Sistema Inteligente de Reconocimiento de Texto Natural Manuscrito

Alumno1, Alumno2, Alumno3…, director1 de TT, director2de TT
Escuela Superior de Cómputo I.P.N. México D.F.
Tel. 57-29-6000 ext. 52000 y 52021. E-mail: alumno1@ipn.mx, alumno2@ipn.mx, alumno3@ipn.mx

Resumen — **El trabajo terminal presentado consiste en el desarrollo de un sistema de cómputo que permite el reconocimiento de texto manuscrito basado en técnicas de procesamiento de imágenes, inteligencia artificial, redes neuronales artificiales, reconocimiento de patrones, sistemas de reconocimiento de texto manuscrito natural, e ingeniería de software, se logra el reconocimiento de la escritura manuscrita con una eficiencia superior al 80%.**

Palabras clave — **Reconocimiento de patrones, redes neuronales, texto natural, procesamiento de imágenes,**

I. Introducción

En los sistemas de reconocimiento de caracteres tiene lugar el proceso de conversión de imágenes de texto impreso o texto manuscrito (números, letras y símbolos) a un formato de fácil procesamiento por la computadora (tal como el código ASCII). Los sistemas de reconocimiento se pueden clasificar como Sistemas de Reconocimiento Óptico de Caracteres (OCR, por sus siglas en inglés), son caracterizados por el reconocimiento de solamente caracteres impresos. Los Sistemas de Reconocimiento Inteligente de Caracteres (ICR, por sus siglas en inglés), realizan el reconocimiento de caracteres manuscritos separados y el reconocimiento OCR, sin embargo tiene ciertas limitaciones cuando se encuentra algún tipo de manuscrito, como el escrito natural cursivo. Los sistemas de Reconocimiento de Manuscrito Natural (NHR, por sus siglas en inglés), permiten a las computadoras leer y reconocer manuscritos con un alto grado de exactitud. Estos últimos son hoy una de las nuevas tecnologías que se han desarrollado, logrando un mejor impacto sobre la lectura de documentos.

La diferencia con los sistemas OCR e ICR radica en que NHR utiliza una serie de algoritmos complejos para comparar y reconocer cada uno de los caracteres. Los sistemas de reconocimiento pueden ser divididos en dos categorías principales: el reconocimiento dinámico y el reconocimiento estático. En el reconocimiento dinámico, también conocido como reconocimiento en línea, las formas o caracteres son identificados en el momento en que se escriben. La información es dinámica y unidimensional, consiste de una secuencia de trazos representados en una escala temporal, el problema en este tipo de reconocimiento esta dado en el tiempo de ordenamiento de los trazos proporcionados por la interfaz con la computadora.

Uno de los medios utilizados para la introducción de los datos en forma dinámica son los lapiceros electrónicos que actúan sobre una tarjeta digital encargada de transferir la información a la computadora. En el reconocimiento estático o reconocimiento fuera de línea, el proceso de escritura es desconocido. La información no tiene una naturaleza temporal, es bidimensional y consiste de una imagen de caracteres digitalizada. La imagen puede ser proporcionada por una cámara de video, un scanner o algún otro dispositivo de captura de imágenes. La diversidad de paquetes computacionales que se encuentran disponibles para llevar acabo la automatización de documentos, utilizando tecnologías OCR, ICR o NHR, es amplia, y se han desarrollado sistemas OCR que en la actualidad logran tazas de reconocimiento del 99.99%, cuando los documentos no tienen diferentes tipos de letras, la distribución del texto es uniforme sobre el documento, la calidad del papel a reconocer es bueno y la imagen proporcionada tiene una buena resolución.

Por otro lado, aún cumpliendo con las mejores condiciones requeridas para llevar acabo el reconocimiento, los sistemas ICR y NHR logran en ocasiones tazas de reconocimiento menores del 87%. La taza de reconocimiento en estos últimos, se ve disminuida debido a la gran variedad de manuscritos que se realizan, cada persona tiene rasgos especiales en su forma de escribir, el texto no es uniforme y existen caracteres sobrepuestos.

II. Metodología

El sistema inteligente de reconocimiento texto natural manuscrito está constituido por cuatro módulos principalmente; adquisición de la imagen, tratamiento de imagen, reconocimiento e interpretación de imagen y editor de texto. Básicamente, el *bloque de adquisición de imagen* está formado por un escáner, o un pad electrónico o una cámara digital, que funcionan como una interfaz a través del cual la imagen es adquirida para el posterior análisis efectuado por el sistema de cómputo. El *bloque de tratamiento de la imagen* consiste de seis procesos por los que pasa la imagen: umbralización, eliminación de información no necesaria (ruido), detección del área del caracter, segmentación del caracter, conversión de la imagen a ceros y unos, y por último la compresión de la imagen a una matriz propuesta de 10x10 pixeles, que será la entrada a un conjunto de redes neuronales artificiales.

El *bloque del reconocimiento de los caracteres*, se planteó resolverse por medio de redes neuronales artificiales, por medio de un perceptrón multicapa entrenado por el algoritmo de retropropagación [1] [2].

Por último se ubica el *bloque de editor de texto*, el cual se orienta hacia una aplicación específica que es la adquisición de información manuscrita de encuestas a través de un formato previamente establecido para grandes volúmenes de información.

A. Tratamiento Digital de la Imagen.

En este módulo se llevan a cabo los siguientes seis procesos [3] [4]:

1) *Características del Documento a Reconocer*: En el documento a reconocer se establecieron ciertas condiciones iniciales tales como el fondo de color blanco, los caracteres de color negro y guías, esto es necesario para que la imagen se pueda tratar y reconocer dichos caracteres (ver Fig. 1).

2) *Conversión a niveles de gris:* El tamaño de la paleta depende del número de bits por pixel de éste. El valor del pixel sirve como un índice dentro de la paleta para una imagen con escala de grises los valores de RGB serán iguales y servirán como una intensidad (Tabla 1).

3) *Binarizado:* El problema principal del Binarizado consiste en encontrar un umbral adecuado para minimizar la pérdida de información. En el Método de Otsu, un análisis del histograma, permite determinar dinámicamente el umbral óptimo para la imagen. Aplicándose el método de selección automática basado en el análisis del histograma, en este algoritmo, se considera al histograma como una función de densidad de una dimensión, y para cada nivel de gris (0 a 255) se determina el valor de la función, dado por (1) [5].

TABLA I
PALETA DE COLORES EN NIVELS DE GRIS;
N ES UN NÚMERO NEUTRO UTILIZADO POR EL FORMATO

R	G	B	N
0	0	0	0
1	1	1	0
2	2	2	0
.	.	.	.
.	.	.	.
.	.	.	.
255	255	255	0

Fig. 1. Ejemplo de un formato.

$$P_i = \frac{n_i}{N_T} \qquad i = 0,1,2...,255 \tag{1}$$

en donde:

n_i = número de pixeles en la imagen con nivel *i*,

N_T = número total de pixeles en la imagen.

El momento de orden uno (media) para todo el histograma está dado por (2).

$$ml = \mu_T = \sum_{i=1}^{L}(i)(P_i) = \sum_{i=1}^{L}(i)\left(\frac{n_i}{N_i}\right). \tag{2}$$

Al binarizar, se divide en 2 clases C0 y C1 con niveles $0 \rightarrow$ k y K + $1 \rightarrow$ L, se obtienen 2 momentos acumulativos, ver (3) y (4).

$$\omega(k) = \sum_{i=1}^{L} P_i \tag{3}$$

$$\mu(k) = \sum_{i=1}^{L}(i)(P_i) \tag{4}$$

que representan la probabilidad de la ocurrencia de la clase y la media de la clase respectivamente [6].

El umbral óptimo seleccionado es aquel que maximiza la separabilidad entre las clases dada por (5).

$$\sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \quad (5)$$

4) *Eliminación de ruido en la imagen*: El ruido se puede considerar como una función f(x, y), sumada o mezclada con una imagen i(x, y) de tal manera que se puede ver el resultado como una función g(x, y); esto es:

$$g(x, y) = f(x, y) + i(x, y) \quad (6)$$

Haciéndose la hipótesis de que en cada par coordenado (x,y) el ruido es una función sin correlación y tiene un valor medio de cero. Esto nos arroja la siguiente suposición, que al poder inferir que el ruido es una función que se puede encontrar su ecuación y conocer su comportamiento, esto a su vez deriva en que es posible, no sólo quitarlo sino también introducirlo.

Uno de los tipos de ruido que nos interesa eliminar, es aquel en donde aparecen pequeños objetos (manchas), las cuales aparecen aleatoriamente en toda la imagen; estas manchas tienen la característica de que varían en tamaño, densidad y forma [6] [7].

5) *Segmentación de caracteres*: Consiste en recorrer toda la imagen de izquierda a derecha y de arriba hacia abajo, de tal manera que por cada vez que se encuentre un punto o pixel en la imagen con nivel de gris 0 (cero), se considere para ser parte de un objeto [3] [7].

Al hallar el pixel negro lo que se hace es, etiquetarlo con un color o nivel de gris distinto a 0 ó 255 (negro o blanco), y tomándolo como centro se busca en sus vecinos un pixel de color negro. Esta búsqueda se hace en sentido de las manecillas del reloj y comenzando desde la posición 0 (cero), como lo muestra la Fig. 2.

Sí se encuentra un píxel negro durante la búsqueda entonces se etiqueta con el mismo nivel de gris que el pixel central, y ahora ese nuevo pixel etiquetado se vuelve el pixel central.

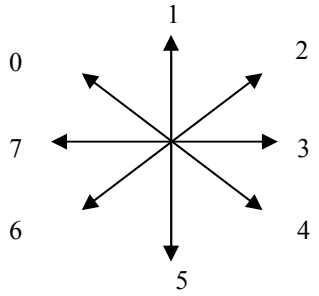


Fig. 2. Sentido de la búsqueda

6) *Escalamiento:* Este proceso consiste en transformar la serie de objetos (caracteres) a un tamaño estándar, el cual se propuso de 10x10 para cada uno de los objetos. El algoritmo para realizar el escalamiento es el siguiente:

Primero se deben contabilizar los pixeles en la imagen como se muestra en la Fig. 3.

Una vez que ya se tiene determinado la cantidad de pixeles negros y blancos, lo que sigue es aplicar (7):

$$\text{No. de Pixeles en } x = \frac{(\text{No. de Pixeles en } X)(\text{No. de Total de Pixeles en } x)}{\text{No. de Total de Pixeles en } X} \quad (7)$$

Donde:

No. de pixeles en x es la cantidad de pixeles a pintar en el eje *x* en la nueva imagen.

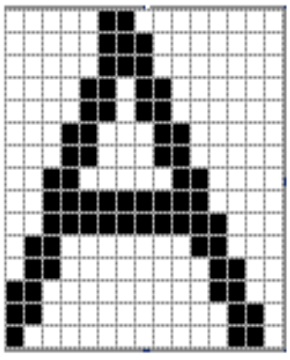
No. de pixeles en X, es la cantidad de pixeles en la imagen original en el eje *x*.

Total de pixeles en x, es la cantidad total de pixeles en el eje *x* en la nueva imagen.

No. Total de pixeles en X, es total de pixeles en el eje *x* en la imagen original.

La Ecuación (7) se debe aplicar primeramente en el eje x lo que dará como resultado una imagen con la escala deseada en el eje *x*, posteriormente se realizará el mismo procedimiento pero ahora en sobre el eje *y* para así obtener la imagen escalada.

El módulo de reconocimiento e interpretación de la imagen consta de tres etapas (ver Fig. 4) descritas de la siguiente manera.



N B N B N B

0, 5, 2, 8
0, 5, 3, 7
0, 5, 3, 7
0, 4, 2, 1, 2, 6
0, 4, 2, 1, 2, 6
0, 3, 2, 3, 2, 5
0, 3, 2, 3, 2, 5
0, 2, 2, 5, 2, 4
0, 2, 9, 4
0, 2, 9, 4
0, 1, 2, 7, 2, 3
0, 1, 2, 8, 2, 2
2, 9, 2, 2
2, 10, 2, 1
1, 11, 2, 1

Fig. 3. Contabilización de los pixeles

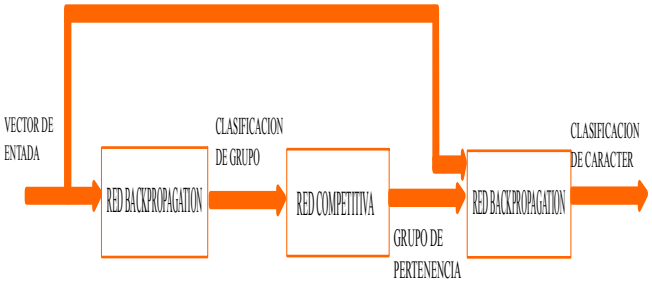


Fig. 4. Diagrama a bloques del módulo de reconocimiento e interpretación de la imagen

En la primer etapa del módulo, el vector de entrada llega a 10 redes de tipo perceptron multicapa, donde cada red clasificará al grupo pertenece dicho vector. Cada una de las 10 redes están entrenadas con diferentes patrones de todos los grupos.

Dichos grupos están previamente analizados de tal forma que en cada grupo existen vectores con características similares es decir esta primera etapa nos ayudará a extraer las características importantes del vector de entrada.

En la segunda etapa una vez que ya se obtuvieron las diez clasificaciones de cada red, funcionarán como entrada a una función de competencia que se basa en las redes de tipo competitivas que como sabemos su funcionamiento se basa en la regla "the winner take all" (el ganador toma todo) la cual indicará finalmente cuál es grupo ganador con el objetivo de trabajar únicamente con un grupo reducido de caracteres y no con todos, como se había planteado en un inicio.

Para la tercer y última etapa del módulo se utilizará nuevamente un perceptrón multicapa para realizar la clasificación final del vector de entrada, pero se tendrá en cuenta la salida de la función de competencia para tomar únicamente la red correspondiente a su grupo, y se propagará el vector de entrada inicial. La salida será únicamente el patrón reconocido

B. Reconocimiento de los Caracteres

1) *Datos de entrada:* Una vez seleccionado el tipo de red, debe tomarse en cuenta el vector de entrada P_j, que será de 100 valores, esta entrada consta de valores binarios (0,1), que surgen a partir de la imagen obtenida por el proceso previo del tratamiento de la imagen [8].

$$P_j = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_{100} \end{bmatrix} \qquad j = 1,2,3\ldots,100 \qquad (8)$$

2) *Datos de Salida para la primera etapa:* La red proporciona a la salida 20 valores binarios que al ser codificados dará como resultado el grupo al que pertenece el vector de entrada, tomando en cuenta que en la posición en la que aparezca el número 1, esta posición indicará a qué grupo de letras es al que posiblemente pertenezca. Por ejemplo:

Si tenemos una salida a = [0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0] el grupo al que esta red está asignando el vector de entrada es el 4.

3) *Número de neuronas de las capas ocultas*. Cuando se diseña una red neuronal por lo general siempre se parte de utilizarse una sola capa oculta, porque para la mayoría de los problemas es suficiente. Por ello el diseño inicial fue de solamente una capa oculta [8] [9].

En cuanto al número de neuronas de la capa oculta existe una regla para obtenerlas llamada "Regla de la Pirámide Geométrica" mostrada en (9), donde, el número de neuronas sigue una forma piramidal, con un número decreciente de neuronas de la entrada hacia la salida. Esta regla es óptima para aplicarse al tipo de red a utilizar.

$$número\_neuronas\_ocultas = \sqrt{mn} \qquad (9)$$

donde:

n es el número de neuronas de entrada.

m es el número de neuronas de salida.

Para nuestro sistema en la primera serie de redes a las que entra n = 100 entradas y m = 20 salidas, utilizando la ecuación antes mencionada obtuvimos 45 neuronas en la capa oculta, pero dado que la red no alcanzo la convergencia en varios ciclos de entrenamiento se tuvo que aumentar el número de neuronas propuestas por la regla de la pirámide y finalmente se utilizaron 50 neuronas en la capa oculta.

4) *Datos de salida de las redes utilizadas en la tercera etapa:* La red debe dar a la salida cinco valores binarios que al ser codificados darán como resultado el carácter reconocido, tomando en cuenta la posición en la que aparezca el número 1, también depende del grupo al que haya pertenecido [9] [10].

Ejemplo 1: Si tenemos que el vector de entrada que fue clasificado en el grupo 1 y al utilizar la red correspondiente a dicho grupo arrojó una salida a = [0 0 0 1] a ser codificado en forma binaria obtendríamos que el caracter H fue reconocido.

Ejemplo 2: Si tenemos que el vector de entrada que fue clasificado en el grupo 4 y al utilizar la red correspondiente a dicho grupo arrojó una salida a = [0 1 0 0] a ser codificado en forma binaria obtendríamos que el caracter D fue reconocido.

5) *Capas ocultas y número de neuronas de las capas ocultas*: En estas redes también se utilizó la regla de la pirámide para determinar el número de neuronas a ocupar en la capa oculta que sería de 22 neuronas, pero dado que la red convergió se fueron disminuyendo el número de neuronas hasta llegar a 8 neuronas donde la red sí convergió con todas las redes que se utilizan en la etapa 3 de este módulo [9] [11].

6) *Arquitectura de las redes de la primera etapa:* Como se ve en la Fig. 5 se utilizó una perceptrón multicapa que constó de tres capas: entrada, consta de cien entradas, cada una de las cuales representa una de las características del caracter a reconocer. Sólo se tiene una capa oculta que consta de cincuenta neuronas cuyas funciones de activación son sigmoidales con salidas acotadas entre 0 y +1. Los umbrales se actualizan al igual que los pesos durante el entrenamiento, cada vez que el patrón evaluado no es clasificado correctamente. Cada una de las neuronas recibe todas las entradas de la capa anterior y sus salidas se conectan a todas las neuronas de la siguiente capa (red totalmente conectada). La red neuronal cuenta con 20 salidas que nos dirá el grupo al que pertenece el carácter que ha sido enviado. Para ello la función de activación es de tipo sigmoidal de manera que se obtengan valores de 0 ó 1 para cada una de las salidas.

7) *Arquitectura de las redes en la tercera etapa:* Se utilizó un perceptrón multícapa entrenado por retropropagación similar al mostrado en la Fig. 5. el cual se describe a continuación [1] [11].

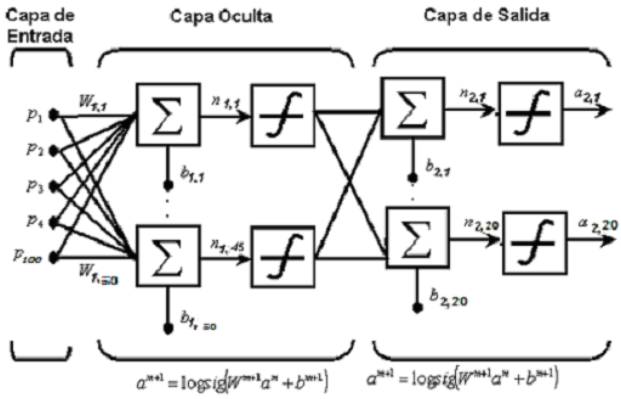


Fig. 5. Arquitectura de las redes de la primera etapa

Consta de cien entradas, cada una de las cuales representa una de las características del carácter a reconocer. Sólo se tiene una capa oculta que consta de 8 neuronas cuyas funciones de activación son funciones sigmoidales que arroja salidas entre 0 y +1. Los umbrales se actualizan al igual que los pesos durante el entrenamiento, cada vez que el patrón evaluado no es clasificado correctamente. Cada una de las neuronas recibe todas las entradas de la capa anterior y sus salidas se conectan a todas las neuronas de la siguiente capa (red totalmente conectada). La red neuronal cuenta con 5 salidas que indicará el carácter al que pertenece el vector que ha sido enviado. Para ello la función de activación es de tipo sigmoidal de manera que se obtengan valores de 0 o 1 para cada una de las salidas.

III. Resultados

Se realizó la digitalización de 70 documentos de distintas personas, basado en el formato que se observa en la Fig. 2. Después de efectuarse el análisis de lo que habían escrito comparado con lo que el sistema había reconocido se determinó que el sistema tiene una eficiencia del 81%.

Cabe mencionar que el 80% de los documentos que fueron analizados por el sistema pertenecían a personas cuya escritura no fue entrenada en el sistema obteniéndose resultados satisfactorios.

El sistema fue entrenado con texto manuscrito del tipo letra de molde sin considerar ningún adorno personal en los caracteres.

El mayor índice de error que presentó el sistema fue con los caracteres Q, O y D lo anterior debido a las características muy similares entre ellos.

IV. Conclusiones

Hoy en día el reconocimiento de patrones por algoritmos de inteligencia artificial basado en las redes neuronales artificiales, es una de las áreas de estudio más amplias y sobretodo con más desarrollo en la industria computacional.

Con este proyecto terminal se dio solución al problema recuperar información manuscrita para fines estadísticos basado en el reconocimiento de texto natural por medio de las redes neuronales artificiales.

El sistema ha sido programado para poder adquirir información a partir de un scanner, cámara digital, Pad o desde un archivo, lo que le da flexibilidad ante los potenciales usuarios. Además de que el campo de aplicación de este proyecto es muy amplio, como el reconocimiento del texto en solicitudes, sobres, encuestas, cheques, etc.

Se desarrolló un sistema de cómputo con la capacidad de reconocer texto natural presentado una eficiencia del 81%, el cual puede ser aplicado para recopilar información manuscrita de encuestas para fines estadísticos.

RECONOCIMIENTOS

Los Autores agradecen a la Escuela Superior de Cómputo del Instituto Politécnico Nacional por el apoyo recibido y las facilidades otorgadas para el desarrollo del presente trabajo terminal.

REFERENCAS

[1] Hagan, Demuth, Beale,"Neural Network Design", PWS Publishing, 1996.
[2] Timothy Masters, , "Practical Neural Network Recipes", Edit, Morgan Kaufman, USA, 493 pp.
[3] Richard E. Woods; C. Rafael Gonzalez, "Tratamiento Digital de imágenes", Addison Wesley-Díaz de Santos.
[4] "Pattern Analysis and Machine Intelligence", IEEE Transactions On, Volúmen 16, enero 1994, páginas ; 98 – 106.
[5] R. Spiegel Murray, "Probabilidad y Estadística", Edit. Mc Graw Hill, Colombia, 372 pp.
[6] S. Pandya Abhijit, Robert B. Macy, "Pattern Recognition with Neural Networks in C++", Edit. IEEE y CRC Press, Florida USA, 1996, 410 pp.
[7] Randy Crane, "A Simplified Approach to Image Processing", Edit. Prentice-Hall, Inc – Hewlett Packard Company, New Jersey, 1997, 317 p.p.
[8] K Mehrotra, C.K. Mohan, S. Ranka, "Elements of Artificial Neural Networks" M.I.T. Press, 1997, ISBN 0-262-13328-8.
[9] Kevin Swingler, , "Applying Neural Networks : A practical guide", Edit. Morgan Kaufman, 1996, USA, 303 pp.
[10] K. Fukushima, S. Miyake, IEEE Transactions On, Systems, Man and Cybernetics, "Neocognitron : A NN model for mechanism of visual pattern recognition", Número 5, Volúmen 13, 1983, páginas 826 – 834.
[11] K. Fukushima, Knowledge and data Engineering IEEE Transactions On, "Neocognitron : A hierarchical NN capable of visual pattern recognition NN", Volúmen 1, 1988, páginas : 119-130.
.
.
.
[20]