

ALGORITMO DE PREDICCIÓN DE ÍNDICES UV-B BASADO EN CASOS, MEDIANTE MEDICIONES DE MÍNIMA DISTANCIA EUCLIDIANA

UV-B INDEX PREDICTION ALGORITHM BASED ON CASES BY MINIMUM EUCLIDEAN DISTANCE MEASUREMENTS

¹Wilson Ferriol Mamani Yanapa; ²José Herrera Quispe

RESUMEN

El presente trabajo propone un nuevo modelo de predicción de índices de radiación UV-B a partir de datos estadísticos anteriores. Esta propuesta extiende el modelo de proceso estocástico propuesto por Herrera (2013) donde se cambian los parámetros del componente determinístico por una función de similitud que usa la distancia euclidiana multidimensional ponderada y retardos de tiempo. Se adjunta un componente aleatorio heredado de un modelo Auto-regresivo de orden 1. El modelo se aplica en la simulación de escenarios climáticos con altos índices de radiación UV-B. Se busca evaluar el porcentaje de respuesta de la radiación solar, además de validar la efectividad del algoritmo para datos proporcionados por la dirección regional de salud de Arequipa.

Palabras Clave: algoritmo de predicción, razonamiento basado en casos, radiación UV-B.

ABSTRACT

It is proposed a new model for predicting levels of UV-B radiation from previous statistical data. This proposal extends the model of stochastic process proposed by Herrera (2013) where the deterministic component parameters are changed by a function of similarity using the weighted multidimensional Euclidean distance and time delays; it is attached a random component inherited an Auto-regresivo model of order 1. The model is applied in the simulation of climate scenarios with high levels of UV-B radiation, to assess the response rate of solar radiation, as well as validate the effectiveness of the algorithm for data provided by the regional Directorate of Arequipa health.

Key Words: prediction algorithm, case-based reasoning, uv-b radiation

I. INTRODUCCIÓN

El agujero de la capa de ozono apareció en los años 80 sobre la Antártida durante las primaveras del hemisferio sur, debido a la acumulación de gases freón, los cuales están relacionados con la atmósfera a través de diversas actividades humanas [2]. La baja concentración de ozono en la capa correspondiente forma agujeros que permiten el ingreso de mayor cantidad de radiación ultravioleta de cortas longitudes de onda (llamadas UV-B) Los bordes del agujero abarca el sur del continente australiano, Nueva Zelanda y el continente sudamericano [3].

Durante las últimas décadas, las campañas públicas de salud han sido implementadas difundir los riesgos de la radiación UV. En una parte importante de los programas de prevención se recomienda el uso estricto de implementos de protección, con el objetivo de reducir el cáncer de piel. La protección solar es de particular importancia en distintos grupos de riesgo (pacientes con xeroderma pigmentosum,

síndrome nevus de células básicas o receptores de trasplantes de órganos), quienes tienen alta probabilidad de desarrollar cáncer de piel.

Esto ejerce una significativa influencia en numerosos procesos que toman lugar en la atmósfera y la biósfera. La radiación UV afecta la salud humana (induce Eritema o quemadura de sol, daños en el ADN, cáncer de piel y cataratas, favorece la síntesis de vitamina D). Produce degradación de materiales (polímeros, pigmentos y tintes). Induce modificaciones en la química atmosférica y ciclos biogeoquímicos [5]; afectando plantas y ecosistemas terrestres [6]; fotosíntesis impares en muchas especies de plantas e incrementa la susceptibilidad de las plantas a enfermedades [7], afectando el cultivo y desarrollo de cosechas [8], daños en ecosistemas acuáticos, reduce la productividad de fitoplancton e induce cambios en comunidades microbiológicas e invertebradas, los cuales pueden afectar el ciclo de nutrientes en los suelos [8].

La Organización Meteorológica Mundial (OMM),

¹Ingeniero Electrónico. Aspirante al grado de Magister en Ingeniería de Sistemas e Informática de la Escuela de Posgrado de la Universidad Nacional Jorge Basadre Grohmann. Tacna-Perú.

²Doctor en Ciencias de la Computación, Maestro en Ciencias: Ingeniería de Sistemas-Gerencia en Tecnología de la Información. Universidad Nacional San Agustín, Cátedra CONCYTEC-UNSA. Arequipa-Perú.

la Organización Mundial de la Salud (OMS) y el Programa de Naciones Unidas para el Medio Ambiente (PNUMA) han propuesto la creación de un índice simple que informa a la población del riesgo, este es el Índice Ultravioleta (IUV) que varía entre 0 a 16 y tiene cinco rangos: Bajo (1 a 2), moderado (3 a 4), alto (5 a 7), muy alto (8 a 10) y extremadamente alto (11 a ms) [9].

En adición, el análisis de varios parámetros de exposición solar ha demostrado contundentemente una asociación entre el desarrollo de melanomas malignos y las intensas y cortas exposiciones de radiación UV, principalmente en la niñez [10].

Según el Ministerio de Salud del Perú, cada año son diagnosticados más de cuatro mil casos de cáncer de piel, debido, entre otros factores, a los elevados índices de radiación ultravioleta a los que están sobreexpuestas las personas.

El Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), realiza monitoreo de la radiación UV-B en diferentes ciudades del país y publica diariamente el índice IUV [12].

Por ello es de vital importancia la implementación de modelos computacionales capaces de pronosticar los índices de radiación UV-B, a fin de estudiar el comportamiento y realizar acciones de prevención en la población local.

II. MATERIALES Y MÉTODOS

2.1. Conceptualización sobre algoritmo de predicción UV-B

Algoritmo

Es un conjunto prescrito de instrucciones o reglas bien definidas, ordenadas y finitas que permite realizar una actividad mediante pasos sucesivos que no generen dudas a quien deba realizar dicha actividad [14]

Muchos algoritmos son ideados para implementarse en un programa. Sin embargo, los algoritmos pueden ser implementados en otros medios, como una red neuronal, un circuito eléctrico o un aparato mecánico y eléctrico.

Ejemplo:

```
...
i:=1;
read(n);
while i < n do begin
  (* cuerpo del bucle *)
  I := i + 1
end;
...
```

Predicción

El término predicción puede referirse tanto a la acción y al efecto de predecir [15] como a las palabras que manifiestan aquello que se predice; en este sentido, predecir algo es anunciar por revelación, ciencia o conjetura algo que ha de suceder [15].

La predicción constituye una de las esencias claves de la ciencia, de una teoría científica o de un modelo científico. Así, el éxito se mide por el éxito o acierto que

tengan sus predicciones [16].

Radiación UV

En el área de la fotobiología, se define como radiación ultravioleta al intervalo de radiaciones electromagnéticas menores a 400nm [17]. La radiación ultravioleta se puede subdividir en cuatro intervalos:

- La que produce envejecimiento (aging) denominada UV-A (315 a 400nm),
- La que produce quemadura (burning) denominada UV-B (280 A 315nm),
- La que los autores denominan cosmic UV-C (200 a 280nm) y finalmente,
- La denominada "vacío" vacuum (debido a que como es absorbida por el oxígeno y nitrógeno solo es posible que se propague en condiciones de vacío) y corresponde a fotones con longitud de onda menor a 200nm.

2.2. Sistema basado en casos

Un Sistema Basado en Casos para la toma de decisiones en un Sistema de Decisión S como se presenta en (1), y se define en términos de un par $(U, X \cup Y)$ donde U es un conjunto finito no vacío de objetos, eventos llamados casos, mientras X y Y son dos conjuntos finitos, no vacíos, de atributos o propiedades llamados rasgos predictores y objetivos respectivamente.

$$S = (U, X \cup Y) \quad (1)$$

Cada rasgo predictor $x_i \in X$ puede ser considerado una función, como la que aparece en (2), que mapea elementos de U en el conjunto $M_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, el cual se denomina conjunto de valores del rasgo predictor X_i .

$$x_i: U \rightarrow M_i \quad (2)$$

De manera análoga cada rasgo objetivo $y_j \in Y$ puede ser considerado una función que mapea elementos de U en el conjunto $N_j = \{y_{j1}, y_{j2}, \dots, y_{jn}\}$ que se denomina conjunto de valores del rasgo predictor Y_j . Dicha función puede verse en (3)

$$y_j: U \rightarrow N_j \quad (3)$$

Las componentes de un Sistema Basado en Casos son: la base de conocimiento o base de casos, el módulo de recuperación de casos y el módulo de adaptación de las soluciones.

Base de casos

La base de casos contiene las experiencias o casos con los cuales el sistema realiza sus inferencias y puede ser representada a través de una tabla, cuyas columnas son etiquetadas por variables o atributos que representan los rasgos predictores y objetivos, mientras que sus filas representan los casos. En la tabla 1, se representa la estructura de la base de casos de nuestro problema. El conjunto U está formado por los casos O_1, O_2, \dots, O_m ; el conjunto X por los atributos x_1, \dots, x_n , mientras que el conjunto Y lo integra el atributo y_1 .

Tabla N°01. Representación de la estructura de la base de casos.

	Rasgos predictores	Rasgo objetivo
Casos	$x_1 \dots x_n$	y_t
O_1	$x_1(O_1) \dots x_n(O_1)$	$y_1(O_1)$
O_2	$x_1(O_2) \dots x_n(O_2)$	$y_1(O_2)$
...
O_m	$x_1(O_m) \dots x_n(O_m)$	$y_1(O_m)$

En nuestro problema, el conjunto U estaría integrado por los 365 días del año estudiados; el conjunto X lo integrarían los índices de radiación registrados anteriormente (rasgos predictores), almacenados en una base de casos y el conjunto Y estaría formado por los índices que siguen a un grupo de vector con rasgos predictores anteriores.

Módulo de recuperación

El proceso de recuperación consiste en determinar los casos de la base más sencilla, de ellas se cuenta el número de rasgos predictores similares entre ambos. Otra técnica utiliza un conjunto de heurísticas que determinan las variables de mayor importancia (peso) en la determinación del rasgo objetivo y se formula una función que teniendo en cuenta esto se considera la semejanza o diferencia entre cada uno de los rasgos predictores.

A continuación se describe el algoritmo del proceso de recuperación.

Entrada: O_n, O_i (O_n : nuevo problema y O_i : caso de la base). Salida: $\beta(O_n, O_i)$ (Medida de semejanza o distancia entre O_n y O_i)

Para cada caso O_x se requiere:

i) Buscar el caso O_x en los valores de los rasgos predictores x_i ($x_i(O_n)$ y $x_i(O_i)$).

ii) Calcular una medida de la semejanza $\delta_i(x_i(O_n), x_i(O_i))$ entre estos valores.

Así se define que δ se obtiene a partir de la siguiente ecuación de distancia euclidiana definida por (4):

$$\delta_i(x_i(O_n), x_i(O_i)) = \sqrt{(\sum (x_i(O_n) - x_i(O_i))^2)} \quad (4)$$

Posteriormente se evalúan las distancias entre cada uno de los rasgos predictores y los casos, obteniéndose un rasgo objetivo como respuesta $\beta(O_n, O_i)$, que es el pronóstico con menor distancia euclidiana o con mayor semejanza.

Finalmente $Y_i(O_n)$ recibe el valor del índice pronosticado.

2.3. Algoritmo de predicción UV-B

Los casos están contenidos en un archivo *csv, de fácil manipulación, pero sin un formato para el procesamiento, entonces se requiere definir la longitud de los rasgos predictores (tamaño del vector), posteriormente se organizan los casos según el tamaño del vector. El presente algoritmo solicita ingresar valores de índice del día, luego internamente procesa los datos actuales, con los datos

de conocimiento referencial y obtiene una respuesta de pronóstico UV-B. El algoritmo está contenido en un bucle while loop, para hacer pronósticos continuos hasta presionar 0 para salir.

Código Matlab:

```

clc
clear
% carga data de conocimiento
dato=csvread('\2007.csv');

disp('* PREDICCIÓN DE VALORES UV-B *')
resp=input('Continuar? si(1)/no(0):');
while (resp == 1)
hora=input('Ingrese la HORA: ');
data_hora=dato(hora+1,2:end);
vector=5
ind=0;
for i=1:vector
    in=input('Valor de Indice: ');
    ind=[ind in];
end
ind(:,1)=[];
m_rasp=r_predictor(vector,data_hora);
min_dist=m_distancia(vector,m_rasp,ind);

disp('El pronostico es: ')
respuesta=min_dist
respuesta=input('Continuar? si(1)/no(0): ');
clc
end

```

III. RESULTADOS Y DISCUSIÓN

Se utilizaron casos de 3 años anteriores, registrados diariamente por el Ministerio de Salud, que forman una parte del conocimiento a consultar, como parte de los rasgos predictores, a fin de obtener el rasgo objetivo (Predicción), obteniéndose lo que se observa en la Figura N°01.

Se obtienen también las curvas de respuesta para los índices de radiación UV-B introducidos y mostrados en la Figura N°02.

Mediante pruebas de eficiencia en respuesta de pronóstico, se probaron diferentes tamaños del rasgo predictor (tamaño de vector), así se obtuvieron las mejores respuestas con vectores de tamaño 5 a 10 rasgos predictores con 10 pruebas para cada test, como se muestra en la Figura N°03. Donde se observa un promedio de 15 por ciento de diferencia entre el pronóstico y el valor real, es decir aproximadamente más del 85% de efectividad en el pronóstico.

IV. CONCLUSIONES

Durante el desarrollo del algoritmo fue necesario tres etapas: Bosquejo, Seudocódigo y Programación. Durante el bosquejo, se entiende el problema a resolver, se consulta la fuente bibliográfica para explicar las variables de

```

**** PREDICCIÓN DE VALORES UV-B *****
Continuar? si(1)/no(0): 1
Ingrese la HOPA a Analizar: 12

vector =

5

Ingrese Valor de Índice: 0.7
Ingrese Valor de Índice: 0.6
Ingrese Valor de Índice: 0.5
Ingrese Valor de Índice: 0.6
Ingrese Valor de Índice: 0.7
El pronostico para los datos ingresados es:

respuesta =

0.6410

Continuar? si(1)/no(0):
    
```

Figura N°01: Interfaz del usuario Matlab

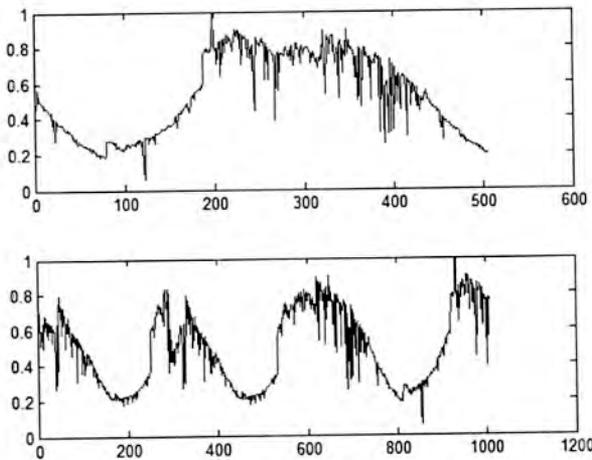


Figura N°02: Curvas de respuesta

Ingrese el numero maximo de pruebas: 10

Numero Pruebas	Tamaño Vector en % 3	4	5	6	7	8	9	10
1.0000	4.0189	18.3099	8.1395	6.0000	5.4167	44.0101	24.9084	57.2917
2.0000	1.5544	7.2000	10.3139	4.3062	7.2539	3.1250	2.6820	22.9773
3.0000	15.3846	52.2463	177.1930	5.2632	25.4902	36.6142	11.2727	5.8323
4.0000	5.9859	5.4688	8.0977	14.3639	14.1026	0.4762	5.1230	269.5652
5.0000	1.5625	3.5014	10.2672	1.4815	20.6897	3.6773	67.9245	7.5853
6.0000	1.3986	10.0649	0.3984	25.4601	2.2222	14.3695	11.9084	2.9907
7.0000	16.4804	1.8315	0	3.1000	40.9574	166.6667	9.7744	40.7821
8.0000	38.8889	11.6667	2.2346	2.2495	2.2727	8.2759	7.4074	4.2654
9.0000	11.8421	2.6667	0.9317	5.8407	6.9913	31.9149	15.0538	32.6531
10.0000	3.5928	4.8000	3.6697	25.8065	3.7383	28.8344	10.9155	36.4286

promedio =

5.5000	10.0709	11.7756	22.1246	9.3872	12.9135	33.7964	16.6970	48.0372
--------	---------	---------	---------	--------	---------	---------	---------	---------

Figura N°03: Test de algoritmo para diferentes tamaños predictores

entrada sus comportamientos y variables de salida. En la etapa de pseudocódigo se migra el problema resuelto en código “borrador”, revisando la lógica y secuencia de solución. Finalmente en etapa de programación se implementa en lenguaje Matlab.

Matlab es una herramienta informática de procesamiento. Sirve efectivamente durante el modelamiento de técnicas de predicción y en diversos campos de la investigación científica. Posee numerosas herramientas TOOLBOX capaces de simular diversos escenarios.

El algoritmo propuesto es una solución posible para analizar los índices de radiación UV-B, en ciudades costeras de Perú.

El presente modelo ofrece respuesta con más del 85% de efectividad; sin embargo es necesario continuar el estudio de este tipo de algoritmo para ajustar mejores resultados.

REFERENCIAS BIBLIOGRÁFICAS

[1] Herrera, José, 2013. Tesis Doctoral “Modelo Estocástico a partir de razonamiento basado en casos para la generación de series temporales”, Doctorado en Ciencias de la Computación. Universidad Nacional San Agustín. Arequipa-Perú.

[2] Chubachi, S., 1984. Preliminary result of ozone observations at Syowa station from February 1982 to January 1983. Memoir of National Institute of Polar Research, Special Issue 34. In: Proceeding of Sixth Symposium of Polar Meteorology, 13e19.

[3] Longstreth, J., de Gruijlb, F.R., Kripkec, M.L., Abseckd, S., Arnolde, F., Slaperf, H.I., Veldersg, G., Takizawah, Y., van der Leun, J.C., 1998. Health risks, Journal of Photochemistry and Photobiology B 46, 20-39.

[4] M.F. Holick, Evolution and function of vitamin D, Recent Results Cancer Res, 164 (2003)3-28.

[5] UNEP, 2007. UNEP report on environmental effects of ozone depletion, Photochem. Photobiol. Sci. 6, 201-332.

[6] Heisler, G.M., Grant, R.H., Gao, W., Slusser, J.R., 2003. Ultraviolet radiation and its impacts on agriculture

and forests, *Agric. For. Meteorol.* 120, 37.

[7] Caldwell, M.M., Bjorn, L.O., Bornman, J.F., Flint, S.D., Kulandaivelu, G., Teramura, A.H., Tevini, M., 1998. Effects of increased solar ultraviolet radiation on terrestrial ecosystems, *J. Photochem. Photobiol. B* 46, 40-52.

[8] Kakani, V.G., Reddy, K.R., Zhao, D., Sailaja, K., 2003. Field crop responses to ultraviolet-B radiation, *Agric. For. Meteorol.* 120, 191-218. doi:10.1016/j.agrformet.2003.08.015.

[9] OMS, OMM, PNUMA-Organización Mundial de la Salud, Organización Meteorológica Mundial, Comisión de Protección contra la Radiación No Ionizante. Índice UV Solar Mundial: Guía Práctica, Ginebra, Suiza 2003.

[10] A. Osterlind, M.A. Tucker, B.J. Stone, O.M. Jensen, The Danish case control study of cutaneous malignant melanoma. II. Importance of UV-light exposure, *Int. J. Cancer* 42 (1988) 319-324.

[11] Dr. Alfaro, Cliff, dermatólogo del Hospital de la Solidaridad de Lima. Entrevista RPP, 09-01-2013

[12] SENAMHI Servicio Nacional de Meteorología. Vigilancia de la radiación ultra violeta-B en ciudades importantes del país. Lima, Per, 2012 vol.:8.

[13] Joo F. Escobedo a, Eduardo N. Gomes a, Amauri P. Oliveira b,*, Jacyra Soares b, Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu, Brazil, 2008.

[14] Brassard, Gilles; Bratley, Paul (1997), *Fundamentos de Algoritmia*, Madrid: PRENTICE HALL

[15] DRAE, Vigésima segunda edición

[16] Ferrater, José. *Diccionario de filosofía*, Ariel, Barcelona, p. 2883.

[17] Cabrera, Sergio S. *Radiación Ultravioleta y Salud*, Editorial Universitaria, 2005.

Correspondencia:

Wilson Ferriol Mamani Yanapa: coronel8421@hotmail.com

José Herrera Quispe: jherreraq@unsa.edu.pe

Fecha de Recepción: 25/06/2014

Fecha de Aceptación: 09/07/2014