matemática: Una publicación del ICM – ESPOL
2009, Vol. 7, No. 1

MODELAMIENTO PARA LA IDENTIFICACIÓN DE SINTOMAS RELEVANTES A TRAVES DE LA PROGRAMACIÓN ENTERA

Herrera Daza Eddy[1]

Resumen. *Este trabajo propone un modelo determinístico de programación entera binaria, para la identificación de un número mínimo de síntomas relevantes que están asociados a un conjunto finito de enfermedades, por medio de un índice de relación entre los síntomas y las enfermedades; de tal forma que cada enfermedad puede caracterizarse mejor, teniendo en cuenta el nivel del síntoma asociado a cada enfermedad. La solución que se presenta, utiliza el método de enumeración implícita, para la solución del modelo, ya que es muy ajustable cuando se tiene un número reducido de variables, con la ayuda de un software elaborado para tal efecto. Finalmente se plantea de manera muy general y como trabajo a futuro, el modelo a seguir cuando se considera que las relaciones entre los síntomas y las enfermedades no son deterministicas.*

Palabras Claves: Programación entera, síntomas relevantes, modelo determinístico.

1. INTRODUCCIÓN

Gran parte de las aplicaciones de la matemática y la estadística pueden ser descritas como el proceso de modelización. En términos de Henry (1997) "*un modelo es una interpretación abstracta, simplificada e idealizada de un objeto del mundo real, de un sistema de relaciones o de un proceso evolutivo que surge de una descripción de la realidad* "(Pág. 78). En investigación de operaciones la modelización encuentra una aplicación en la optimalidad, la cual juega un papel importante en la resolución de problemas con un objetivo, en los que se trata encontrar de un conjunto factible de soluciones, la que da el mejor valor al objetivo.

Algunos problemas de optimización, se componen de variables de decisión que son matemáticamente enteras y/o binarias, lo que los convierte en problemas de tipo enteros y/o combinatoriales respectivamente, este tipo de problemas en investigación de operaciones son tratados a través de la programación entera, la cual tiene una gran espectro de aplicación en modelos estructurados, como el del agente viajero, en problemas de programación de inversiones, de localización de industrias, coste fijo etc., ver [4].

A diferencia de los problemas con variables reales, el número de soluciones de un modelo de programación lineal entera es finito, por lo que podría plantearse la posibilidad de encontrar la solución mediante la exploración de todas las soluciones posibles. Sin embargo, el número de soluciones a explorar para un problema puede ser muy elevado, aún utilizando una buena maquina; ya que para un problema con n variables binarias, se tiene 2^n soluciones posibles , es decir que si tenemos un problema con $n = 30$ tendremos $2^{30}= 1073741824$ soluciones posibles.

Los métodos de solución en programación entera se dividen en: Algoritmos de programación entera pura, en el cual están los métodos, que utilizan la lógica de Branch and bound (ramificación y acotamiento) [1] y están incorporadas a la mayoría de programas informáticos que en algunas ocasiones emplea tiempos grandes de resolución y en los algoritmos para problemas de programación entera mixta, donde están entre otros los métodos de enumeración implícita que resultan adecuados cuando el número de variables es bajo.

El método de enumeración implícita o parcial, considera conjuntos parciales de posibles soluciones que sean candidatos (factibles) para determinar entre ellas la solución al problema (optimiza el objetivo), prescindiendo de aquellas que no lo sean.

2. PROBLEMA DE IDENTIFICACIÓN DE SÍNTOMAS RELEVANTES

Considerando que tenemos a la mano un número finito de enfermedades, como también un número finito de síntomas, que están relacionadas a través de un nivel de asociación, y asumiendo que dichas relaciones entre las enfermedades y los síntomas son de carácter deterministicas, se da el siguiente modelo

Problema

Sea $\{D_1,D_2,\ldots,D_n\}$ un conjunto conocido de posibles enfermedades. Considérese que los médicos, al identificar las enfermedades asociadas a un conjunto de pacientes, basan su decisión normalmente en un conjunto finito de síntomas $\{S_1 ,S_2 , \ldots ,S_m\}$. Considérese que se quiere identificar S_a un número mínimo de síntomas, de tal manera que cada enfermedad puede distinguirse de las otras de acuerdo con los niveles de los síntomas en el conjunto S_a.

Sea $D=\{D_1,D_2,\ldots,D_n\}$ conjunto de enfermedades

Sea $S=\{S_1,S_2,\ldots,S_m\}$ conjunto de síntomas

[1] *Herrera Daza Eddy, Pontificia Universidad Javeriana. (e_mail: eherrera@javeriana.edu.co)*

n : número de enfermedades (cardinal de D)
m : número de síntomas (cardinal de S)
c_{ij} : el nivel del síntoma j asociado a la enfermedad i
d_{ijk} : denota si hay o no discrepancia entre las enfermedades i y la enfermedad k debido al síntoma j , en términos de los síntomas incluidos en S_a , donde se define como:

$$d_{ijk} = \begin{Bmatrix} 1 \text{ si } c_{ij} \neq c_{kj} \\ 0 \text{ si } c_{ij} = c_{kj} \end{Bmatrix}$$

a : nivel mínimo requerido de discrepancia

$$X_j : \begin{Bmatrix} 1 \text{ si el sintoma } j \text{ esta presente en } S_a \\ 0 \text{ si el sintoma } j \text{ no esta presente en } S_a \end{Bmatrix}$$

Modelo

$$z(\min): \sum_{j=1}^{j=m} x_j \quad (1)$$

Sujeto a :

$$\sum_{j=1}^{m} x_j d_{ijk} \geq a \quad \forall i,k \in \{1,2,...,n\} \text{con } i \neq k \quad (2)$$

Observaciones

Se tiene que $\sum_{j=1}^{m} d_{ijk} = 0,1,2,3,...,\binom{n}{2}$ mide la discrepancia entre las enfermedades D_i y D_k en términos de los síntomas incluidos en el subconjunto S_a y $a > 0$ es el nivel de discrepancia deseado.

A medida que el valor de a es mayor, mayor será el número de síntomas requeridos., es decir el cardinal del subconjunto S_a debe ser suficiente, para así poder distinguir las enfermedades.

$\sum_{j=1}^{m} x_j d_{ijk}$ coincide con el número de síntomas en S_0 que toman distintos valores para el de enfermedades D_i y D_k, y a es el número mínimo, para cualquier par (D_i, D_k) de enfermedades, necesario para tener un subconjunto aceptable S_a.

Si las enfermedades han de identificarse con alguna carencia de información; es decir cuando $a = 0$, el conjunto S_0 puede resultar inservible. Por tanto, normalmente se emplea un valor $a > 0$.

El objetivo del problema es determinar el subconjunto mínimo del conjunto $S_{ai} \subseteq S$, de manera que la enfermedad i tenga síntomas diferentes comparados con el resto de enfermedades. Este subconjunto se denomina el "conjunto de síntomas relevantes para la enfermedad i".

3. SOLUCIÓN

Se genera una matriz de unos y cero, comparando los niveles c_{ij} y c_{jk}, es decir la matriz generada muestra si hay o no discrepancia entre la enfermedad i y la enfermedad j, con esta matriz se puede seleccionar los síntomas relevantes asociados para identificar cada enfermedad y una vez seleccionados éstos, pueden determinar los síntomas relevantes asociados a la enfermedad i, minimizando (1), teniendo el valor a, que cumpla con la restricción dada (2).

Considérese el conjunto de enfermedades $D= \{D_1,D_2,...,D_n\}$ con $n = 5$ y el conjunto de síntomas $S= \{S_1,S_2,...,S_m\}$ con $m = 8$.

Considérese asimismo que los síntomas asociados a las diferentes enfermedades son los que aparecen en la tabla I.

TABLA I
Modelamiento para la identificación de síntomas relevantes a través de la programación entera
Nivel de asociación entre enfermedades y síntomas

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
D_1	2	3	1	1	1	2	1	2
D_2	1	1	1	1	3	1	2	1
D_3	3	4	2	3	1	1	3	2
D_4	2	2	2	2	2	1	2	3
D_5	1	1	1	2	1	1	1	2

Para cualquier valor de $a > 0$ se puede generar una tabla binaria de discrepancia, realizando $\binom{n}{2}$ combinaciones posibles, como se muestra en la siguiente tabla (Tabla II).

TABLA II

Modelamiento para la identificación de síntomas relevantes a través de la programación entera

Matriz binaria de la comparación de discrepancia

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	$\sum_j d_{ijk}$
D_1 Vs D_2	1	1	0	0	1	1	1	1	6
D_1 Vs D_3	1	1	1	1	1	0	1	0	7
D_1 Vs D_4	0	1	1	1	1	1	1	1	7
D_1 Vs D_5	1	1	0	1	0	1	0	1	4
D_2 Vs D_3	1	1	1	1	1	1	1	1	7
D_2 Vs D_4	1	1	1	1	1	0	0	1	6
D_2 Vs D_5	0	0	0	1	1	0	1	1	4
D_3 Vs D_4	1	1	0	1	0	1	1	1	6
D_3 Vs D_5	1	1	1	1	1	1	1	0	7
D_4 Vs D_5	1	1	1	0	1	0	1	1	6
$\sum_{i=k} d_{ijk}$	8	9	6	8	8	6	8	7	

De esta tabla se genera la (Tabla III), la cual se forma con los valores de acuerdo a las sumatorias en j , la cual muestra las discrepancias por síntoma, es decir se obtiene la cantidad de veces que esta un 1 (uno), cada vez que se compare una de las enfermedades con las demás a través de un síntoma. Como se observa para el síntoma 7 vemos que hay un 1 (uno) en tres de las cuatro veces que se compara la enfermedad dos, y tres veces en la enfermedad 5 y así sucesivamente.

TABLA III

Modelamiento para la identificación de síntomas relevantes a través de la programación entera

Sumatoria de discrepancias por síntoma

Enfermedad	S_1	S_2	S_4	S_5	S_7
D_1	3	4	3	3	3
D_2	3	3	3	4	3
D_3	4	4	4	3	4
D_4	3	4	3	3	3
D_5	3	3	3	3	3

Como se observa para el síntoma 7 vemos que hay un 1 (uno) en tres de las cuatro veces que se compara la enfermedad dos, y tres veces en la enfermedad 5 y así sucesivamente. De esta tabla se extrae un vector de síntomas relevantes por enfermedad, el cual corresponde a los síntomas por enfermedad que tengan el máximo valor. Así se obtiene los síntomas factibles de ser relevantes para todas las enfermedades:

Síntomas factibles

Enfermedad 1: 2
Enfermedad 2: 5
Enfermedad 3: 1, 2, 4, 7
Enfermedad 4: 2
Enfermedad 5: 1, 2, 4, 5, 7

Una vez seleccionados los síntomas en cada enfermedad, se procede a determinar los síntomas relevantes asociados a la enfermedad i, comparando si cada uno de los síntomas relevantes en cada enfermedad, se encuentra dentro del conjunto S_a, esto determinara el subconjunto mínimo. Para poder determinar este subconjunto, se necesita tener en cuenta (1) y (2), el mínimo de síntomas por enfermedad que es 1, ya que el nivel mínimo de discrepancia es a =1, esto genera un vector síntomas relevantes para cada enfermedad:

Enfermedad	**Síntomas relevantes**
Enfermedad 1:D_1	{2}
Enfermedad 2:D_2	{5}
Enfermedad 3:D_3	{2}
Enfermedad 4:D_4	{2}
Enfermedad 5:D_5	{2,5}

Solución del Problema: El número mínimo de síntomas que se necesitan para distinguir las cinco enfermedades es uno, aún en ausencia de dos síntomas (S_3 y S_6).También se puede observar que el síntoma 2 es suficiente para identificar las enfermedades D_1 ,D_3 y D_4 y el síntoma 5, es suficiente para identificar la enfermedad D_2; sin embargo son necesarios los síntomas 2 y 5 para identificar la enfermedad D_5 , por lo tanto el conjunto mínimo de síntomas relevantes queda conformado por los síntomas S_2 y S_5.

Otros Casos

Nivel de discrepancia	**Conjunto mínimo requerido**	**Número de Síntomas**
a = 2 ó 3 ó 4 ó 5	{1,2,4,5,7}	5
a = 6	{1,2,4,5,6,7,8,}	7
a ≥ 7	{1,2,3,4,5,6,7,8}	8

Caso probabilístico

En la mayor parte de las aplicaciones, la incertidumbre es lo común; por ejemplo, una pregunta típica en diagnóstico médico es: dado que el paciente presenta un conjunto finito de síntomas, ¿cuál de las enfermedades posibles de un conjunto también finito, es la que tiene el paciente? Esta situación implica un cierto grado de incertidumbre puesto que: Los hechos o datos pueden no ser conocidos con exactitud tanto como por el paciente como por el médico. Por ejemplo, un paciente puede no estar seguro de haber tenido fiebre o el grado de temperatura, cuando va a consultar a un

médico, por ello, hay un cierto grado de incertidumbre en la información asociada a cada paciente, que puede ser producida por su subjetividad, imprecisión, ausencia de información, errores en la mediciones , datos ausentes, etc.

Puede considerarse entonces que, las relaciones entre las enfermedades y los síntomas no son deterministas, puesto que un mismo conjunto de síntomas puede estar asociado a diferentes enfermedades. De hecho, no es extraño encontrar dos pacientes con los mismos síntomas pero diferentes enfermedades.

Las relaciones entre las enfermedades y los síntomas no son deterministas siempre, puesto que un mismo conjunto de síntomas puede estar asociado a diferentes enfermedades. De hecho, no es extraño encontrar dos pacientes con los mismos síntomas pero enfermedades distintas. Lindley (1987), dice: *"La única descripción satisfactoria de la incertidumbre es la probabilidad. Esto quiere decir que toda información incierta debe estar en forma de una probabilidad, que varias incertidumbres deben ser combinadas usando las reglas de la probabilidad, y que el cálculo de probabilidades es adecuado para manejar situaciones que implican incertidumbre. En particular, las descripciones alternativas de la incertidumbre son innecesarias".*

Modelo.

Supuestos:

- Se dispone de una base de datos con información sobre N pacientes
- Sea $\{D_1, D_2, \ldots, D_n\}$ un conjunto finito de n enfermedades
- Un paciente puede tener una y sólo una de n enfermedades, $D_1, D_2, \ldots, D_n$
- Un paciente puede tener ninguno, uno, o más de un conjunto finito de m síntomas $\{S_1, S_2, \ldots, S_m\}$
- Sea D una variable aleatoria que toma como valores las enfermedades $D_1, D_2, \ldots, D_n$
- Los síntomas son variables binarias,

$$S_i = \begin{Bmatrix} 1 \text{ si la enfermedad } i \text{ esta presente} \\ 0 \text{ si la enfermedad } i \text{ no esta presente} \end{Bmatrix}$$

4. OBSERVACIONES

Cualquier variable aleatoria en el conjunto $\{D, S\}$, define una partición del conjunto formado por todos los paciente, en una clase disyunta y exhaustiva de conjuntos. Entonces, combinando las enfermedades y los síntomas, cada paciente puede clasificarse en una y sólo una región de combinación. El modelo más general posible se basa en especificar directamente la función de probabilidad conjunta, es decir, asignar un valor numérico (parámetro) a cada una de las posibles combinaciones de valores de las variables. Pero, la especificación directa de la función de probabilidad conjunta implica un gran número de parámetros por lo que, no hay ordenador en el mundo capaz de almacenarlo incluso para un valor de n pequeño .Sin embargo, en la mayor parte de las situaciones prácticas, muchos subconjuntos de variables pueden ser independientes o condicionalmente independientes. En tales casos, se pueden obtener simplificaciones del modelo más general teniendo en cuenta la estructura de independencia de las variables, esto da a lugar a una reducción importante del número de parámetros; y a los siguientes modelos:

1. El Modelo de Síntomas Dependientes (MSD).
2. El Modelo de Síntomas Independientes (MSI).
3. El Modelo de Síntomas Relevantes Independientes (MSRI).
4. El Modelo de Síntomas Relevantes Dependientes (MSRD).

5. CONCLUSIONES

Determinar el número mínimo de síntomas relevantes, asociados a una enfermedad, es de gran importancia ya que indudablemente mejora el diagnostico médico y da lugar a un coste mínimo de diagnóstico.

Las pruebas que se visualizan, utilizando los registros médicos son muy prometedoras; como se ve en los trabajos de Peter Kolezar "Testing for Visión Loss in Glaucoma Suspects", [6] el cual realiza una prueba utilizando un modelo también determinístico, con grandes resultados.

La idea a futuro es utilizar el modelo en conjunto con los registros médicos, considerando el caso determinístico, como probabilístico, con miras a desarrollar un sistema experto, de apoyo para las decisiones médicas.

En caso del modelo probabilístico existen trabajos apoyados en la aplicación de Redes Bayesianas, el cual permite realizar cualquiera de los tipos posibles de inferencia probabilística, o sea, causal, diagnostica, intercausal o mixta, como es el caso de del trabajo [8] "An experience in the use of statitics networks for medical diagnosis: a brazilian experience"

En este caso la tecnología utilizada es ideal para el manejo de la incertidumbre, muy común en medicina, además de eso, modela el conocimiento del especialista de dominio de una forma intuitiva.

REFERENCIAS BIBLIOGRÁFICAS Y ELECTRÓNICAS

[1]. **COOK, W.** (1989). Cunningham,W, Pulleyblank, Schrijver, A., Combinatorial Optimization, John Wiley &.Sons.

[2]. **HENRY, M.** (1997). Notion de modéle et modélization en l'enseignement. En *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.

[3]. **HILLIER,F., LIEBERMAN,G.,** (2000). Introduction to Operations Research,7ma ed., McGraw-Hill.

[4]. **MAROTO,C., ALCAZAR JAVIER.** (2002). Investigación operativa: Modelos y técnicas de optimización, Ed. Universidad Politécnica de Valencia.

[5]. **RIOS, S.** (1996), "Investigación operativa: Programación lineal y aplicaciones ",Centro de Estudios Ramón Areces, Editorial Ramón Areces.

[6]. **JSTOR.** (2007), **"Peter Kolesar :Testing for Vision Loss in Glaucoma Suspects",** [en línea] http://links.jstor.org/sici?sici=00251909%28198005%2926%3A5%3C439%3ATFVLIG%3E2.0.CO%3B2-0, Fecha de última visita: febrero de 2008.

[7]. **MASSACHUSETTS INSTITUTE OF TECHNOLOGY.** (2007), "Material Didactico: Programación Entera "[en línea], http://mit.ocw.universia.net/15.053/s02/lecture-notes/index.html, Fecha de última visita: febrero de 2008.

[8]. **PORTAL DE EVIDENCIAS.** (2008), "Evidencias clínicas: An experience in the use of statitics networks for medical diagnosis: a brazilian experience", http://evidences.bvsalud.org/modules/dia/index.php, Fecha de última visita: junio de 2008.

[9]. **UNIVERSIDAD DE CANTABRIA.** (2006), "Departamento de Matemáticas y Computación"Escuela Técnica Superior de Ingenieros, (2006), "Departamento de Matemáticas y Computación" http://departamentos.unican.es/macc/personal/profesores/castillo/Libro/Chap2.pdf, Fecha de última visita: Mayo de 2008.