Sistema de Predicción y Clasificación para la Utilización de Recursos Humanos para el Área de Emergencias de un Hospital

Fabricio Echeverría[1], Jennifer Avilés[2], Tatiana Navarro[3], Chrystiam Toapaxi[4]

[1] Magíster en Sistemas de Información Gerencial 2006
[2] Ingeniera en Computación Sistemas Tecnológicos 2006
[3] Ingeniera en Computación Sistemas Tecnológicos 2006
[4] Ingeniero en Computación Sistemas Tecnológicos 2006

Resumen

La calidad deficiente en la atención hacia los pacientes en el área de emergencia de un hospital, así como los recursos humanos limitados que laboran en la misma, son las razones principales para aplicar las técnicas de minería de datos que ayuden a mejorar la toma de decisiones en la industria de la salud.

El Sistema de Predicción y Clasificación (SISPRED) está enfocado en mejorar el funcionamiento del área de emergencias, satisfaciendo las necesidades y expectativas de los pacientes con respecto a los tiempos de atención, lo que se vería reflejado en un servicio de mejor calidad.

Para lo cual, se establece un modelo de predicción que determine la posible concurrencia de pacientes al día siguiente, en base a la información previa obtenida en el diario accionar de una institución, descubriendo así un conjunto de patrones, con lo que podrían establecer predicciones de las posibles emergencias. Además, de desarrollar un modelo de clasificación que permita conformar tres grupos de doctores analizando el nivel de desempeño de cada uno, y por último un modelo de planificación que forme guardias de trabajos óptimas con los grupos previamente formados. Al mismo tiempo se ofrece automatizar el registro de ingreso de pacientes, logrando de esta manera reducir la desorganización ocasionada por la acumulación de documentos en papel.

Abstract

The faulty quality in the attention toward the patients in the emergency's area in a hospital, as well as the limited human resources that work in the same one, are the main reasons to apply the techniques of data mining to help a improve the support of the decision in the care's industry .

The Prediction and Classification System (SISPRED) is focused to improve the functionally of the emergency's area, to satisfy the patients' requirements and expectations, it would be reflected in a service of better quality.

For that reason, it establishes a prediction model that determines the possible patients' concurrence the following day, based on the previous information obtained in the daily job of one institution, discovering a group of patterns, with which could establish predictions for taking decisions of the emergencies that could be happen. Besides developing a classification model that allows conforming three doctors' groups to analyze the performance level of each one, and finally setting up a model of scheduling that makes optimum job's guards with the formed groups previously. At the same time it offers to automate the record of patient, thus reducing the disorganization caused by the accumulation of documents in paper.

1. Introducción

El sector de la salud ha sido objeto de estudios, para tratar la forma en que los servicios hospitalarios se brindan al público. En estos estudios, se aprecia la carencia de un sistema que permita al director del área de emergencia, predecir el número de pacientes al siguiente día y realizar una adecuada planificación de los recursos médicos, a esto se suma el déficit del personal médico y la poca organización del área de emergencia, lo cual se ve reflejada en un servicio deficiente.

Tomamos como parte de nuestro estudio al Hospital Guayaquil Dr. Abel Gilbert Pontón, quien colaboró con el desarrollo de nuestra tesis y proporcionó los datos que serán objeto de las técnicas de minería, en donde pudimos constatar que no se lleva un registro del área de emergencia al igual que en otros hospitales públicos locales, ya que el presupuesto destinado hacia ellos es muy bajo y no les permite contar con insumos básicos para atender las emergencias presentadas diariamente, mucho menos para invertir en un sistema que les ayude en la toma de decisiones para mejorar el rendimiento de los servicios que proveen.

En el área de emergencia, se archiva en papel: el registro de personas que acude, la especialidad en la que son atendidas, el doctor que realiza el chequeo médico u operación. Por esta razón, el tratar de recaudar información para elaborar informes que le sean de utilidad al personal encargado de administrar esta área, les resulta una tarea tediosa que toma gran inversión de tiempo y es propenso a errores de tipo humano.

Otra problemática que hemos tomado en consideración es la dificultad a la cual se enfrenta el director de esta área cada mes, al tener que conformar los grupos eficientes de doctores que atenderán por guardias a los pacientes. Eficiente sería formar grupos médicos estándares con conocimiento y experiencia equilibrados, pero esta situación no siempre sucede, puesto que en la actualidad la conformación de grupos médicos se la realiza sin tomar ninguna consideración en el nivel de preparación o años de experiencia del doctor. El sistema colabora con el director de emergencia creando guardias médicas de trabajo más eficientes y productivas como parte de la estrategia de organización.

Los hospitales están bajo una gran presión por reducir costos, y sin embargo, continuar brindando cada vez mejor calidad, a pesar de contar con un presupuesto bajo. Por esta razón, los hospitales deben buscar herramientas que puedan ser efectivas en el control de recursos. Los pacientes siempre exigen mejores niveles de servicio, no solo, en cuanto al servicio en sí, sino también a la manera en que se lo proporciona.

La innovación ayudará a la gerencia hospitalaria a conseguir una ventaja. La implementación de un sistema que ayude al director del área de emergencia a tomar decisiones en cuanto a la previsión de la demanda hospitalaria, mejorará la utilización de los recursos existentes.

2. Contenido

La solución para la problemática en la gestión de las emergencias hospitalarias la hemos planteado en tres puntos importantes: En la predicción de la cantidad de emergencias esperadas, en la segmentación del personal médico que atiende a los pacientes en grupos de acuerdo a su nivel de desempeño y en la planificación de las guardias médicas. Esto lo realizamos mediante la implementación de un modelo predictivo, un modelo clasificativo y un modelo planificador (algoritmo genético).

MODELO PREDICTIVO

El modelo predictivo está enfocado en estimar o predecir el valor de una variable, a partir de valores conocidos de otras variables. Para nuestro caso particular, el modelo generado se encarga de predecir el número de pacientes que asistirán a solicitar servicios médicos al área de emergencias de una institución hospitalaria. Los parámetros utilizados para el modelo de regresión son: fecha, día de la semana, mes, especialidad y si ese día es feriado.
Para obtener las predicciones para el área de emergencia hospitalaria hemos implementado un algoritmo de Regresión Dicotómica Lineal Múltiple, el cual emplea variables cuantitativas y variables cualitativas. Las variables cualitativas utilizadas para este modelo son:

Variables Cualitativas
Día de la Semana
Mes
Especialidad

Tabla 1 Variables Cualitativas

Cada una de estas variables cualitativas deben ser estandarizadas o codificadas para su posterior utilización en el modelo predictivo, esto se da debido a que el contenido de este tipo de variables generalmente no es el más idóneo y la codificación es necesaria para reducir el numero de variables, optimizando el modelo.

La codificación utilizada para las variables cualitativas es la siguiente:

Especialidad	**d1**	**d2**	**d3**
Medicina General	0	0	1
Cirugía	0	1	0
Pediatría	0	1	1
Ginecología	1	0	0

Tabla 2 Codificación de la variable Especialidad

Día de la Semana	**d1**	**d2**	**d3**
Lunes	0	0	1
Martes	0	1	0
Miércoles	0	1	1
Jueves	1	0	0
Viernes	1	0	1
Sábado	1	1	0
Domingo	1	1	1

Tabla 3 Codificación de la variable Día de la Semana

Mes	**d1**	**d2**	**d3**	**d4**
Enero	0	0	0	1
Febrero	0	0	1	0
Marzo	0	0	1	1
Abril	0	1	0	0
Mayo	0	1	0	1
Junio	0	1	1	0
Julio	0	1	1	1
Agosto	1	0	0	0
Septiembre	1	0	0	1
Octubre	1	0	1	0
Noviembre	1	0	1	1
Diciembre	1	1	0	0

Tabla 4 Codificación de la variable Mes

Luego de la correspondiente estandarización obtenemos el siguiente modelo predictivo para ser aplicado en el sistema SISPRED:

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 m_1 + \beta_5 m_2 + \beta_6 m_3 + \beta_7 m_4 + \beta_8 .feriado + \beta_9 e_1 + \beta_{10} e_2 + \beta_{11} e_3$$

Donde las variables que intervienen para el diseño del modelo de regresión son: d1, d2, d3; correspondientes al día de la semana, m1, m2, m3, m4, correspondientes al mes, feriado que es una variable que nos indica si un día es feriado, y e1, e2, e3, e4; correspondientes a la especialización. Cada una de estas variables: d1, d2, d3, m1, m2, m3, m4, e1, e2, e3, son variables cuyo tipo de dato es binario. Los coeficientes β del modelo serán calculados por medio de la regresión dicotómica lineal múltiple.

El modelo de regresión dicotómica lineal múltiple con p variables predictoras y basado en n observaciones tomadas es de la forma:

$$y_i = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots\ldots \beta_p x_{pi}$$

para i = 1,2,….n. Si escribimos el modelo para cada una de las observaciones, obtendremos ecuaciones lineales de la forma:

$$y_1 = \beta_o + \beta_1 x_{11} + \beta_2 x_{21} + \ldots\ldots \beta_p x_{p1} +$$
$$y_2 = \beta_o + \beta_1 x_{12} + \beta_2 x_{22} + \ldots\ldots \beta_p x_{p2}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_n = \beta_o + \beta_1 x_{1n} + \beta_2 x_{2n} + \ldots\ldots \beta_p x_{pn}$$

Estas ecuaciones lineales pueden ser escritas en forma matricial como:

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot\cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}$$

El vector Y es un vector columna n dimensional que contiene a cada una de las observaciones tomadas para realizar la predicción, X es una matriz que corresponde a cada una de las variables que intervienen en el modelo predictivo y β es el vector de coeficientes de regresión a ser estimados.

De donde los β (estimadores mínimos cuadráticos) se obtienen a partir de la ecuación:

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$$

MODELO CLASIFICATIVO

Los modelos clasificativos permiten detallar la manera en que los datos se distribuyen o agrupan dependiendo de sus características.

El algoritmo de clasificación que utilizamos conocido como k-means o k-medias, identifica un número de grupos, dentro de una gran cantidad de datos donde a simple vista sería muy difícil poder identificarlos. Este modelo encuentra tres niveles o categorías de médicos (ocultos en los datos). Dependiendo del desempeño de los doctores los niveles serán: Alto, Medio y Bajo.

Primeramente se definen el número de grupos a ser calculados por el algoritmo, por lo tanto para nuestra solución serán tres grupos correspondientes a los tres niveles de médicos. Inicialmente, los grupos se generan de manera aleatoria y luego comienza a operar el algoritmo. Se calculan los centroides (promedio entre los elementos que conforman el grupo) y se procede a calcular las distancias euclídeas desde cada elemento a los centroides de cada uno de los grupos, asignando cada elemento al grupo que esté más próximo. Además cuando se asigna un nuevo elemento a un grupo se recalculan las coordenadas del centroide del grupo. El proceso es iterativo y termina cuando se han definido los grupos más homogéneos posibles.

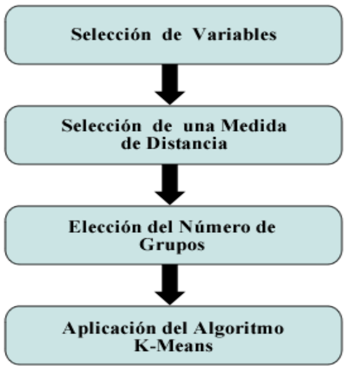

Figura 1 Proceso del Modelo Clasificativo

Para seleccionar las variables sobre las cuales se basan la agrupación, fue necesario realizar investigaciones y entrevistas, logrando identificar las que se consideran de mayor relevancia para calificar al personal médico. La inclusión de una o más variables irrelevantes puede distorsionar una solución de agrupación que de otra forma podría ser útil. En nuestro caso, las variables que intervienen en la aplicación del algoritmo clasificativo son:

Total de Consultas Atendidas: Consiste en el número de atenciones médicas registradas por el doctor. Este valor se actualiza diariamente, a medida que el médico vaya registrando más atenciones a pacientes en la base de datos.

Nivel de Estudio: Indica el nivel académico alcanzado por el médico como: Titulado, Titulado con Postgrado y Titulado con Masterado.

Número de Certificados: Muestra el número de certificados que posee un médico. Esta variable es muy importante puesto que en algunos hospitales públicos el número de seminarios o cursos a los que ha asistido un doctor son de gran relevancia al momento de hacer concursos de méritos, grados de ascenso, designación de cargos, entre otras distinciones.

Años de Experiencia: Indica el tiempo de experiencia que posee el médico.

Luego se necesita alguna medida para evaluar las diferencias y similitudes entre los doctores. La estrategia más común consiste en medir las similitudes en términos de la distancia entre los pares de doctores. Los doctores con distancias reducidas entre ellos son más parecidos entre sí, que aquellos que tienen distancias mayores. La medida de distancia que utilizamos es la euclídea.

$$D_{ij} = \sqrt{\sum_{k=1}^{n}\left(x_{ki} - x_{kj}\right)^2}$$

Dij distancia entre los casos i y j
Xkj valor de la variable Xk para el caso j

MODELO PLANIFICADOR

El propósito del modelo planificador es establecer los días en que los doctores deberán realizar guardia, una tarea que era realizada por el director del área sin ayuda de ningún sistema; la regla del hospital hace énfasis en que un doctor no debería realizar 2 guardias seguidas, para resolver este problema se utiliza un algoritmo genético el cual utilizando los grupos de doctores clasificados conforma la guardia del mes colocando tres doctores (escoge uno de cada clasificación) aleatoriamente en cada guardia formando una planificación base.

En este caso los genes que conforman al cromosoma, esta constituido por el día del mes y un doctor de cada especialidad.

Luego de haber realizado el cruzamiento se produce la mutación, lo cual consiste en alterar un gen del cromosoma, las mutaciones son quienes permiten crear nuevos individuos, escapar de los mininos locales, abastecernos de nuevo material genético, y en ciertos casos mejorar los procesos de elitismo en busca de mejores soluciones.

En los algoritmos genéticos se desarrollan criterios de elitismo (función objetivo) es decir, por medio de penalizaciones para determinar su aptitud; los cromosomas con el menor número de penalizaciones serán los más aptos. Este criterio se aplica con la finalidad de mantener al mejor individuo de cada población y copiarlo a la siguiente de esta manera se asegura mantener en cada nueva población al cromosoma con mejor aptitud.

Pseudo código de la función objetivo:

```
GENES gen_anterior, gen_actual;
gen_anterior = Obtener el primer gen del cromosoma
Inicializar fitness = 0;
Por cada gen en el Cromosoma
Inicio
        gen_actual = Obtener el gen siguiente
        SI(gen_anterior.Doctor1 = gen_actual.Doctor1)
                fitness = fitness - 30;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor2 = gen_actual.Doctor2)
                fitness = fitness - 30;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor3 = gen_actual.Doctor3)
                fitness = fitness - 30;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor1 = gen_actual.Doctor1) Y gen_anterior.Doctor2 =
        gen_actual.Doctor2)
                fitness = fitness - 100;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor1 = gen_actual.Doctor1 Y gen_anterior.Doctor3 =
        gen_actual.Doctor3)
                fitness = fitness - 100;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor2 = gen_actual.Doctor2 Y gen_anterior.Doctor3 =
        gen_actual.Doctor3)
                fitness = fitness - 100;
        SINO
                fitness = fitness + 10;
        SI(gen_anterior.Doctor1 = gen_actual.Doctor1 Y gen_anterior.Doctor2 =
        gen_actual.Doctor2 Y gen_anterior.Doctor3 = gen_actual.Doctor3)
                fitness = fitness -  500;
        SINO
                fitness = fitness - 10;
        gen_anterior = gen_actual;

Fin
```

Para esta planificación en especial se combinan las planificaciones bases para conformar otras nuevas (cruzamiento) y además se reasignan los doctores para un día especifico (mutación), las planificaciones existentes son evaluadas por la función objetivo y se selecciona las mejores para realizar nuevamente el proceso de cruzamiento y mutación hasta que el valor de esta función objetivo converja (se estabilice).

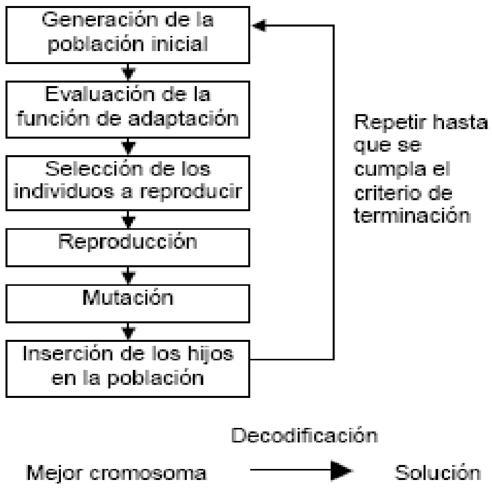

Figura 2 Proceso del Modelo Planificador

3. Conclusiones

Al haber realizado un estudio del problema de gestión hospitalaria en el área de emergencia, se puede plantear algunas conclusiones que se enfocan en la incertidumbre que genera la afluencia de pacientes para los directivos responsables del correcto funcionamiento del área en cuestión. Además, de la determinación de las técnicas y algoritmos de minería de datos utilizados en el proceso de análisis, diseño e implementación del sistema SISPRED, lo cual ha sido muy enriquecedor en cuanto a los conocimientos adquiridos y al desarrollo de la habilidad para encontrar soluciones a problemas que están siendo presentadas a la minería de datos.

- El uso de algoritmos predictivos proporciona diversas ventajas que se resumen en la optimización de los grupos de recursos humanos del área y la reducción de costos, que se verá reflejado en una atención más rápida y efectiva de los pacientes.

- La minería de datos es útil para mejorar la gestión hospitalaria, ayudando a los responsables de la planificación del área de emergencias a tener una visión mucho más clara y profunda para la toma de decisiones acerca de la organización del personal médico.

- Lo indispensable en el modelo planificador es la optimización en la conformación de las guardias médicas, en base a los grupos de médicos ya clasificados, lo cual prestará mejor atención en el área.

4. Referencias

1. Hair, J.F.; Anderson, R.E.; Tatham, R.L. and Black, Análisis Multivariante, Prentice-Hall, Madrid, 1999.
2. F. Tusell, Análisis Multivariante, EDITORIAL, 2005.
3. VARELA RODRIGO, Innovación empresarial, Prentice Hall, 2001.
4. Luis Carlos Silva Ayaguer, Ed. Díaz de Santos, Excursión a la regresión logística en ciencias de la salud, Madrid, 1995.
5. David W. Hosmer, Stanley Lemeshow, Ed.John Wiley, Applied Logistic Regression, New York, 1989.
6. Víctor Abraira Santos, Alberto Pérez de Vargas Luque, Métodos Multivariantes en bioestadística, Ed. Centro de Estudios Ramón Areces, Madrid, 1996.
7. MONTGOMERY, D. C. y RUNGER G. C., Probabilidad y Estadística Aplicadas a la Ingeniería, Mc Graw-Hill, México, 1996.
8. José Manuel Molina López y Jesús García Herrero, Técnicas de Análisis de Datos, Universidad Carlos III, Madrid, 2004.
9. Los datos fuentes fueron tomados Hospital Guayaquil Dr. Abel Gilbert Pontón, basados en la Ley de Transferencia Pública.