

Modelos de diagnóstico: Diferencial, aditivo y probabilidad condicional

María Eugenia Ambrois¹,
Isabel Morales, Andrea Mattiozzi² and Franco Simini²
Franca Oppici, Pablo Ahumada, Claudio Garuti y Mario Sandoval³

¹*Estudiante del XXX Seminario de Ingeniería Biomédica 2021.*

²*Docentes del XXX Seminario de Ingeniería Biomédica 2021.*

³*Directores de Medical Sapiens y conferencistas del XXX Seminario de Ingeniería Biomédica 2021.*

Abstract— The purpose of this study is to describe the operational modality of medical diagnosis methods such as additive, differential and Bayes Probability. Analytic Hierarchy Process and Decision Trees models are used as an example for additive and differential diagnosis. Whereas that Decision Trees need large databases to be reliable, Analytic Hierarchy Process method is based in experts opinion and algebra. For Bayes Probability test sensibilities and medical probabilities are the main components. All models are used in real life, and which one to use depends on the medical experts. However, there is evidence Bayesian Networks included in Bayes Probability successfully diagnose more than 90% of medical diagnosis cases in which it was proven.

Keywords— Medical Diagnosis, Multiple Criteria Problems, Analytic Hierarchy Process, Decision Trees, Bayesian Network.

Resumen— La presente monografía pretende describir la modalidad operativa de los esquemas de diagnóstico médico aditivo, diferencial y teorema de probabilidad de Bayes. Se utilizarán los métodos Analytic Hierarchy Process, Árboles de Decisión como ejemplo de modelos aditivo y diferencial respectivamente. Mientras que los Árboles de Decisión necesitan de bases de datos grandes para ser un método confiable, el Analytic Hierarchy Process se basa en la opinión de expertos y algebra matricial. Por otro lado, la sensibilidad de los tests aplicados y las probabilidades médicas son los principales componentes del modelo de Probabilidad Condicional. Todos los modelos son utilizados en la vida real, y cual usar es decisión de los especialistas en medicina. Sin embargo, existe evidencia que indica que las Redes Bayesianas incluidas en el método por Probabilidad de Bayes diagnostica correctamente más del 90% de los casos de diagnóstico médico en los que fue probado.

Palabras clave— Diagnóstico médico, Problemas Multicriterio, Proceso de Análisis Jerárquico, Árboles de Decisión, Redes Bayesianas.

INTRODUCCIÓN

El diagnóstico médico se puede describir como el conocimiento técnico en el que se encuentra el estado de salud de una persona, si es sano para mejorarlo, si es una enfermedad para descubrir las causas e intentar curarlo [1]. Existen varios factores que se deben tener en cuenta para realizar una hipótesis correcta de la causa de los síntomas del paciente, como el interrogatorio entre médico-paciente (anamnesis), el examen físico y la estrategia para determinar el diagnóstico [2].

El diagnóstico puede ser catalogado como diferencial según [3] si se basa en la identificación de una enfermedad mediante la eliminación de posibles causas que presenten síntomas similares. El médico puede realizar el procedimiento manualmente, al basarse en bibliografía y antecedentes hospitalarios para definir cuáles son las decisiones que se desprenden de los resultados de los exámenes médicos. A su vez, se van descartando y agregando hipótesis de la causa raíz. Este modelo se representa mediante árboles de decisión. Sin embargo, también se puede utilizar este método mediante inteligencia artificial utilizando algoritmos de generación de árboles.

Distintas enfermedades pueden compartir síntomas, por ello “no se puede afirmar que una persona tiene una enfermedad solamente con saber que se tiene un síntoma de esta.” [4]. Es por esto que se utilizan otras estrategias para llegar a un diagnóstico, como los Métodos de Decisión Multicriterio (MCDM), de los cuales se describirá el método aditivo mediante Analytic Hierarchy Process (AHP) introducido por Thomas Saaty en 1980, y por otro lado el método de Probabilidad Condicional con Redes Bayesianas, basadas en el Teorema de Bayes, donde se obtiene la solución con mayor probabilidad dada las evidencias obtenidas y sus incertidumbres.

I. MATERIALES Y MÉTODOS

Para el desarrollo de los modelos operativos de las técnicas de inteligencia artificial del diagnóstico diferencial, aditivo y probabilidad condicional se eligieron algoritmos mencionados por la bibliografía consultada como los más utilizados y representativos de los mismos. Como ya se mencionó anteriormente estos son; Modelo por Árboles de Decisión, Analytic Hierarchy Process y Redes Bayesianas.

A. ÁRBOLES DE DECISIÓN

Los árboles de decisión son modelos de predicción con aprendizaje inductivo, es decir, se respaldan en bases de datos que los ayudan a predecir comportamientos. En medicina, estas bases de datos consisten en bibliografía como [3] o antecedentes hospitalarios.

El nodo raíz es el motivo principal de consulta, con el cual se inicia el proceso. A partir de él, se van desprendiendo otros nodos de decisión que corresponden a las preguntas acerca del atributo anterior, estos se les llama nodos internos. Las ramas que salen de los nodos se encuentran etiquetadas por los posibles valores de esta característica. El camino a seguir en el árbol es determinado por las decisiones (resultados de test, síntomas) tomadas en cada nodo anterior. Los nodos finales corresponden a una decisión. [5]

Un ejemplo tomado de [6] se muestra en la figura a continuación, donde para un paciente con problemas respiratorios se tiene un árbol con resultados de CT (CTF), X-Ray (XRF), dolor en el pecho (CPT), exámenes de sangre (BTF). El médico va a pedir un examen de Rayos X si el dolor en el pecho es “1”, mientras que si es “2” pedirá un examen de sangre.

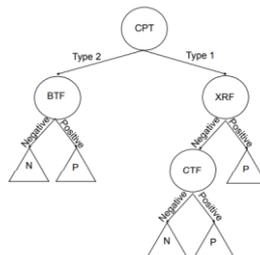


Figura 1. Árbol de decisión para ejemplo obtenido de [6]

Para describir el algoritmo que utiliza este modelo se introducirán los siguientes conceptos:

1. Conjunto de Entrenamiento: Conjunto de vectores que contienen valores de atributos determinados. Consiste de información previamente obtenida con la que se puede predecir el comportamiento de un sistema mediante sus propios valores a los atributos determinados. [6]
2. Algoritmo de Inducción: El algoritmo de inducción toma el conjunto de entrenamiento y forma un modelo que generaliza la relación entre los atributos y los resultados obtenidos. [6]

El algoritmo comienza con el nodo raíz, donde se desprende la primera etapa a la que se le llama proceso de clasificación. La misma consiste en elegir un atributo que divida el conjunto de entrenamiento en dos o más subconjuntos. Para cada partición se genera un nuevo nodo y así sucesivamente. La segunda etapa, o etapa de clasificación, consta de clasificar los objetos nuevos por clases. [5]

Existen varios algoritmos de clasificación, siendo uno de los más utilizados el ID3 [5] [6]. Éste determina cual es el atributo que brinda mayor información sobre la clase terminal o cual es el atributo con menos entropía. La entropía se define como una medida de incertidumbre promedio donde p_i es la probabilidad de que el atributo S pertenezca a la clase i :

$$Entropía (S) = \sum_{i=1}^c -p_i \cdot \log_2 p_i \quad (I)$$

EJEMPLO

[5] Comprende un estudio en el que se ponen a prueba varios algoritmos de decisión, entre ellos el ID3. Para ello utilizaron dos bases de datos que contienen información acerca de tumores para la detección del cáncer de mama. Cada base de datos consiste en datos históricos sobre las mismas variables de entrada sobre el cáncer de mama, algunas de ellas se muestran en la figura a continuación.

Característica observada (variable)	Valores que puede tomar la variable	Definición
Edad	uno, dos o tres	Toma valor de "uno" si la paciente tiene menos de 50 años, "dos" si está entre 50 y 70 años y "tres" si tiene más de 70 años.
Dimensión celular	verdadero o falso	Toma valor de "verdadero" si la mayoría de las células epiteliales de la paciente se encuentran dentro de grupos adhesivos y valor de "falso" si la mayoría están dentro de grupos cohesivos.
Lumina intracitoplasmática	verdadero o falso	Toma valor de "verdadero" si en las células epiteliales de la paciente está presente esta característica y "falso" si está ausente.

Figura 1. Variables de entrada para la base de datos del ejemplo obtenido de [5]

En primer lugar dividieron cada una en dos subconjuntos; uno para ser utilizado como conjunto de entrenamiento, y el otro para probar la capacidad de calificar de los algoritmos. Es así que concluyeron que el 93.04% de los casos fueron clasificados correctamente por el algoritmo ID3. En segundo lugar, utilizaron una base de datos como conjunto de entrenamiento y otra base de datos para probar el algoritmo. Esta vez el 82.60% de los casos fueron clasificados correctamente.

B. REDES BAYESIANAS

El objetivo es modelar un problema de diagnóstico médico como una red bayesiana. Las mismas se basan en el Teorema de Bayes, que permite calcular la probabilidad de un suceso A, dado un suceso B del cual tenemos información y su generalización se describe como [4] :

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{P(B/A_1) \cdot P(A_1) + \dots + P(B/A_i) \cdot P(A_i)} \quad (II)$$

Un problema de diagnóstico modelado como una red bayesiana queda definido por las siguientes variables y está constituida por los mismos conceptos que la teoría de grafos [4]:

1. Variable principal X "Tener la enfermedad" que puede tomar los valores "Si"(S) o "No"(N)
2. Variables de Evidencia E_i ; $i=1, \dots, r$ que apoyan si se tiene o no la enfermedad.

Donde el resultado es determinado por la hipótesis con mayor probabilidad teniendo en cuenta un conjunto de observaciones y condiciones [7]. Puede ser expresado matemáticamente como la ecuación:

$$Diagnostico = \max(P(D_i/E)) \quad (III)$$

Donde $P(Di/E)$ es la probabilidad de la enfermedad Di dada la evidencia E. Siendo esta ultima un conjunto de síntomas, tests de laboratorio, exámenes físicos, entre otros [7].

Para resolver la red en primer lugar se debe fijar la variable principal y la modalidad de interés, por ejemplo $X=S$ que significa que el paciente tiene la enfermedad. Luego, se determina la probabilidad a priori del suceso de interés $P(S)$, se calcula la distribución conjunta, es decir, $P(S \cap E1 \cap E2 \cap \dots \cap Er)$ y por último se calcula la probabilidad a posteriori, es decir $P(S | E1 \cap \dots \cap Er)$ [4].

Un problema genérico en diagnóstico medico se representa en la figura a continuación, donde $O \in \{O_1, O_2, \dots, O_n\}$ son posibles orígenes de la enfermedad (diagnostico sindrómico), $S \in \{S_1, S_2, \dots, S_k\}$ son síntomas de la enfermedad y $P \in \{P_1, P_2, \dots, P_m\}$ son resultados de pruebas. [4]

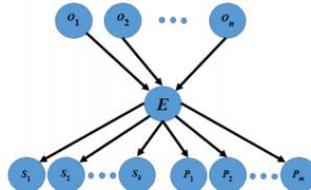


Figura 2. Grafo usual de red bayesiana para un problema de diagnóstico médico obtenida de [4]

EJEMPLO

Se usará este método mediante un ejemplo utilizado en [4]. Se plantea que un paciente acude a una consulta médica y el facultativo le aplica un test para determinar si tiene cierta enfermedad. Se dice que el test tiene una sensibilidad del 95%, lo que significa que da positivo el 95% de las veces que se aplica a una persona enferma, y tiene una especificidad del 98% , lo cual significa que da negativo el 98% de las veces que se aplica a un sano. También se supone 0.5% de la población tienen la enfermedad. El ejemplo continua diciendo que tras aplicar el test el resultado da positivo y se plantea la pregunta “¿Cuál es la probabilidad de que dicho paciente padezca realmente la enfermedad?”.

Se definen dos variables aleatorias: tener o no tener la enfermedad (E) y el resultado del test (T).



Figura 3. Grafo de red bayesiana asociada al ejemplo obtenida de [4]

La variable E tiene dos posibles modalidades y sus probabilidades son estar enfermo (S) o no estar enfermo (N): $P(S) = 0.005$ y $P(N) = 0.995$.

Además, la variable T también puede tener dos modalidades: positivo o negativo. Las probabilidades de que de positivo si es enfermo, y dar negativo si es sano están dadas en el problema, y de ellas se desprenden las probabilidades conjuntas (ver tabla 1).

TABLA I.

PROBABILIDADES CONJUNTAS OBTENIDAS DE [4]

	+	-
S	$P(S, +) = P(+ S)P(S) = 0.00475$	$P(S, -) = P(- S)P(S) = 0.00025$
N	$P(N, +) = P(+ N)P(N) = 0.0199$	$P(N, -) = P(- N).P(N) = 0.9751$

Por lo tanto [4]: $P(S|+) = 0.1927$

[4] Concluye que la probabilidad de que el paciente padezca la enfermedad ha pasado del 0.5% (probabilidad a priori) al 19% (probabilidad a posteriori.)

Como se vio anteriormente para el algoritmo ID3 sobre árboles de decisiones en diagnóstico diferencial, [5] comprueba las clasificaciones realizadas por varios algoritmos utilizando dos bases de datos sobre cáncer de mama. Este experimento también se realiza utilizando Redes Bayesianas. Para el primer caso ya mencionado, éste modelo logra clasificar 94.35% de casos correctamente, mientras que para el segundo caso 85.71%.

C. ANALYTIC HIERARCHY PROCESS AHP

El modelo de diagnóstico por evidencia aditiva a diferencia del método diferencial no elimina enfermedades durante el proceso de diagnóstico. Se basa en la aplicación de métodos de toma de decisión multicriterio (MCDA) y se hará referencia específica al método AHP (Analytic Hierarchy Process) creado por Thomas Saaty. [8]

El método ANP (Analytic Network Process) es creado por Thomas Saaty en 2001 para generalizar el modelo AHP de tomas de decisión. El ANP permite incluir todos los factores y criterios con sus dependencias y feedback mientras que AHP estructura los problemas jerárquicamente. [9]

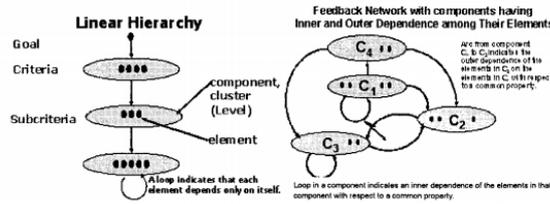


Figura 4. Diferencia entre ANP y AHP obtenida de [9]

El método AHP según [10] considera ciertos parámetros de evaluación y un conjunto de opciones dentro de las cuales se debe determinar la mejor decisión. Se genera un peso para cada criterio de evaluación siendo los criterios más importantes los más pesados. Luego se realizan matrices de comparaciones entre los elementos y utilizando algebra matricial se llega a una optimización que se toma como resultado. La escala de comparación creada por [10] permite definir el peso entre dos opciones en una escala de 1 a 9:

TABLA II.

ESCALA FUNDAMENTAL OBTENIDA DE [10]

1	Igual de importantes
3	Moderadamente preferida
5	Fuertemente preferida
9	Extremadamente importante
2,4,6,8	Valores intermedios

Se va creando la matriz A de tamaño mxm, donde m es la cantidad de criterios considerados. Cada entrada aij es la importancia del criterio i frente el j según la Escala Fundamental. Luego se normaliza la matriz A, es decir, se divide cada término de la matriz sobre la suma de sus columnas:

$$\overline{a_{jk}} = \frac{a_{jk}}{\sum_{l=1}^m a_{lk}} \tag{IV}$$

Finalmente, se consigue un vector m-dimensional de prioridad de criterio al promediar los valores de cada fila:

$$w_j = \frac{\sum_{l=1}^m \overline{a_{jl}}}{m} \tag{V}$$

Para cada criterio se crea una matriz B, de tamaño nxn, donde n es la cantidad de opciones. Cada valor bij representa el peso de la opción i sobre la opción j para el criterio determinado. Se aplican los mismos pasos que para la matriz A, de normalización y obtención de un vector s. Se obtienen m vectores s. [10]

Una vez que se obtienen S y w, el método AHP propone obtener un vector v multiplicando S y w, donde vi representa el puntaje global asignado a la opción i:

$$v = S \cdot w \tag{VI}$$

Se ordenan los puntajes de cada opción en orden decreciente, la mejor decisión corresponderá al mayor puntaje.

Durante el proceso pueden surgir discordancias de criterios provenientes del error humano, es por esto que el último paso es chequear la consistencia de los juicios emitidos mediante la Proporción de Consistencia. Definida como el cociente entre el Índice de Consistencia (CI) y el Índice Aleatorio (RI), y debe ser menor a 10%. [11]

$$CI = \frac{\lambda_{\max} - m}{m - 1} \quad (\text{VII})$$

Donde λ_{\max} es el mayor valor propio de la matriz traspuesta de la matriz de comparaciones pareadas y m el rango de la matriz. El Índice Aleatorio es un índice de consistencia de una matriz aleatoria [11]:

TABLA III.

ÍNDICE ALEATORIO RI OBTENIDO DE [10]

m	2	3	4	5	6	7	8	9	10
RI	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.51

II. RESULTADOS

	DIFERENCIAL	ADITIVO	PROBABILIDAD CONDICIONAL
INTELIGENCIA ARTIFICIAL	Árbol de Decisión	AHP	Redes Bayesianas
ALGORITMO	Inductivo	Matemático	Matemático
PESO DE LA EVIDENCIA	Histórico	Decisor	Sensibilidad de Test, Datos Historicos
SOLUCION	Mayores coincidencias	Resultado de una maximización	Mayor probabilidad
CALCULO DE CONFIABILIDAD	No	Consistencia (CI/RI)	P(N/+)
REPRESENTACION GRAFICA	Si	Si	Si
EN SUS REPRESENTACIONES GRAFICAS			
NODOS	Atributos: Enfermedades, Síntomas, Tests	Criterios y Alternativas (Enfermedades)	Alternativas (Enfermedades), Síntomas, Resultados de Tests
LINKS	Valores que pueden tomar los atributos	Dependencias wij (de 1 a 9)	Probabilidades (de 0 a 1)

III. CONCLUSIONES

Existen varios métodos para definir un diagnóstico médico. El especialista puede realizarlo manualmente mediante métodos diferenciales, o mediante métodos aditivos, diferenciales y de probabilidad apoyándose en inteligencia artificial.

Los sistemas diferenciales computarizados se basan en arboles de decisiones, el más utilizado es el ID3 que tiene como ventaja su sencillez. Sin embargo, la capacidad predictiva de este modelo depende notoriamente del tamaño del conjunto de entrenamiento, que consiste en una gran planilla de información sobre valores de atributos relevantes respecto enfermedades ya diagnosticadas. De acuerdo a [6] se espera que la capacidad sea pobre si el conjunto de entrenamiento es pequeño.

El diagnóstico es un problema de la vida real, con múltiples criterios y variables, por ello se utilizan Métodos de Toma de Decisión Multi Criterio. Los mismos consisten en ordenar una toma de decisión y hallar la mejor solución según la teoría en la que se basen, mediante un algoritmo matemático. Se describieron los modelos AHP y Redes Bayesianas.

En el modelo AHP a diferencia del modelo diferencial, el decisor le da un valor de peso a cada criterio, considerando si es o no más importante que otro. También se genera una representación más realista con ANP al permitir definir conexiones entre los atributos, eliminando la estructura lineal o jerárquica de otros métodos. En este modelo un especialista define la importancia de cada criterio según su experiencia profesional y bibliografía relevante.

Por otro lado, las Redes Bayesianas se basan en las sensibilidades de tests, probabilidades médicas estudiadas o históricas. Permite modelar problemas caracterizados por la incertidumbre.

Por más que es evidente la necesidad de una gran base de datos para el método diferencial por Árboles de Decisiones, los datos históricos son importantes para todos los métodos. En el modelo AHP los especialistas ponderan los criterios según su opinión personal, que está condicionada por su experiencia histórica en la materia, así como en los Árboles de Decisión, mientras más experimente el especialista, mejor categorizara los criterios y mejor será el resultado. En las Redes Bayesianas se utilizan probabilidades que han sido conseguidas por investigaciones científicas y datos históricos, asimismo, mientras más información se tenga mayor su complejidad pero más confiable será el diagnóstico [4].

Por lo tanto, qué modelo es mejor depende de cada problema de diagnóstico y la base de datos que se tenga. Los modelos son comparables entre sí solo si las bases de datos son comparables. En la bibliografía consultada, el modelo bayesiano obtiene mejores resultados que el diferencial por más que ambos clasifican correctamente más del 80% de los casos

REFERENCIAS

- [1] P. L. Entralgo, *Diagnóstico Médico, Historia y Teoría*, Barcelona: Salvat, 1982.
- [2] J. Diaz Novas, B. Gallego Machado y A. Leon Gonzalez, «El Diagnóstico Médico: Bases y procedimientos,» *Revista Cubana de Medicina General Integral*, vol. 22, n° 1, pp. 1-11, 2006.
- [3] J. Laso Guzmán, *Diagnóstico Diferencial en Medicina Interna*, Madrid: Elsevier, 2018.
- [4] V. D. Estruch, F. J. Boigues, A. Vidal y J. I. Pastor, «Redes bayesianas y diagnóstico médico,» *Modelling in science education and learning*, vol. 12, n° 2, 2019.
- [5] R. Erandi Barrientos, N. Cruz Ramirez y H. G. Acosta Mesa, «Árboles de decisión como herramienta en el diagnóstico,» *Rev Med UV*, vol. 2, pp. 19-24, 2009.
- [6] L. Rokach y O. Maimon, *Data Mining With Decision Trees: Theory and Applications*, World Scientific, 2014.
- [7] A. Onisko, «Bayesian Networks: A Practical Guide to Applications,» Bialystok, O. Pourret, P. Naim, B. Marcot, 2008, pp. 15-32.
- [8] J. C. Osorio Gómez y J. P. Orejuela Cabrera, «El proceso de análisis jerárquico (AHP) y la toma de decisiones multicriterio,» *Scientia Et Technica, Red de revistas científicas de América Latina, el Caribe, España y Portugal*, vol. XIV, n° 39, pp. 247-252, 2008.
- [9] T. Saaty y L. Vargas, *Decision Making with the Analytic Network Process*, Pittsburgh: Springer, 2006.
- [10] T. Saaty, *The Analytic Hierarchy Process*, Nueva York: McGraw-Hill, 1980.
- [11] B. Muñoz y M. G. Romana, «Aplicación de métodos de decisión multicriterio discretos al análisis de alternativas en estudios informativos de infraestructuras de transporte,» *Pensamiento Matemático*, vol. VI, n° 2, pp. 27-46, 2016.