11° encuentro de investigación · inaoe

4-5 de noviembre de 2010

INSTITUTO NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA

XI

ENCUENTRO DE INVESTIGACIÓN

MEMORIA TÉCNICA

Tonantzintla, Puebla, México.

4 y 5 de noviembre de 2010

COMITÉ ORGANIZADOR:

Dr. Roberto Murphy Arteaga, Director de Formación Académica
Dr. Alberto Carramiñana Alonso, Coordinador de Astrofísica
Dr. Sergio Vázquez y Montiel, Coordinador de Óptica
Dr. Arturo Sarmiento Reyes, Coordinador de Electrónica
Dr. Aurelio López López, Coordinador de Cs. Computacionales
Dr. Divakara Mayya Yalia, Representante Docente de Astrofísica
Dr. Francisco Renero Carrillo, Representante Docente de Óptica
Dr. Pedro Rosales Quintero, Representante Docente de Electrónica
Dr. Ariel Carrasco Ochoa, Representante Docente de Cs. Computacionales
Martha Aurora Olmos y Flores
Gabriela López Lucio
José Luis Toxqui Olmos
Ricardo Toxqui Olmos
Fabiola Vázquez Tecuanhuey
Cecilia Tecuatl Cuautle
María Guadalupe Coyotl Tecuatl
María Esther Montes Tecanhuey
Rocío Leticia Rodas Fernández
Rosario Tlahuel Tello
Landy del Carmen Ríos Morales
Catalina Montes Tecuanhuey

EDICIÓN:

Dr. Roberto Stack Murphy Arteaga
Martha Aurora Olmos y Flores
Gabriela López Lucio
José Luis Toxqui Olmos
Ricardo Toxqui Olmos

CALLE LUIS ENRIQUE ERRO NO 1
SANTA MARÍA TONANTZINTLA, PUEBLA
C.P. 72840, TEL (222)247-27-42
http://www.inaoep.mx

Redes Neuronales Auto-organizables y Recurrentes para reconocimiento temporal: el caso de palabras manuscritas

Rodolfo Luna Pérez, Pilar Gómez Gil
INAOE
Coordinación de Ciencias Computacionales
rodolfo.luna@inaoep.mx, pgomez@inaoep.mx

RESUMEN

El reconocimiento de texto manuscrito fuera de línea es una tarea importante debido a que existen grandes volúmenes de datos contenidos en documentos escritos. Sin embargo, la variedad que existe en los estilos de escritura dificulta las tareas involucradas en el reconocimiento. En este trabajo se presenta un método de reconocimiento de palabras fuera de línea basado en 2 redes neuronales artificiales: los mapas auto-organizados (SOM) y las redes recurrentes simples (SRN). El sistema crea prototipos de trazos manuscritos utilizando la red SOM. Durante el reconocimiento, la SOM elige prototipos para los trazos de la palabra manuscrita que se proporcionan a una red SRN a partir de una secuencia de entrada uno a uno, la cual determina la clase correcta . El método muestra mejor desempeño en que otros reconocedores básicos basados en redes neuronales al aplicarse al reconocimiento de palabras de la base de datos .

I. INTRODUCCIÓN

El reconocimiento de texto manuscrito consiste en transformar un lenguaje de una representación gráfica a una representación simbólica [**1**]. Generalmente se clasifica en reconocimiento en línea y reconocimiento fuera de línea. El reconocimiento en línea cuenta con información espacio temporal, la cual es obtenida durante la escritura, mientras que en el reconocimiento fuera de línea sólo se cuenta con una imagen que contiene texto. El reconocimiento de texto manuscrito puede ser a nivel de caracteres, palabras ó líneas completas de texto. El reconocimiento de palabras consiste en encontrar, entre las palabras de un léxico definido, la más compatible con los datos manuscritos [**2**]. Las técnicas de reconocimiento de palabras pueden dividirse en dos categorías: analíticas y globales [**3**]. El enfoque analítico consiste en dividir la imagen en caracteres o pseudo-caracteres (partes de un carácter) para después identificar cada uno de estos segmentos y utilizando información del contextual y un diccionario reconocer la palabra. El enfoque global trata a una palabra como una sola entidad, no se hace segmentación, el reconocimiento se realiza utilizando características de la imagen completa evitando los problemas de la segmentación, sin embargo es necesario modelar cada palabra.

El presente trabajo se enfoca en un método analítico para el reconocimiento fuera de línea de palabras aisladas, independiente del escritor y con vocabulario definido basado en dos tipos de redes neuronales artificiales: redes recurrentes y auto-organizadas. Las redes neuronales recurrentes permiten conexiones de una neurona a sí misma, a o a otras neuronas de la misma capa o incluso a neuronas de capas anteriores. Estas conexiones, llamadas recurrentes, pueden almacenar información obtenida de entradas anteriores, por lo cual se puede usar información contextual del pasado. Un modelo básico de redes recurrentes es el SRN (*Simple Recurrent Network*) [**5**]. Los mapas auto-organizados [**6**], también conocidos como redes SOM (*Self-organizing maps*) son redes neuronales que generan grupos a partir de datos no etiquetados. La capa de salida de la red forma un mapa, generalmente bidimensional, en el cual cada elemento es el centroide de un grupo. A causa del algoritmo no supervisado y competitivo utilizado para entrenar la red SOM, el mapa de salida tiene grupos ordenados por similitud, es decir, las neuronas en la capa de salida representan a datos similares a los de sus vecinos.

El artículo está organizado de la siguiente manera: la sección II presenta el modelo de reconocimiento

propuesto, la sección III describe brevemente la manera en que el sistema se entrena, la sección IV resume los experimentos y resultados obtenidos al aplicar el modelo a la base de datos IAM-DB [7], y la sección V muestra algunas conclusiones de la investigación.

II. MODELO PROPUESTO

El modelo de reconocimiento de palabras propuesto consta de cuatro etapas: pre-procesamiento de las imágenes, segmentación, generación de vectores de características basado en una SOM y clasificación de palabras usando una SRN (Figura 1). Cada etapa se explica con detalle a continuación.

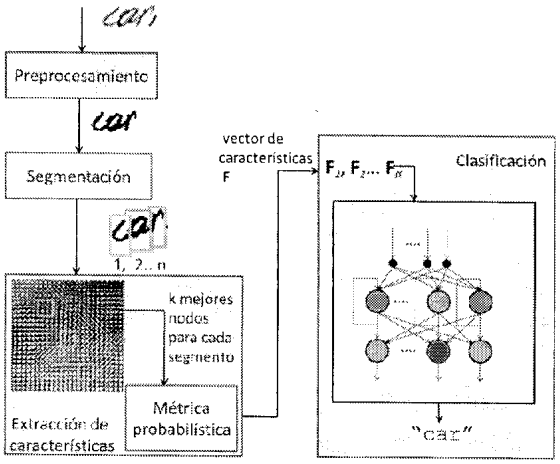

Figura 1 Esquema general del método propuesto.

1. El pre-procesamiento de imágenes consiste en la eliminación de ruido, binarización de la imagen y corrección de la inclinación vertical de los trazos, este último utilizando el método propuesto por Ding y colaboradores [8]. La Figura 2 (a) muestra una imagen tomada de la base de datos IAM-DB correspondiente a la palabra "*action.*" La Figura 2 (b) es la imagen resultante después de aplicar el preprocesamiento a la imagen original.

2. Una vez pre-procesada una imagen se segmenta automáticamente de la siguiente manera: cada columna de la imagen se recorre de abajo hacia arriba, hasta encontrar el primer pixel negro, a fin de localizar todos los pixeles inferiores de la imagen. Enseguida se buscan los máximos locales y son utilizados como puntos de segmentación. La Figura 2 (c) muestra en negro los pixeles inferiores, y las líneas verticales indican los puntos de segmentación correspondientes a los máximos locales del borde formado por los pixeles inferiores.

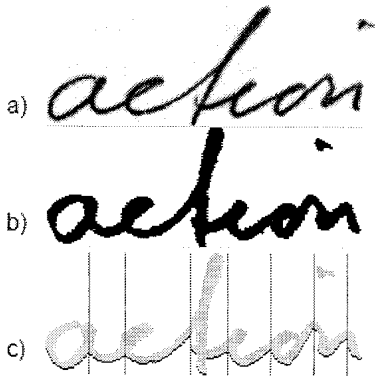

Figura 2 a) Ejemplo de imagen original obtenida de la IAM-DB [7]. b) Imagen resultante del pre-procesamiento. c) Segmentación automática.

Por último los segmentos son normalizados a un mismo tamaño. Los segmentos obtenidos del pre-procesamiento se representan con vectores derivados de la concatenación de los valores correspondientes a cada renglón de la imagen.

3. Estos segmentos se usan para la generación de vectores de características, proceso basado en una red SOM y una métrica probabilística. En la fase de entrenamiento, una SOM con mapa de salida de neuronas, es entrenada a partir de los segmentos. Durante la fase de clasificación, cada segmento es presentado a la red, que obtiene una neurona ganadora . Además se seleccionan las k-1 neuronas con mayor activación en el mapa, identificadas como con donde . De esta forma se tiene información sobre los k grupos que mejor se ajustan al segmento de entrada . Para cada neurona se calcula la probabilidad de que dicha neurona represente al segmento utilizando una métrica basada en la distancia Euclidiana entre la neurona ganadora y las neuronas dentro del mapa, definida por:

(1)

Donde es un vector bidimensional definido por las coordenadas de la neurona ganadora en la red SOM al presentar el segmento , es el vector

bidimensional de las coordenadas de la i-ésima neurona con mayor activación en el mapa SOM, $i = 1..k$, cuando el segmento t es presentado a la res SOM, $||.||$ es la distancia Euclidiana entre sus argumentos. Para cada segmento t:

a) $\sum_{i=1}^{k} p_t(\mathbf{m_{ti}}) = 1$ (2)

b) si $||\mathbf{m_{ti}} - \mathbf{c}|| = ||\mathbf{m_{tj}} - \mathbf{c}||$ entonces

$p(\mathbf{m_{ti}}) = p(\mathbf{m_{tj}})$ (3)

En resumen, el nuevo vector de características para el t-ésimo segmento está definido como:

$$\boldsymbol{F_t} = (c_{t1}, c_{t2}, p(\boldsymbol{c_t}), m_{t21}, m_{t22}, p(\boldsymbol{m_{t2}}), m_{t31}, m_{t32}, p(\boldsymbol{m_{t3}}), \dots, m_{tk1}, m_{tk2}, p(\boldsymbol{m_{tk}})) \quad (4)$$

F_t es un vector con dimensión igual a $3k$.

4. Para la etapa de clasificación se utiliza una SRN, la cual es alimentada con una secuencia de vectores $\boldsymbol{F}$ que representan a los segmentos de una palabra. Cada segmento es introducido uno a la vez. Cuando el último segmento de la palabra a clasificar es presentado a la red, la neurona con mayor activación en la capa de salida representa a la clase a asignar a la palabra, tomada ésta de un léxico definido, siendo la que mejor representa a la secuencia de segmentos de entrada. La SRN consiste de una capa de entrada, una oculta con conexiones recurrentes a sí misma y de salida que contiene una neurona por cada palabra que se desea reconocer. Los nodos de la capa de salida usan la función de activación $softmax$, definida como:

$$softmax(out_i) = \frac{\exp(out_i)}{\sum_{j=1}^{n} out_j} \quad (5)$$

Donde n es el número de neuronas en la capa de salida y out_i es la activación de la i-ésima neurona. Al utilizar $softmax$ la suma de la activación de las neuronas en la capa de salida es igual a 1, lo que puede interpretarse como la probabilidad de que la secuencia de entrada sea representada por la neurona de la capa de salida.

III. ENTRENAMIENTO DEL SISTEMA

La red SOM se entrena a partir de los segmentos obtenidos de las imágenes pre-procesadas del conjunto de entrenamiento; de esta forma se generan prototipos para los segmentos de palabra. La SRN es entrenada de manera supervisada para recibir secuencias de vectores F_t (ecuación 4) correspondientes a cada palabra. Cuando el último segmento de cada palabra es alimentado a la SRN, la neurona de la capa de salida que mejor representa a la clase a asignarse deberá valer uno. La salida deseada d_i para cada nodo de salida i cuando se presenta el segmento $F_t$que no es el último segmento de una palabra se calcula como:

$$d_i = \begin{cases} \frac{x}{n}, si\ i\ corresponde\ a\ la\ palabra \\ 0, en\ otro\ caso \end{cases} \quad (6)$$

Donde x es la posición del segmento $F_t$en la palabra y n es el número de segmentos de la palabra. Se utilizó el algoritmo de retropropagación a través del tiempo para entrenar la red. Para evitar que la red aprendiera el conjunto de datos de entrenamiento como una sola serie de tiempo, el ajuste de los pesos de la red se realizó después de presentar a la red los segmentos correspondientes a una sola palabra.

IV. EXPERIMENTOS Y RESULTADOS

Para probar el método se hicieron experimentos para un vocabulario de 10 palabras, que se muestran en la Tabla 1. El conjunto de entrenamiento se formó con 20 imágenes de cada palabra, mientras que el conjunto de prueba contenía 5 imágenes por palabra.

Tabla 1 Vocabulario utilizado para los experimentos.

Vocabulario									
a	and	are	as	at	be	but	bye	can	for

Se realizaron experimentos con 3 diferentes tipos de redes neuronales:

1) Red alimentada hacia adelante *(feed-forward)* (FF): Los datos de entrada fueron segmentos de palabra representados por la concatenación de los valores binarios de cada uno de los renglones de la imagen (2,800 nodos de entrada). Este modelo se utilizó como caso base para las comparaciones realizadas.

2) Red SOM-FF: Los datos de entrada fueron construidos a partir de la red SOM con k=5 (315 nodos de entrada).

3) SOM-SRN (método propuesto). Los vectores de entrada son construidos a partir de la red SOM, y los vectores son presentados a una SRN uno a uno (15 nodos de entrada).

Se realizaron 20 experimentos para 100 configuraciones distintas de cada red. Se evaluó el desempeño del sistema calculando el error al reconocer las palabras del conjunto de entrenamiento y prueba y calculando la métrica "exactitud de palabra" [4]. La **Tabla 2** y la Tabla 3 muestran los resultados de la configuración de red que obtuvo el mejor desempeño para cada caso.

Tabla 2 Porcentaje de error obtenido.

Caso	*Porcentaje de error*	
	Conjunto de entrenamiento	**Conjunto de Prueba**
FF	83.30% ±9.07%	85.90% ±5.89%
SOM-FF	12.15% ±5.04%	48.50% ±5.10%
SOM-SRN	**1.65%±0.58%**	**29.60%±4.06%**

Tabla 3 Exactitud de palabra obtenida.

Caso	*Exactitud de palabra*	
	Conjunto entrenamiento	**Conjunto de Prueba**
FF	19.13% ±10.89%	16.66% ±9.58%
SOM-FF	89.05% ±4.65%	54.96% ±4.8%
SOM-SRN	**98.67% ±0.85%**	**73.92% ±3.09%**

Los resultados muestran que SOM-SRN (método propuesto) tiene un mejor desempeño que los otros 2 reconocedores. Comparando los resultados obtenidos por la red FF con los del caso SOM-FF se puede observar que el uso de una red SOM ayuda mejorar el reconocimiento.

V. Conclusiones

En este artículo se presentó un método novedoso de reconocimiento de palabras basado en una red SOM para crear modelos de segmentos de palabras de manera no supervisada y una red SRN para la clasificación de secuencias compuestas de dichos segmentos. Los resultados muestran que el uso de SOM ayuda a mejorar el reconocimiento. Además el método propuesto mostró un mejor desempeño que los otros dos reconocedores presentados en el reconocimiento de 10 palabras.

VI. Agradecimientos

R. Luna agradece al CONACyT el apoyo otorgado a través de la Beca para estudios de Maestría # 224412.

VII. Referencias

[1] Rejean Plamodon and Sargur Srihari, "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 01, pp. 63-84, 2000.

[2] Alessandro Vinciarelli, "A survey on off-line Cursive Word Recognition," *Pattern Recognition*, vol. 35, no. 7, pp. 1433 - 1446, 2002.

[3] Abderrahmane Namane, Abderrezack Guessoum, and Patrick Meyrueis, "New Holistic Handwritten Word Recognition and Its Application to French Legal Amount," *Sprinnger'Verlag Berlin Heidelberg*, pp. 654-663, 2005.

[4] Alex Graves et al., "A Novel Connectionist System for Unconstrained Handwriting Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 855-868, 2009.

[5] Jeffrey L. Elman, "Finding Structure in Time," *Congnitive Science*, vol. 14, pp. 179-211, 1990.

[6] Teuvo Kohonen, "Automatic formation of topological maps patterns in a self-organizing system," , 1981, pp. 214-220.

[7] U. Marti and H. Bunke, "A full English sentence database for off-line handwriting recognition," *Proc. of the 5th Int. Conf. on Document Analysis and Recognition*, pp. 705 - 708, http://www.iam.unibe.ch/fki/databases/iam-handwriting-database, 1999.

[8] Yimei Ding, Wataru Ohyama, Fumikata Kimura, and M Shridhar, "Local Slant Estimation for Handwritten English Words," in *Proceedings of the 9th Int'l Workshop on Frontiers in Handwriting Recognition*, 2004.