

REDES NEURONALES EN OCR CON RECONOCIMIENTO DE IMAGENES

Callisaya Uchani Zenón Amado
amado113@gmail.com

ABSTRACT

Systems that, from a written or printed on paper or similar, creating a text file on a storage medium for computer systems are called OCR (Optical Character Recognition), or optical character recognition. This paper presents an OCR system with the feature of being able to "learn" through a neural network, patterns of characters representing the possible variations (size and font) of how the different characters that can appear in documents in the future and with the same network, able to "recognize" and to convert written text on paper to text stored representing the possible in an ASCII file.

Keywords: OCR, neural networks, shape recognition, *backpropagation*.

RESUMEN

Los sistemas que, a partir de un texto escrito o impreso en papel o similar, crean un fichero de texto en un soporte de almacenamiento informático, se denominan Sistemas de OCR (*Optical Character Recognition*), o de Reconocimiento óptico de Caracteres. En esta ponencia se presenta un sistema OCR con la característica de poder "aprender", mediante una red neuronal, patrones de caracteres que representen las posibles variaciones (tamaño y tipo de letra) de la forma de los diferentes caracteres impresos que pueden aparecer en los documentos para, en el futuro y con la misma red, poder "reconocerlos" y realizar la conversión el texto escrito en papel a texto almacenado en un fichero ASCII.

Palabras claves: OCR, Redes Neuronales, Reconocimiento de Formas, *Backpropagation*.

1. INTRODUCCION

Para el ser humano es relativamente sencillo reconocer los patrones de la escritura manuscrita siendo que esta tarea es sumamente difícil para la computadora.

Un ejemplo típico de red neuronal aplicada al reconocimiento de patrones son los OCR o programas de reconocimiento óptico de caracteres escritos. De acuerdo con Werbos (1998) el 50% de los sistemas de OCR se basan en redes neuronales.

Un programa de OCR consiste en un software que trata de convertir la imagen digitalizada de un documento manuscrito, fotocopiado, etc., en un archivo de texto plausible de ser utilizado por algún procesador de texto.

Los programas de OCR diseñados sobre bases algorítmicas son los mas populares hasta el momento, pero poseen un grado apreciable de inexactitud cuando el texto presenta "ruido", es decir cuando el original contiene manchas (como las producidas al fotocopiar una pagina) o símbolos mezclados con el texto (un dibujo).

Los programas de OCR basados en el reconocimiento de patrones (como las redes neuronales) son capaces de leer y reconocer textos manuscritos o escritos a máquina, de imprenta o de impresora, convirtiéndolos en texto ASCII, editable por el usuario con cualquier programa de tratamiento de textos convencional.

Resulta muy útil para documentos en otros idiomas que pueden ser digitalizados usando el OCR para traducirlos después con un software apropiado, o para digitalizar formularios escritos, especialmente cuando hay que hacerlo en grandes volúmenes.

En documentación manuscrita se requiere una separación entre los caracteres al igual que entre las palabras sucesivas. En la actualidad no se dispone de instrumentos para la segmentación. La separación de campos tienen que estar prevista en el formulario antes de iniciar el proceso. También la calidad de cumplimentación de formularios incidirá en la calidad del reconocimiento.

Para la optimización de los resultados, el modulo OCR debe disponer de un modulo de video corrección. Este modulo debe de estar situado entre el primer modulo para la manipulación física de los documentos y su escaneo y un último modulo de manipulación sobre bases de datos.

El modulo OCR dispone de tres componentes funcionales.

Sub modulo de localización de campos y extracción de caracteres. Permitirá el tratar grandes volúmenes de documentos.

Sub modulo de reconocimiento de caracteres individuales. se realiza mediante la técnica neuronal de empleo de Perceptores Multicapa.

Sub modulo de validación e interpretación parcial. Ante una información concreta se pueden cruzar datos almacenados anteriormente. Para una optimización de las consultas se pueden utilizar índices.

Previo a la video corrección la biblioteca puede:

Seleccionar imágenes de los campos a tratar, dados como coordenadas sobre la imagen del correspondiente documento, así como el resultado del reconocimiento parcial de los mismos.

Definir para las mismas y sobre cada documento un orden de presentación en video corrección.

Incorporar mecanismos para la interpretación de campos pendientes de un documento en función del resultado de la video corrección de otros anteriores.

Es frecuente que se queden caracteres sin identificar. La productividad de OCR no esta tanto en no tener un solo error, si no en su capacidad de procesar enormes cantidades de documentos.

APRENDIZAJE Y RECONOCIMIENTO DE CARACTERES CON UNA RED NEURONAL

Las Redes Neuronales son sistemas de computación que permiten la resolución de problemas que no pueden ser descritos fácilmente mediante un enfoque algorítmico tradicional, como, por ejemplo, ocurre en el reconocimiento de formas. Con las redes se expresa las soluciones en un problema, no como una secuencia de pasos, sino como la evolución en un sistema inspirado en el funcionamiento del cerebro y dotado, por tanto, de cierta "inteligencia". Tal sistema no es sino la combinación de una gran cantidad de elementos simples de proceso (nodos o neuronas) interconectados que, operando de forma masivamente paralela, consiguen resolver el problema.

Las conexiones sirven para transmitir las salidas de unos nodos a las entradas de otros. El funcionamiento de un nodo es similar al de las neuronas biológicas presentes en el cerebro. Suele aceptarse que la información memorizada en el cerebro esta relacionada con los valores sintéticos de las conexiones entre las neuronas. De igual forma, se dice que las redes neuronales tienen la capacidad de "aprender" mediante el ajuste de las conexiones entre nodos.

Existen muchos tipos de redes neuronales, un modelo utilizado en una gran variedad de aplicaciones, entre las que se incluyen las de reconocimiento de formas [5][6], es la red multicapa con conexiones unidireccionales hacia delante (*feedforward*). En la figura 2 se muestra la red de este tipo utilizada en el sistema OCR para el reconocimiento de caracteres.

En este tipo de estructura es posible distinguir una capa de entrada, varias capas de neuronas ocultas y una capa de salida.

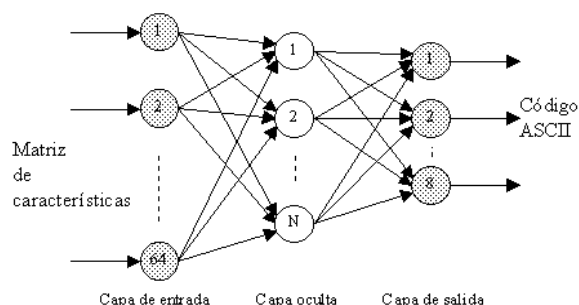
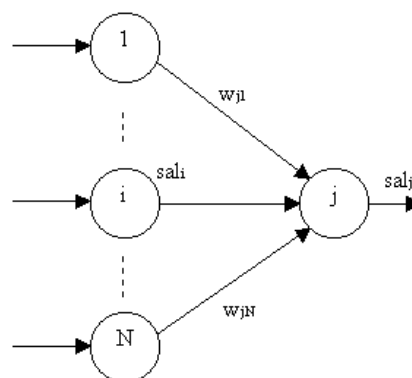


Figura 2.- Estructura de la red neuronal utilizada

Cada una de las neuronas de la red recibe como entrada la suma de los valores de salida de las neuronas a ella conectadas multiplicados por el peso de la conexiones correspondiente. Al resultado de esta suma se le aplica una función de transferencia (f) para obtener el valor que genera a la salida.

$$sal_j = f\left(\sum_{i=1}^N w_{ji} sal_i\right)$$



Una función de transferencia que habitualmente se utiliza en este tipo de redes es la denominada "sigmoideal":

$$y=f(x)=1/(1+e^{-x})$$

Cuando se trabaja con redes neuronales como la que se utiliza en este trabajo, se distingue entre una primera etapa de aprendizaje o entrenamiento de la red y una segunda etapa de funcionamiento u operaciones la red. En la primera se realiza el ajuste de los pesos en funciones lo que se desea que almacene la red. Se suele distinguir entre dos tipos de aprendizaje: supervisado o no supervisado.

La red utilizada en el sistema de OCR, tiene un aprendizaje supervisado, el cual consiste en el ajuste de los pesos para que "aprenda" a relacionar un conjunto de patrones de entrada con las salidas que se desea que la red genere ante cada uno de dichos patrones. El entrenamiento concluye cuando la diferencia entre las salidas deseadas ante cada patrones las que se obtienen realmente con los valores calculados de los pesos es suficientemente pequeños.

Por ejemplo, el aprendizaje de un alfabeto de 37 caracteres con un tipo de letra Courier, con 12 neuronas ocultas converge en tan solo 20 iteraciones, reconociendo posteriormente un texto de prueba sin ninguno. Con 10 neuronas y los mismos patrones de aprendizaje no lo hace en 60, y con 20 en 9 iteraciones.

En lo que se refiere a la capa de salida, como se pretende que la red genere el código ASCII de 8 bits del carácter que se presenta a la entrada, se necesitan 8 neuronas con salida binaria. Otra posible estructura de salida, aunque no se ha implementado aquí podría utilizar tantas neuronas de salida como caracteres a reconocer, es decir una salida para cada tipo de carácter del alfabeto que se quiera utilizar.

2. FIGURAS/LEYENDAS

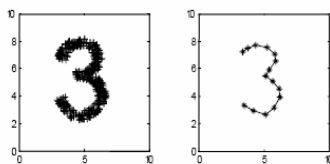


Fig.1. Resultado de la red para el numero 3

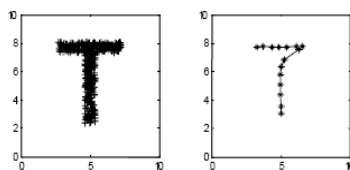


Fig.2 Resultado de la red para la letra T

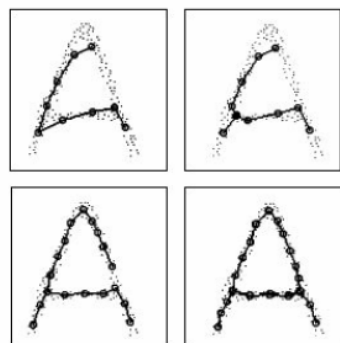


Fig.3 Resultado de la red para la letra A

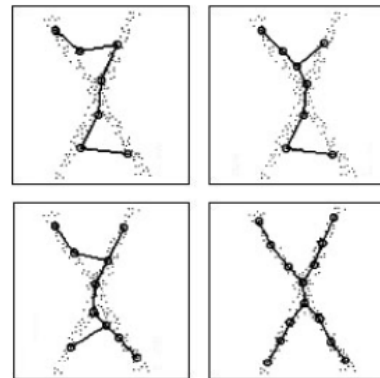


Fig.4 Resultado de la red para la letra X

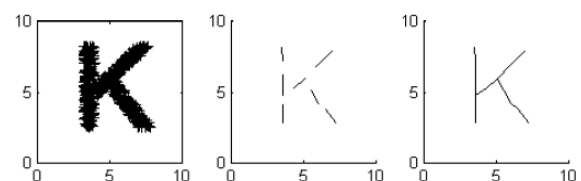


Fig. 5 Resultados de la letra K con 7 segmentos

CONCLUSION

En este artículo se ha propuesto una red neuronal competitiva no supervisada basada en segmentos con fase de expansión. Hemos visto cómo la red expuesta sin la fase de expansión es capaz de proporcionar segmentos que nos indican la dirección predominante de los datos. Sin embargo es muy sensible a la inicialización de los pesos sinápticos de las neuronas y al orden de introducción de los patrones.

3. REFERENCIAS

1. Amin, A., Al-sadoun, H., Fischer, S.: Hand-printed arabic carácter recognition system using an artificial network. Pattern Recognition, 29(4):663-675, 1996.
2. Bayro-Corrochano, E.J., Pahm, D.T.: Selforganizing neural-network-based pattern clustering method with fuzzy outputs. Pattern Recognition, 27:1103-1110, 1994.