

Aplicación de Técnicas de Minería de Datos para Análisis de Mediciones Antropométricas del Pie

Sandoval Palomares José J.⁽¹⁾; Maldonado Vega María⁽¹⁾. Cuevas de la Rosa Francisco Javier⁽²⁾

⁽¹⁾ CIATEC.A.C.

Omega #210 Industrial Delta, León Gto. México. C.P.37545. Tel: 01 (477) 7100011 ext. 1505.

jsando@ciatec.mx.

⁽²⁾ CIO. A.C.

Lomas del bosque 115 Col Lomas del campestre. C.P 37150 Tel 01 (477) 4414213 ext. 333

fcuevas@cio.mx

Resumen

A partir de una base de datos antropométricos de los pies de individuos sanos y con padecimiento de diabetes fueron analizados a través de técnicas de minería de datos MD, con el propósito de establecer posibles reglas de asociación. La técnica de MD están soportado en datos de significancia estadística por la aplicación de Chi cuadrada. Este estudio corresponde a un acercamiento a los cambios que podrían estar manifiestos antropométricamente en los pies de los diabéticos considerando únicamente los datos de medidas externas de antropometría de pies, el índice de masa corporal, datos generales de registro y el tiempo de cursar con la enfermedad y son comparados con personas sanas sin el padecimiento. Todos los datos fueron analizados a través del programa Weka^{MR} para establecer las posibles reglas de asociación y de discriminación. Los factores de alta asociación derivados del análisis indican que la diabetes y la obesidad son los factores de alta repetitividad y estrecha interrelación, a estos se las sumaron cambios como las diferencias en los diámetros del empeine y altura del primer dedo del pie, siendo más evidente en individuos diabéticos por más de 10 años con la enfermedad.

Palabras clave: Minería de datos, reglas de asociación, antropometría, diabetes.

1. Introduction

El pie es un sistema altamente organizado, consta de 29 articulaciones, 26 huesos y 42 músculos para realizar coordinada y armónicamente los movimientos, el soporte, la marcha y guardar el equilibrio. Al ir creciendo el pie del ser humano madura para adaptarse a nuestra forma bípeda de

caminar, así mismo su estructura y dimensiones se modifican hasta llegar a un pie maduro.

Previos estudios reportan que no existen deferencias significativas en las mediciones de longitud, ancho y circunferencias entre el pie derecho e izquierdo del mismo individuo, no así en su volumen, donde sí se encuentran diferencias [1]. La funcionalidad, estructura, morfología o dimensiones del pie se pueden ver afectadas al sufrir alguna alteración en el pie, ya sea por traumatismo, amputación, malformación o enfermedad durante la vida del individuo o por enfermedad.

En el caso particular de la diabetes, es una enfermedad con efectos notorios hacia las extremidades inferiores, de tal manera que se tiene diversas manifestaciones debida a neuropatías, macro antipatía, micro e infecciones, que inicialmente predisponen y finamente conllevan a las alteraciones funcionales y morfológicas del pie diabético [2].

En una población de mexicana de diabéticos, sospechamos de diferencias antropométricas en los pies con respecto a una población sana. Por cual se ejecutaron medidas antropométrica en las cuales se medirán circunferencias en los pies serán analizas por métodos de minería de datos así como procedimiento de agrupaciones para valorar la relación o diferencias que pudiera existir, de manera que podamos comprobar o refutar la relaciones de estos grupos y sus mediciones.

2. Metodología

2.1 Participantes y adquisición de datos

La muestra de estudio se conformo por un total de 382 voluntarios mexicanos de la ciudad de León Guanajuato. Se formaron dos grupos principales, el grupo control que no presento a la fecha notificación medica de diabetes y el grupo de

estudio conformado por diabéticos. Ese realizaron a los dos grupos, mediciones antropométrica de ambos pies: empeine, recio, longitud del pie, ancho del pie y altura del dedo u ortejo uno o mayor las medidas expresadas en milímetros (figura 1). El criterio de inclusión para el estudio considero a personas que no presentaran deformidades importantes, lesiones, úlceras en sus pies o problemas en sus capacidad de caminar, tanto en el grupo control como el grupo diabético, para este estudio se considero solo las mediciones realizadas al el pie derecho de los individuos, ya que no se encontraron diferencias significativas entre las mediciones antropométricas de ambos pies.

Referente a padecimiento de la diabetes se consideró rango de 1 a < 5, 5 a < 10 y > 10 años de padecer la enfermedad. Se realizo el cálculo de índice de masa corporal IMC= kg/ m², se considero peso normal a BMI ≤ 25 kg/ m² y sobre peso a aquellos que tenía un BMI > 25 kg/ m².

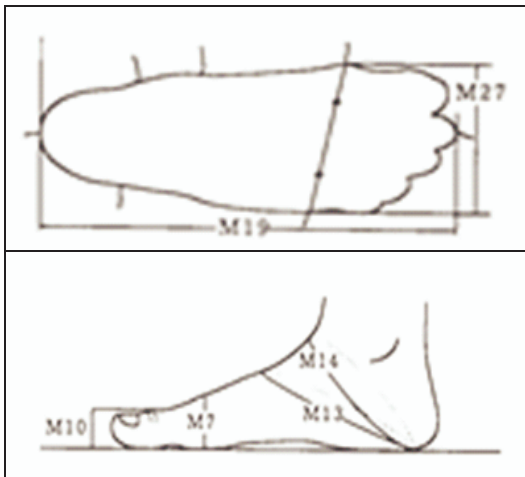


Figura 1: mediciones antropométrica a del pie realizadas en el estudio

2.2 breve revisión de técnicas de minería de datos utilizadas en este estudio

La minería de datos define técnicas y algoritmos para ser utilizados en problemas de tratamiento de datos, con el fin de encontrar conocimiento no trivial y generar una interpretación útil. Las tareas dominadas por la minera de datos son la clasificación, estimación, predicción y agrupamiento.

Las reglas de asociación son usadas para identificar patrones asociaciones de ocurrencia o no ocurrencias en el conjunto de datos, para este propósito sobre las reglas se determina las medidas

de control, soporte (*support*) también llamado cobertura, confianza (*confidence*) e interés (*Lift*), conocida esta ultima como factor de certeza, para cuantificar las relaciones entre los atributos de contenidos en los datos. Las reglas tiene la forma “Si *antecedente*, Entonces *Consecuente*”. Si se tiene un conjunto de elementos *A* y otro conjunto *B*, entonces una regla de asociación, toma la forma “ si *A* entonces *B*” (i.e, $A \Rightarrow B$), donde *A* es el *antecedente* y *B* el *consecuente*.

El soporte de una regla de asociación en particular $A \Rightarrow B$, es el promedio de transacciones en la base de datos analizada, que contienen a *A* y *B* y está dada por la ecuación por la ecuación (1), que se determina por el porcentaje de transacciones en base de datos que contienen *A* y contienen *B*. Si el antecedente y el consecuente no ocurren en la transacción su valore es 0, si ocurrieran en todas las transacciones seria 1.

$$support(A \Rightarrow B) = frecuencia(A \cup B) \quad (1)$$

La confianza, también conocida como significancia, (*confidence*) de la regla de asociación es la medición de exactitud de la regla y está dada por, numero de transacciones que contienen a *A* y *B* entre el número total de transacciones de *A*, ecuación (2), dicho en otras palabras, es la probabilidad de aparecer el consecuente bajo las condiciones de contener el consecuente. Si el antecedente y el consecuente son independientes, su valore seria 0, si ocurriera una dependencia total seria 1.

$$confidence(A \Rightarrow B) = \frac{support(A \cup B)}{support(A)} \quad (2)$$

El nivel de confianza (*lift*) confianza dividida por el numero de ejemplo cubiertos de la parte consecuente de la regla y otras más elaboradas. Es una medida de la importancia de la asociación independiente del soporte y está dada por ecuación (3). Valores mayores a 1 satisfacen la condición de que el antecedente incrementa la probabilidad del consecuente, el antecedente y el consecuente serian independiente si, *lift* es exactamente igual a 1.

$$lift = \frac{confidence(A \Rightarrow B)}{support(B)} \quad (3)$$

Confidence, support y lift son no dan por si solas una significancia estadística [3], una manera de determinar esto es utilizar la *chi cuadrada* χ^2 (4), considerando las tres mediaciones anteriores. En este aspecto la Chi cuadrada, determina el nivel de significancia estadística de la dependencia o

independencia entre el antecedente y el consecuente de las reglas de asociación. Por definición chi-cuadrada mide el grado de independencia entre los atributos, la probabilidad de 0.05 con un grado de libertad su valor es 3.84, así, si $\chi^2 > 3.84$, A y B en nuestro caso, tiene una correlación con un 95% de nivel de confianza (5).

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (4)$$

$$\chi^2 = \frac{N(z-xy)^2}{xy(1-x)(1-y)} \quad (5)$$

De misma manera podemos determinar chi-cuadrada en valores de confidence, support y lift, ecuación (6).

$$\chi^2 = n(\text{lift} - 1)^2 \frac{\text{supp} \cdot \text{conf}}{(\text{conf} - \text{supp})(\text{lift} - \text{conf})} \quad (6)$$

Dado que los datos reales pueden conducir a una extracción de reglas poco útiles, es requerido un transformación, que consiste en convertir los datos, comúnmente discretos a nominales, entendiendo con esto datos numéricos (1, 50, 100) a un equivalente no numérico que represente en esencia de interés lo mismo (min, med, max).

2.3 Análisis de los datos

Una vez concentrada la información, se procedió a la discretización de datos la cual fue como sigue, tabla 1.

Tabla 1: atributos y valores discretizados de información de estudios en la base de información.

Atributo	Variables	Elementos
Gender/Genero	NOMINAL	{M,F}
Group/Grupo	NOMINAL	{Control, Diabetic}
Weight/Peso	NOMINAL	{NormalWeight, OverWeight}
Diabetic_years/Años diabetes	NOMINAL	{Control, Diab1-3, Diab3-10, Diab>10}
Age_Grupo/Grupo Edad	NOMINAL	{Age<18, Age=25, Age=30, Age>30}
BMI_category/IMC_catagoria	NOMINAL	{Nor<25, OW<30, OW_I<35, OW_II<40, OW_III=40}
Ball_right/Empeine_Derecho	NOMINAL	{R160-170, R170-180, R180-190, R190-200, R200-210, R210-220, R220-230, R230-240, R240-250, R250-260, R260-270, R270-280, R280-290, R290-300, R300-310, R310-320}
Instep_right/Recio_derecho	NOMINAL	{R160-170, R170-180, R180-190, R190-200, R200-210, R210-220, R220-230, R230-240, R240-250, R250-260, R260-270, R270-280, R280-290, R290-300, R300-310, R310-320}
Length_right/Largo_derecho	NOMINAL	{R160-170, R170-180, R180-190, R190-200, R200-210, R210-220, R220-230, R230-240, R240-250, R250-260, R260-270, R270-280, R280-290, R290-300, R300-310, R310-320}
Width_right/Ancho_derecho	NOMINAL	{R60-65, R65-70, R70-75, R75-80, R80-85, R85-90, R90-95, R95-100, R100-105, R105-110, R110-115, R115-120, R120-125, R125-130}
Toe/Dedo	NOMINAL	{R10-15, R15-20, R20-25, R25-30, R30-35}

En la generación de reglas de asociación se aplico considerando tres niveles tabla 2, esto buscando ir de lo más general a lo particular en relación a los datos.

Tabla 2: relación e niveles y elementos

Nivel	Elementos
Uno	Genero, Grupo, Peso
Dos	Genero, Peso, Diabético_years, Age_Grupo
Tres	Genero, Peso, Diabético_years, Age_Grupo, Recio (ball), Empeine (instep), largo y ancho (length, width) y dedo (toe)

Para la generación de las reglas se utilizo el software Weka, bajo el modelo de reglas de asociación a "priori", las corridas para generar las reglas se realizaron independientemente para cada nivel.

2.4 Método estadístico

La comprobación de significación de estadística se realizo el cálculo comprobatorio de chi-cuadrada a través de los dos métodos Ecuación 4 y 5, identificados como " χ^2 m1" y " χ^2 m2", solo se realizo la comprobación de las reglas generadas previa eliminación de aquellas conceptualmente iguales. En este caso se define dependencia estadísticamente significativa, entre los elementos de las reglas si $\chi^2 > 3.84$.

3. Resultados

3.1 Población

La población total de los individuos analizados se concentro en N=382 correspondientes al 100%, de los cuales 248 individuos fueron del género femenino y 134 masculinos, dentro de estos 139 pertenecieron al grupo control y 243 al grupo diabético, así mismo del grupo control y grupo diabético 39 y 50 respectivamente presentaron peso normal, en cuanto al sobre peso 100 se presentaron en el grupo control y 193 en el grupo diabético, tabla 3.

Tabla 3: Características de individuos de estudio

ATRIBUTO	TOTALES	INDIVIDUOS								
		CONTROL				DIABETICO				
N	382/100	139/36.39				243/63.61				
Años con diabetes	n	1-<5		5-<10		>10				
		108/28.27	52/13.61	83/21.73						
IMC	n	Normal	Sobrepeso	Normal	Sobrepeso	Normal	Sobrepeso	Normal	Sobrepeso	
		39/10.21	100/26.18	16/4.19	92/24.08	12/3.14	40/10.47	22/5.76	61/15.97	
Genero	F	248/64.92	27/7.07	72/18.05	12/3.14	57/14.92	9/2.36	24/6.28	100/26.18	36/9.42
		M	134/35.08	12/3.14	28/7.33	4/1.05	35/9.16	3/0.79	16/4.19	11/2.88

Separación de grupos de estudio (n/%) respecto a la población total

3.2 Reglas de asociación

Primer nivel

El primer nivel constituido por los elementos: Genero, Grupo, Peso, genero tres reglas, la primera y la tercera de ellas presento evidencia

estadísticamente significativa en ambas χ^2 . Una segunda regla, no presento evidencia estadísticamente, tabla 4.

Tabla 4: Reglas primer nivel.

No	Rules	Support A	Support B	Itemset (A∪B)	Support A%	Support B%	Itemset (A∪B)%	Confidence	Lift	χ^2 m1	χ^2 m2
1	Group=Diabetic => Gender=M	243	134	94	0.64	0.35	0.25	0.7	1.11	3.81	4.28*
2	Group=Control Weight=OverWeight => Gender=F	100	248	72	0.26	0.45	0.19	0.72	1.11	2.99	3.03
3	Gender=M Weight=OverWeight => Group=Diabetic	104	243	76	0.27	0.64	0.20	0.73	1.15	5.53*	5.6*

* evidencia estadísticamente significativa $\chi^2 > 3.84$ para denotar dependencia en sus elementos .

Segundo nivel

Este segundo nivel constituido por los elementos: Genero Peso, Diabético_years , Age_Grupo, genero cuatro reglas, en este caso la regla 4 es exactamente igual a la regla 2 del grupo anterior, nuevamente no presentado estadísticamente significativa, es este caso las otras tres reglas restantes si presentaron evidencia estadísticamente significativa en ambas χ^2 , Tabla 5.

Tabla 5: Reglas segundo nivel.

No	Rules	Support A	Support B	Itemset (A∪B)	Support A%	Support B%	Itemset (A∪B)%	Confidence	Lift	χ^2 m1	χ^2 m2
1	Diabetic_years=Diab10 => Age_group=Age50	83	203	69	0.22	0.33	0.28	0.83	1.56	38.3*	37.89*
2	Diabetic_years=Control => Age_group=Age50	139	140	73	0.36	0.37	0.19	0.53	1.43	23.7*	23.45*
3	Diabetic_years=Diab15 => Weight=OverWeight	108	253	92	0.28	0.77	0.24	0.85	1.11	6.06*	5.97*
4	Diabetic_years=Control Weight=OverWeight => Gender=F	100	248	72	0.26	0.45	0.19	0.72	1.11	2.98	3.03

Tercer nivel

El tercer nivel donde ya se presentan las mediciones antropométricas que se constituyo por los elementos: Genero, Peso, Diabético_years , Age_Grupo, ball, instep, length , width and toe, genero trece reglas, en este caso como era de esperarse por la conformación de la información la reglas 1, 2, 7 y 8 se presentaron nuevamente, en este caso de las reglas 8 y 9 no presentaron en el cálculo de ambas χ^2 evidencia estadísticamente significativa, siendo en las restantes 3,4,5 y 6 de ambas χ^2 evidencia estadísticamente significativa, tabla 6.

Tabla 6: Reglas segundo tres.

No	Rules	Support A	Support B	Itemset (A∪B)	Support A%	Support B%	Itemset (A∪B)%	Confidence	Lift	χ^2 m1	χ^2 m2
1	Diabetic_years=Diab10 => Age_group=Age50	83	203	69	0.22	0.33	0.18	0.83	1.56	38.3*	37.89*
2	Diabetic_years=Control => Age_group=Age50	139	140	73	0.36	0.37	0.19	0.53	1.43	23.7*	23.45*
3	instep_right<250 => Toe_right<20.5	90	245	72	0.24	0.64	0.19	0.8	1.25	12.88*	13.08*
4	Diabetic_years=Diab15 Toe_right<20.5 => Weight=OverWeight	79	293	71	0.21	0.77	0.19	0.9	1.17	9.67*	9.58*
5	Diabetic_years=Diab15 => Toe_right<20.5	108	245	79	0.28	0.64	0.21	0.73	1.14	5.52*	5.27*
6	Age_group=Age50 Toe_right<20.5 => Gender=F Weight=OverWeight	135	189	76	0.35	0.49	0.20	0.56	1.14	3.88*	3.98*
7	Diabetic_years=Diab15 => Weight=OverWeight	108	253	92	0.28	0.77	0.24	0.85	1.11	6.06*	5.97*
8	Diabetic_years=Control Weight=OverWeight => Gender=F	100	248	72	0.26	0.45	0.19	0.72	1.11	2.98	3.03
9	Ball_right<146.500 => Weight=OverWeight	84	293	71	0.22	0.77	0.19	0.85	1.1	3.69	3.63

4. Discusiones

El estudio refleja alta influencia del SobrePeso, tanto en el grupo control como en diabético, sin embargo el porcentaje de diabéticos con sobrepeso es mayor al control con sobrepeso, estudios preliminares muestran resultados similares donde el grupo de diabéticos tiene más altos valores de BMI que el grupo control. En todas las reglas generadas, desde la más general conformado por el primer grupo y de manera muy significativa en el grupo final, donde se presenta las mediciones antropométricas, que es el factor sobrepeso es el que contribuye de manera significativa a una modificación en la antropometría del pie, al menos en aquellos en los que aparece.

5. Conclusiones

Los resultados siguen que el efecto de la diabetes no es un factor que influya en la morfología del pie del diabético, acotando esto a diabéticos que no han sufrido trastornos graves por la enfermedad, como es el caso del grupo estudiado.

Por otro lado el sobrepeso si evidencia una dependencia y de manera muy notoria en el dedo gordo y de manera indirecta al empeine. El grupo de trabajo esperaba encontrar evidencia principalmente en el ancho las cuales no se presentaron, sin embargo este estudio muestra evidencias de que el zapato especializado para diabético debe de considerar construcción del calzado basada en características particulares de la población diabética para aumento de su confort y prevención o disminución de problemas asociados al calzado y pie diabético.

La más importante aportación de este estudio viene dada en la aplicación de técnicas de minera de datos en particular en la generación de reglas de asociación, que permitió visualizar los resultados como patrones de asociación y presentarlos de una manera mas grafica en forma de reglas y esto Instituto Tecnológico de Veracruz

soportado por una comprobación estadística como lo es la χ^2 a través de los dos métodos propuestos, que como se mostro dan valores muy acercados, dejando a los investigadores seleccionar el más adecuado a su estudios.

Referencias

- [1] Indranil Manna. "A comparative study of food dimension between adult male and female and evaluation of foot hazards due to using of footwear". Journal physiological anthropology, 20 (4):241-146, 2001.
- [2] Ravindras s. goonetilleke. "Foot anthropometry in Hong Kong". Hong Kong University of science and technology.
- [3] Sergey Brin. "Beyon market baskets, Generalizing association rules to correlations". Department of computer science. Stanford University.