Alta: Sigmoide(50,1)
- Media: Gauss(30,10)
- Baja: Sigmoide(15,-1)

B. Salida

1) *Riesgo*: Para los dos SILD, salida será el riesgo. Como fue mencionado anteriormente, para el SILD para puntos de concentración (SILDPC) tendrá como entrada GC y C, y las entradas para el SILD para áreas(SILDA) son DP, PC y GC. Los valores lingüísticos y la función de membresía para ambos SILD están dadas por lo siguiente:

- Alta: Sigmoide(50,1)
- Media: Gauss(30,10)
- Baja: Sigmoide(15,-1)

C. Reglas

A continuación, se presenta las reglas para cada sistema.

1) *Reglas de SILDA*: Debido a que este sistema contiene 3 variables, realizamos un grafo para contemplar todas las combinaciones y eliminando algunas redundancia obtenemos reglas en la tabla I

GD	PC	DP	Riesgo
Alto	Alto	Medio	Bajo
Alto	Alto	Bajo	Bajo
Alto	Medio	-	Bajo
Alto	Bajo	-	Bajo
Medio	Bajo	Bajo	Bajo
Alto	Alto	Alto	Medio
Medio	Alto	Bajo	Medio
Medio	Medio	Bajo	Medio
Medio	Bajo	Alto	Medio
Medio	Bajo	Medio	Medio
Medio	Alto	Alto	Alto
Medio	Alto	Medio	Alto
Medio	Medio	Medio	Alto
Medio	Medio	Medio	Alto
Bajo	-	-	Alto

TABLE I: Reglas para SILDA

2) *Reglas de SILDPC*: Las reglas obtenidas para este sistema se encuentra registradas en la tabla II.

CGC	Alta	Medio	Baja
Alta	Medio	Alto	Baja
Media	Medio	Alto	Alta
Baja	Alta	Medio	Medio

TABLE II: Reglas para SILDPC

V. DATOS

Se conformo un dataset para realizar la prueba los SILD en el cual estará dividido por Unidades Comunera del Gobierno(UCG). Se usarán los datos de *Cartagena Como Vamos* de población [7] y para extraer el área de las UCG. Por otro lado, La cantidad de contagiados fueron extraídos de la Alcaldía de Cartagena [8]. Los puntos de concentración fueron extraídos Google Maps para la UCG2 y UCG3 para realizar la prueba. Cabe decir que GC y C no fácilmente medible así que en estos campos del dataset se coloca un valor cualquiera para la prueba.

VI. VISOR DE RIESGO

Con el fin de obtener un una herramienta visual para mostrar el riesgo en cada zona, se emplea el uso de la API *Gmaps* para Jupyter Notebook. Está permite usar Google Maps en un Notebook así, el visor está compuesto por un mapa, en donde se mostrarán la zona en un gradiente de verde a rojo. Para un valor de riesgo bajo, la zona será verde mientras que para un valor alto será rojo. Adicionalmente, en los puntos de concentración habrá un gradiente del mismo color en un radio definido en el cual la intensidad del rojo varia de acuerdo al nivel de riesgo en ese punto.

VII. PRUEBA DE SILD

Para observar la variabilidad de la salida de SIDA ingresamos parámetros de prueba no correspondiente a ningún área. En la tabla III se muestra estas pruebas y su resultados. Se observa que cumple con los requerimientos ya que muestra el GC alto y PC dando como un bajo riesgo. Por otro lado, cuando el confinamiento es bajo aun así que la DP y PC son bajos el riesgo de transmisión es alto.

	GC	PC	DP	Riesgo
Pruebas	75	75	20000	68.44
	40	75	40000	73.771
	20	10	10000	74.449
	80	10	40000	20.569

TABLE III: Prueba de SIDA

El SILDPC fue probado con los datos registrados en la tabla IV. A un GC alto y baja C el riesgo es de 30% y para GC bajo y alta C el riesgo es mayor.

	GC	C	Riesgo
Pruebas	70	10	31.048
	60	10	30.843
	60	80	74.462
	20	60	73.771

TABLE IV: Prueba de SLIDPC

VIII. EVALUACIÓN DE DATASET

Para este caso, se utilizaran dos localidades UCG2 y UCG3. La tabla V muestra los datos del dataset para cada localidad. El riesgo para cada localidad, en este caso, depende principalmente de la GC y DP ya que el PC es bajo por la cantidad de contagiados de la zona.

Localidad	GC	GC	Densidad	Riesgo
UCG2	80	0.12	16501.62	38.114
UCG3	60	0.10	29048.27	74.464

TABLE V: Evaluación de riesgo de Localidades

Para cada localidad se identificaron puntos de concentración los cuales tienen como valor de confinamiento el valor de la zona donde se encuentra, es decir que para los puntos de concentración dentro del UCG2 tendrán el confinamiento de 80 en este caso. La tabla VI y VII muestran los valores de riesgo para puntos de concentración de UCG2 y UCG3.

Puntos de Concentración	C	Riesgo
1	50	41.87
2	32	41.87
3	32	41.87
4	23	44.123

TABLE VI: Evaluación de puntos de concentración para UCG2

Puntos de Concentración	C	Riesgo
5	32	73.91
6	33	74.05
7	42	74.41
8	80	74.46
9	23	55.93

TABLE VII: Evaluación de puntos de concentración para UCG3

En la figura 2, se encuentra el visor observe que para la zona UCG2 es un color amarillo el cual representa un valor de riesgo medio y la zona UCG3 rellena de color rojo para indicar alto nivel de riesgo. Además, Observa la intensidad de los puntos de concentración, la cual cada punto varia de acuerdo al riesgo mostrados en la tabla VI y VII.

