

Sistema de Clasificación de los Cuestionarios Llenados en el Proceso de Encuesta, que se Realizan para la Evaluación del Profesor Mediante el Uso de Redes Neuronales

D. Carrera, J. Palomeque, F. Echeverría

Escuela Superior Politécnica del Litoral Campus "Gustavo Galindo V." Km. 30.5, Vía Perimetral, Guayaquil, Ecuador

dcarrera@espol.edu.ec, jpalomeq@espol.edu.ec, pechever@espol.edu.ec

Resumen

La necesidad de discriminación de formularios válidos que se han evaluado en un cuestionario a través del internet, y la necesidad de automatizar este proceso, ha dado apertura para aplicar técnicas de minería de datos que comprende la rama de la Inteligencia Artificial, detectando comportamientos anormales que son difíciles de obtener por medio de métodos manuales. Esta técnica abarca las diferentes teorías que se formaron en la evaluación docente sobre los formularios que se consideran no válidos, y adicionalmente la comprobación de la efectividad de la red neuronal artificial aplicada en el trabajo, mediante dos algoritmos de complejidad más sencilla.

Palabras Claves: *Inteligencia Artificial, Discriminación de formularios, Redes Neuronales Artificiales.*

Abstract

The need of differentiation between valid and not valid forms that are the result of an Internet survey, and the need to automate this process, has given the aperture to apply new data mining techniques that are included in field of Artificial Intelligence, detecting abnormal behaviors that are difficult to obtain by manual methods. The technique presented in this work will include different theories that have been used in the educational evaluation, about forms that are considered valid. In addition, we will verify the effectiveness of the applied artificial neuronal network, by comparing it with two simpler algorithms.

Keywords: *Artificial intelligence, Discrimination of forms, Artificial Neuronal Networks.*

1. Introducción

Durante muchos años se conoce que la recolección de datos conlleva errores ajenos al muestreo que distorsionan los resultados reales, estos errores contienen datos aberrantes que no siempre son de fácil discriminación para el experto. Actualmente uno de los mejores métodos para la discriminación de formularios no válidos, se aplica en el momento de realizar la encuesta a un usuario. Sin embargo, las tendencias tecnológicas, los bajos costos operativos y calidad similar de resultados en la recolección de datos en formato digital, no sería posible aplicar este método.

Junto a estos cambios de tendencias tecnológicas y la tradicional demanda de los profesores con respecto a las suposiciones de diferentes teorías que definen "¿cuáles deberían ser los formularios que son evaluados de manera incorrecta?" para que no sean considerados en el cálculo de indicadores de la evaluación docente, determinan una necesidad de aplicación de diferentes técnicas de minería de datos que permitan detectar estos formularios válidos y no válidos.

Entre estas técnicas de minería de datos, hemos desarrollado una red neuronal para la detección de patrones y la determinación de formularios válidos y no válidos, adicionalmente se desarrollo dos algoritmos de discriminación que utiliza la distancia de Mahalanobis, y la distancia Euclidiana.

2. Metodología

Dentro de la rama de las redes neuronales artificiales [1] podemos encontrar varios tipos de redes [2], para la cual su selección, debe ser basada en los requerimientos del usuario y su aplicación.

De entre los requerimientos del usuario podemos mencionar:

La automatización de la clasificación de formularios.

El proceso de clasificación debe realizarse en lo posible sin ningún ente humano.

Determinar los formularios necesarios para el entrenamiento de la Red Neuronal Artificial (RNA).

Recibido: Junio, 2007

Aceptado: Agosto, 2007

Realizar el proceso discriminación al finalizar el proceso de encuesta de una evaluación docente.

Clasificar un cuestionario de N preguntas.

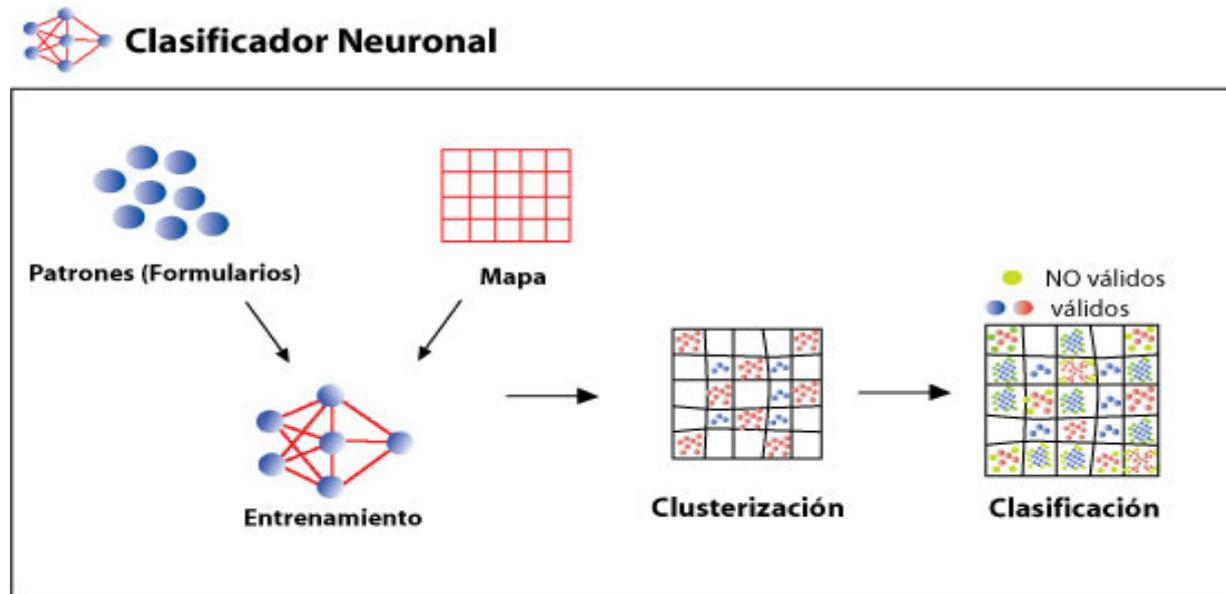


Figura 1. Cuadro General de los formularios válidos y no válidos, en la Red Neuronal Artificial.

Existiendo otros factores de importancia para ser considerados, como son:

- a falta de conocimiento de los formularios válidos o no válidos por un humano.
- a subjetividad de validación de los formularios.
- las diferentes teorías que existen sobre los formularios válidos en la evaluación docente.

Entre las diferentes tipos de redes neuronales artificiales y basadas en los requerimientos del usuario podemos definir dos características principales de la red que son de tipo offline, y de aprendizaje no supervisado [3]. Seleccionando el tipo de red ideal que cumple con las características mencionadas para este trabajo se ha determinado los Mapas Auto-organizativos de Kohonen (SOM).

Este tipo de RNA fue desarrollada originalmente por T. Kohonen en 1982, se ajusta a los requerimientos del usuario ya que como su nombre lo menciona, auto-organiza los patrones (formularios) que ingresan a la red, durante el proceso de entrenamiento de la red, permitiendo crear un hiperplano, cuya característica principal es la de reducir problemas de dimensiones mayores y llevarlas a dos dimensiones manteniendo sus propiedades [4].

Como se puede observar en la figura 1, los patrones deben ser primeramente entrenados para luego realizar una clasificación, considerando que se deberán tomar en cuenta ciertas configuraciones para su funcionamiento óptimo, entre las cuales mencionaremos:

Una función de adyacencia.- Es el grado de medida comparativa en entre una neurona con sus vecinas, cuando es del tipo Gaussian^L presenta un aprendizaje en forma de sombrero mexicano como se lo conoce a este estilo, y obteniendo una onda descendente desde la neurona ganadora hacia sus vecinas, por ende se observa así un aprendizaje blando.

Un factor de aprendizaje.- valor que comprende valores entre 0 y 1, y que se aplican según se realice las pruebas.

Capas de una red.- esta característica muchas veces se define dependiente del tipo de RNA utilizada, para el caso de las redes tipo Kohonen, se construyen dos capas, una capa de entrada y una segunda de salida.

Tamaño de la red.- comprende el número de neuronas y sus diferentes capas que serán utilizados en la red neuronal para resolver el problema que se plantee.

Número de patrones de ingreso.- este valor depende del tamaño de la población, también se lo

denomina tamaño de la muestra. El rango de obtención de muestra se recomienda entre (500, 10000) para poblaciones mayor a 500, y en un 75% de la misma para tamaño de muestra menor a 500.

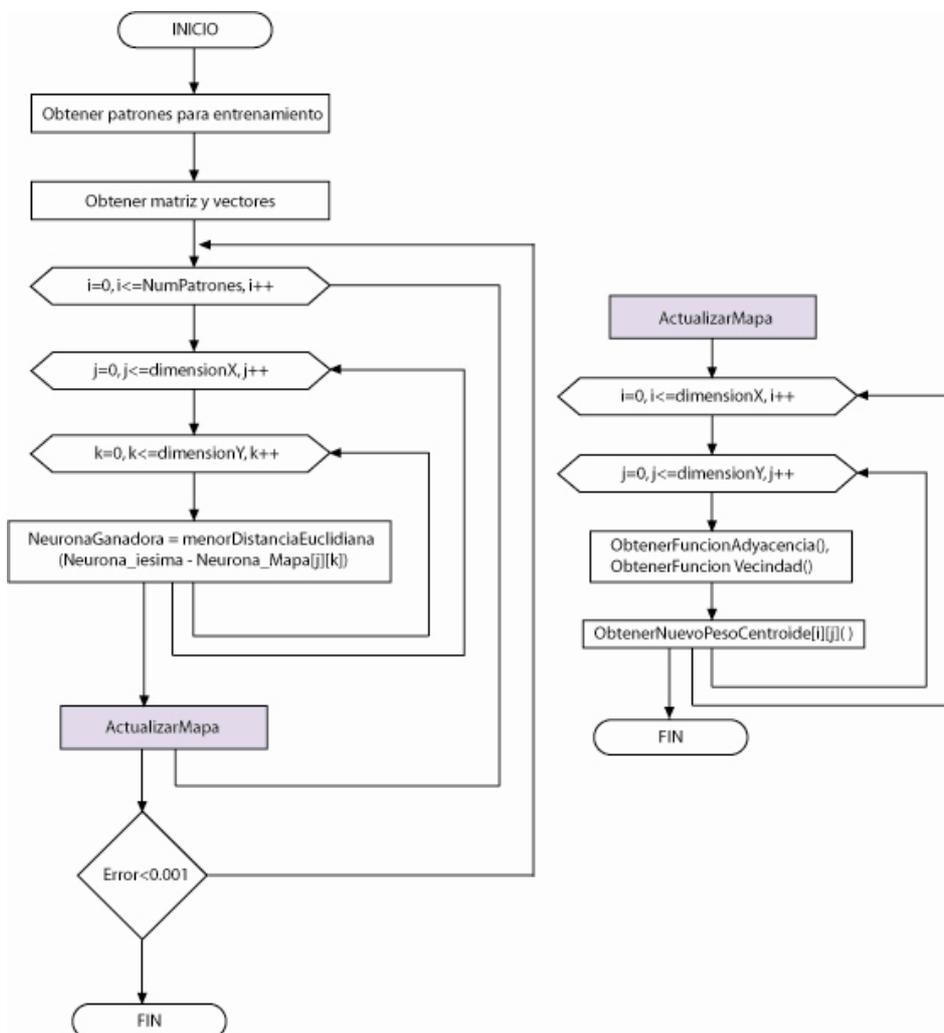


Figura 2. Algoritmo de Entrenamiento de la Red Neuronal Artificial

A continuación detallamos la fórmula de entrenamiento, acorde a lo antes mencionado [5]:

$$m_i(t + 1) = m_i(t) + h_{ci}(t) * [x(t) - m_i(t)]$$

Donde $h_{ci}(t)$ es la función de vecindad, que es igual a

$$h_{ci}(t) = \alpha(t) * \exp\left(-\frac{|r_c - r_i|^2}{2\sigma^2(t)}\right)$$

$\sigma(t)$ Está definido por el ancho de la dimensión del mapa

Donde $\alpha(t)$ es el factor de aprendizaje, y su rango se encuentra entre $0 < \alpha(t) < 1$.

$$\alpha(t) = K * \left(\frac{1-t}{N}\right)$$

Para t = iteración, y N = Número de patrones (formularios), K = coeficiente de aprendizaje

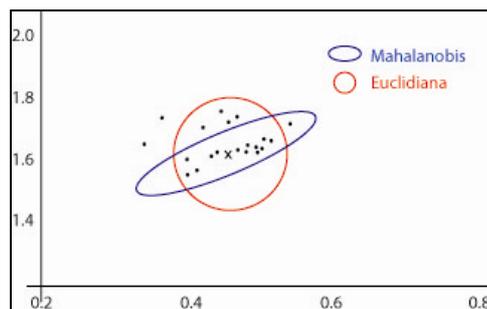


Figura 3. Gráfica de distancia Euclidiana vs distancia Mahalanobis

Donde $\sigma(t)$ y $\alpha(t)$ son funciones mono tónicas decrecientes.

En la figura 2 se detalla el algoritmo utilizado para el proceso de entrenamiento de la red neuronal mencionada [6].

En la figura 3, se puede apreciar la diferencia en la discriminación de los formularios válidos y no válidos, según la distancia utilizada [8].

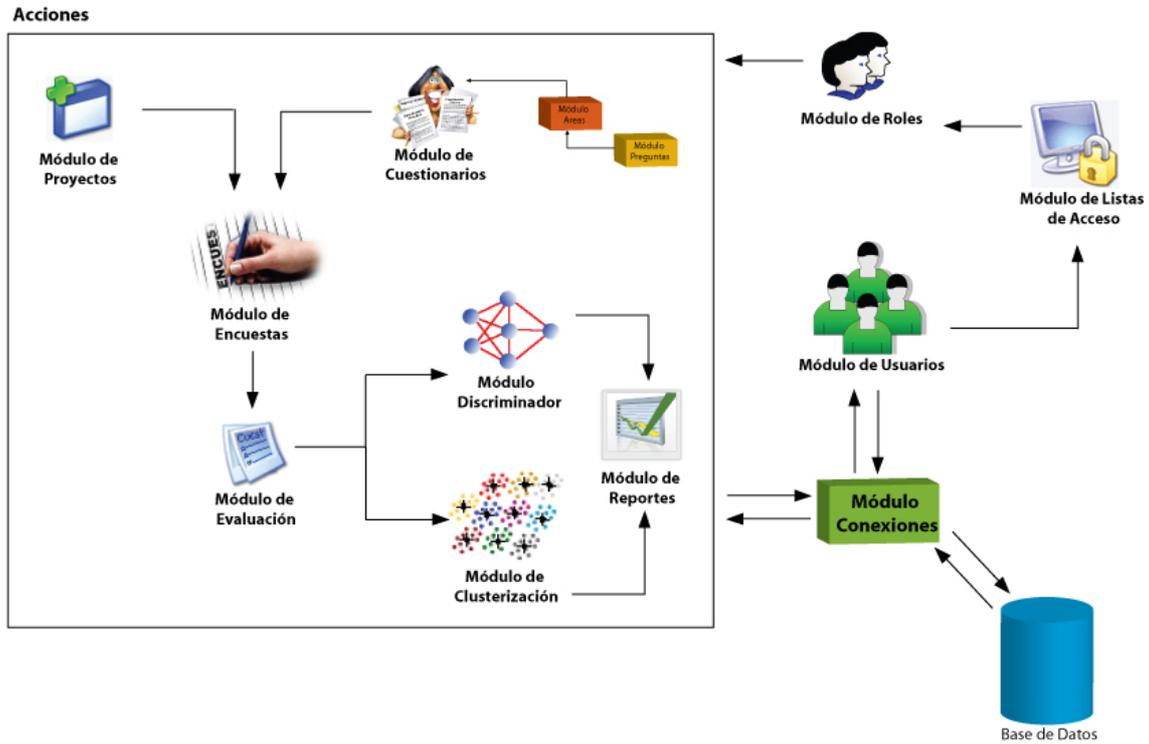


Figura 4. Gráfica que el sistema CENACAD, utilizando la discriminación de formularios

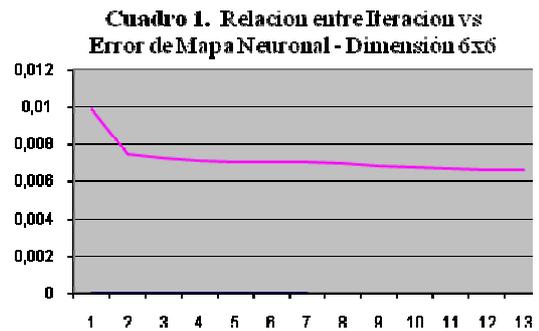
El módulo de Reportes de los formularios discriminados se la realiza mediante una interfaz web (Ver Figura 4), para el usuario administrador, permitiendo presentar informes con datos previamente filtrados y optimizados [7].

3. Evaluación

Durante el proceso de entrenamiento, se ha aplicado una función de error equivalente al promedio de las diferencias entre las distancias del formulario patrón vs la neurona ganadora.

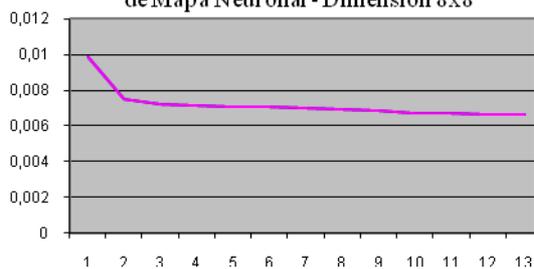
$$\text{Error} = \text{SumaError} / \text{NumPatrones}$$

Tomando como referencia la fórmula antes mencionada, hemos podido encontrar los siguientes cuadros:



*Tiempo de cálculo: 3 días

Cuadro 2. Relacion entre Iteracion vs Error de Mapa Neuronal - Dimensión 8x8

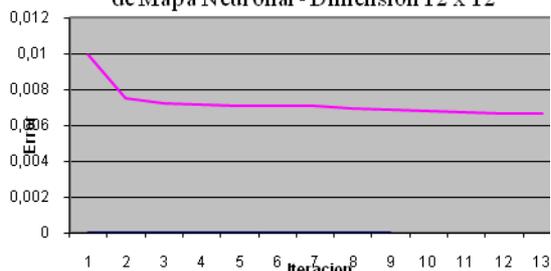


*Tiempo de cálculo: 5 días

En donde se puede representar que para un mapa de dimensión 8x8 y un mapa de dimensión de 6x6, el error converge para valores de 0.005, sin embargo para un mapa de dimensión 12x12, este error converge en 0.0069.

Una vez obtenido estos resultados y acorde a la teoría mencionada sobre la importancia de la dimensionalidad de los mapas, se ha establecido utilizar un mapa de dimensión = 8x8

Cuadro 3. Relacion entre Iteracion vs Error de Mapa Neuronal - Dimensión 12 x 12



*Tiempo de cálculo: 6 días

RESUMEN				
Método discriminante	Paralelos	Suma	Promedio ESPOL	Varianza
Neuronal	1208	103000.229	85.336	103.443
Mahalanobis	1208	103950.869	86.195	75.390
Regresión	1208	104614.708	86.745	67.815
Euclidiano	1208	107655.917	89.267	31.252

Tabla 1. Métodos y sus varianzas

Otro método de evaluación de la red neurona, ha sido realizando una prueba estadística, en la cual se toma como hipótesis que “La calificación del profesor no varía según el método discriminante”, llegando a obtener los valores presentados en la Tabla 1, que determina una mayor varianza en los formularios discriminados por el método neuronal.

Una vez realizado un análisis de varianza de un factor, con una confianza del 95% para una

población N=1208 paralelos evaluados en el 2006 - segundo término, en las materias teóricas, se encuentra como resultado, que sí existe una diferencia significativa con respecto a los métodos (Ver tabla 2), siendo el de menor varianza y mayor promedio el algoritmo euclidiano, de mayor varianza y menor promedio el método neuronal.

ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	10337.59	3	3445.86	49.59	0.00	2.61
Dentro de los grupos	334972.87	4821	69.48			
Total	345310.46	4824				

Tabla 2. Análisis de Varianza

La tabla 2 concluye que se rechaza la hipótesis nula a favor de la alternativa, puesto que el valor de significancia (probabilidad) que muestra la tabla (0.000) es menor a 0.05, es decir que si existe variación en la calificación de los profesores, siendo el método de mayor aceptación el neuronal al ser el de mayor varianza, así el sesgo con respecto a las evaluaciones será menor. [9].

La figura 5, un mapa de dimensión 8x8 y sus respectivas neuronas, incluyen el número de patrones (formularios) que han sido agrupado por su comportamiento en cada neurona, adicionalmente se indica cual es el número de formularios aproximados que serán discriminados. Por ejemplo: en la neurona con id=342, tiene 37 de 41 patrones válidos.

4. Resultados

Discriminador Neuronal

Encuesta "Encuesta de Materias Teóricas 2006 - 25"
Cuestionario " CUESTIONARIO MAT TEORICAS 2006 "

Formularios evaluados: 34488
Formularios válidos: 32900
Formularios NO válidos:1517
Dimensión del mapa : 8 x 8

Calcula validos

8 de 8 (341)	43 de 47 (333)	7 de 7 (325)	774 de 860 (349)	1166 de 1295 (357)	252 de 280 (365)	24 de 26 (373)	96 de 106 (381)
6 de 6 (382)	37 de 41 (342)	27 de 30 (366)	1238 de 1375 (374)	27 de 29 (326)	18 de 20 (350)	67 de 74 (334)	14 de 15 (358)
69 de 76 (375)	2 de 2 (359)	638 de 708 (383)	0 de 0 (343)	33 de 36 (367)	15 de 16 (351)	137 de 152 (327)	2 de 2 (335)
23 de 25 (376)	14 de 15 (336)	12 de 13 (328)	18 de 19 (352)	8 de 8 (344)	149 de 165 (360)	33 de 36 (368)	3 de 3 (384)
9 de 10 (353)	10 de 11 (377)	31 de 34 (345)	40 de 44 (329)	110 de 122 (385)	21927 de 24363 (361)	62 de 68 (369)	41 de 45 (337)
49 de 54 (354)	7 de 7 (338)	93 de 103 (346)	3 de 3 (330)	159 de 176 (378)	145 de 161 (370)	345 de 383 (362)	42 de 46 (386)
20 de 22 (347)	465 de 516 (339)	90 de 99 (355)	37 de 41 (363)	8 de 8 (371)	5 de 5 (387)	62 de 68 (379)	134 de 148 (331)
227 de 252 (380)	1874 de 2082 (372)	9 de 10 (364)	81 de 90 (356)	0 de 0 (348)	4 de 4 (340)	8 de 8 (332)	9 de 10 (388)

Figura 5. Gráfica que el sistema CENACAD muestra la matriz de neuronas con sus respectivos porcentajes de formularios discriminados

La figura 6, detalla los puntajes que tiene cada patrón obtenido en la evaluación docente y que han sido agrupados en la neurona #328, en un cuestionario que tiene 33 preguntas utilizado para

evaluación de profesores que dictan materias clasificados como "teóricos" acorde a la reglamentación de la universidad.

Formularios de la Neurona # 328
 Encuesta " Encuesta de Materias Teóricas 2006 - 2S "
 Cuestionario " CUESTIONARIO MAT TEORICAS 2006 "

id_patron	distancia	Preg #1	Preg #2	Preg #3	Preg #4	Preg #5	Preg #6	Preg #7	Preg #8	Preg #9	Preg #10	Preg #11	Preg #12	Preg #13	Preg #14	Preg #15	Preg #16	Preg #17	Preg #18	Preg #19	Preg #20	Preg #21	Preg #22	Preg #23	Preg #24	Preg #25	Preg #26	Preg #27	Preg #28	Preg #29	Preg #30	Preg #31	Preg #32	Preg #33
314641	2,4643	3	3	2	2	1	4	3	3	4	3	4	3	4	3	4	3	2	3	4	3	3	4	3	3	4	3	3	3	3	3	4	3	3
333746	2,6306	2	3	3	1	1	4	4	4	4	2	2	4	4	2	2	2	2	3	3	3	3	3	4	3	3	3	3	3	3	3	3	3	2
330132	2,6574	3	3	1	1	1	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1
324199	2,6655	2	2	2	1	1	4	4	4	4	3	4	4	4	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	3
321495	2,9993	3	3	1	1	1	5	4	5	4	3	4	3	4	3	4	3	4	3	4	4	3	4	4	3	4	4	3	4	3	4	3	4	3
325534	3,0624	2	2	1	1	1	3	4	4	2	3	2	3	4	3	3	3	3	3	3	5	2	2	3	3	3	2	2	2	3	3	4	5	2
322540	3,5432	3	3	3	1	2	4	5	4	4	3	3	4	4	5	3	5	3	3	4	4	4	3	3	3	3	4	3	3	3	5	5	1	4
308862	3,6685	2	3	2	1	1	5	4	4	4	4	5	5	4	4	3	2	1	3	3	5	4	4	4	3	3	4	5	4	4	4	5	4	3
313805	3,8027	2	2	1	1	1	5	4	5	5	2	2	2	3	3	4	4	3	3	5	4	4	3	4	4	4	4	1	1	5	3	2	3	2
308217	3,9693	3	3	3	1	1	4	5	4	4	4	5	5	5	5	4	4	4	4	5	3	1	2	3	4	4	1	3	4	4	2	2	3	
326002	4,2476	3	3	3	1	2	5	5	5	3	3	5	5	5	3	3	2	4	3	5	3	3	4	4	4	4	3	2	4	3	5	5	3	
328679	4,8967	3	3	3	2	1	5	4	4	4	4	5	5	5	4	4	4	4	4	4	4	3	4	4	4	5	5	4	3	5	3	1	1	3
318630	5,7020	2	3	3	2	2	5	3	5	5	2	5	5	5	5	5	5	5	2	5	5	2	3	1	4	5	5	2	4	5	5	5	5	2

Figura 6. Gráfica que detalla los patrones de una neurona, con sus respectivos formularios válidos y no válidos.

5. Conclusiones

a. Un mayor número de patrones de entrenamiento dará como resultado un mayor tiempo para el proceso.

d. Permite enfocar de manera más precisa los diferentes comportamientos sobre la evaluación docente.

e. El período de entrenamiento de una red puede tomar varios días para finalizar el mismo.

f. En el modelo de Kohonen puede obtener un discriminador para cualquier tipo de encuesta en donde todas sus preguntas utilizan escala Likert en alternativas.

g. Obtener un porcentaje de error relativamente bajo, en comparación a otros métodos.

i. Ayudar a determinar los formularios válidos y no válidos en los diferentes comportamientos que pueda obtener durante una encuesta.

j. Permite entregar resultados concretos a teorías inciertas como por ejemplo: Si un estudiante que pierde una materia, contesta el formulario de diferente manera a los demás estudiantes.

6. Referencias

[1]. **Palacios Burgos, Francisco José.** Redes Neuronales con GNU/Linux. *Herramientas en GNU/Linux para estudiantes universitarios.* [En línea] 12 de Agosto de 2003. [Citado el: 18 de Diciembre de 2006]. http://es.tldp.org/Presentaciones/200304curso-glisa/redes_neuronales/curso-glisa-redes_neuronales.html/.

[2]. **Hilera, José R y Martínez, Víctor J.** *Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones.* [ed.] ALFAOMEGA S.A. Madrid : RA-MA, 2000.

[3]. **Wikipedia.** Wikipedia. *Wikipedia.* [En línea] 2 de Diciembre de 2005. [Citado el: 23 de Agosto de 2006.] http://en.wikipedia.org/wiki/Supervised_learning.

[4]. **Palmer, A, Montaño, J J y Jiménez, R.** Tutorial sobre Redes Neuronales Artificiales: Los Mapas Autoorganizados de Kohonen. *Psicologia.com.* [En línea] 1 de Enero de 2002. [Citado el: 15 de Enero de 2006.] <http://www.bibliopsiquis.com/psicologiacom/vol6num1/3301/#entren32>.

[5]. **Zuluaga A., Camilo M., Acosta B., María Isabel y Salazar I., Harold.** Tutorial de Redes Neuronales. *Tutorial de Redes Neuronales.* [En línea] 30 de Enero de 2000. [Citado el: 30 de Abril de 2005]. <http://ohm.utp.edu.co/neuronales/>.

[6]. **Kohonen Teuvo.** Self Organization Map. [ed.] ISBN S.A. Madrid : RA-MA, 2001.

[7]. **Estévez Valencia, Ph.D Valencia y Achermann Undurraga, Cristóbal.** Visualización de datos mediante el mapa auto-organizativo de kohonen y análisis de componentes. *visualización de datos mediante el mapa auto-organizativo de kohonen y análisis de componentes.* [En línea] 25 de Abril de 1998. [Citado el: 20 de Noviembre de 2006.] <http://www2.ing.puc.cl/~iing/ed431/anales/anal02.htm>.

[8]. **Wikipedia.** Mahalanobis. *Mahalanobis.* [En línea] 15 de Enero de 2006. [Citado el: 10 de Agosto de 2006.] http://es.wikipedia.org/wiki/Distancia_de_Mahalanobis.

[9] **Freund, Miller, Miller.** *Estadística matemática con aplicaciones. Sexta Edición.* Prentice Hall