

Reconocimiento de caracteres manuscritos a partir de señales EMG usando redes neuronales convolucionales y recurrentes de Elman

José Guadalupe Beltrán-Hernández, José Ruiz-Pinales,
Pedro López-Rodríguez, José Luis López-Ramírez

Universidad de Guanajuato, División de Ingenierías,
Departamento de Ingeniería Electrónica,
México

{jg.beltranhernandez, pinales, p.lopez.rodriguez,
jl.lopezramirez}@ugto.mx

Resumen. Hoy en día la tecnología está presente en todos los ámbitos de la vida diaria, y la escritura es sin duda un medio muy importante de comunicación. Además, la escritura y la lectura juegan un papel importante en el desarrollo cognitivo de los niños, tanto así que éstas determinan su posterior éxito académico. Este documento presenta una nueva metodología para el reconocimiento de caracteres manuscritos. La metodología propuesta consiste en el uso de un brazalete Myo para capturar las señales EMG que se generan durante el proceso de escritura y el uso de redes neuronales convolucionales y redes recurrentes para el procesamiento y reconocimiento de caracteres. Los resultados usando una base de datos propietaria grande muestran un porcentaje de aciertos promedio de 94.92 %.

Palabras clave: Señales electromiográficas, aprendizaje profundo, redes neuronales convolucionales, redes de Elman.

Handwritten Character Recognition from EMG Signals using Convolutional and Elman Neural Networks

Abstract. Nowadays, technology is present in almost every aspect of our lives. Reading and writing continue to be an important means of communication. Besides that, they play an important role in the cognitive development of children and later on determine their academic success. This paper presents a new methodology for the recognition of handwriting. The proposed methodology consists in using a Myo armband to capture the EMG signals generated during the writing process and the use of Convolutional Neural Networks and Elman Networks for processing and recognizing characters. The results using a large proprietary database show an average accuracy of 94.92 %.

Keywords: Electromyographic signals, deep learning, convolutional neural networks, Elman networks.

1. Introducción

La escritura y la lectura, son tareas complejas y habilidades importantes para los seres humanos. De hecho, existen estudios que han mostrado la importancia de tales habilidades para el desarrollo cognitivo y el éxito académico de los niños [15]. Sin embargo, resulta que la escritura manuscrita a partir de señales bioeléctricas ha sido poco abordada en el área de las interfaces humano-computadora. Estas interfaces son importantes porque ofrecen una alternativa en aplicaciones clínicas y de asistencia para las personas con o sin discapacidades.

Ya se han logrado desarrollar algunas prótesis neurales capaces de hacer los mismos movimientos que un brazo o una pierna humana. Incluso, ya existen interfaces para juguetes y videojuegos que permiten el control por medio de las señales bioeléctricas de los usuarios. En particular, existen dispositivos comerciales tales como el brazalete Myo que puede usarse para controlar presentaciones, vehículos, etc.

Las señales bioeléctricas tales como las señales EMG (electromiográficas), han sido aplicadas a la detección de trastornos neurológicos [10]. Por ejemplo, el análisis de patrones de escritura se ha utilizado para el diagnóstico de la enfermedad de Parkinson y la disgrafía en niños.

En la literatura hay un cierto número de publicaciones que tratan con las señales EMG generadas por los músculos de las manos y los antebrazos durante la actividad de la escritura [14,6]. En general, el reconocimiento de la escritura a partir de las señales EMG se han abordado utilizando diferentes enfoques: concordancia de plantillas [14], Dynamic Time Warping (DTW) [6,13], Redes Neuronales RBF [10]. Igualmente, la reconstrucción de la escritura a partir de las señales EMG ha sido abordada usando: filtros de Wiener [14], mínimos cuadrados recursivos [8], Redes Neuronales RBF [18], y filtros de Kalman [16].

Las señales EMG son bioseñales complejas, cuya forma se ve afectada por muchos factores externos así como por propiedades de los músculos, la instrumentación que se utiliza para detectarlas, etc. La complejidad puede explicarse por el simple hecho de que hay muchas maneras de realizar una tarea motora (fenómeno de redundancia motora) [16,4,1].

Siempre se ha sabido que las señales EMG contienen una gran cantidad de información sobre los movimientos de las extremidades del cuerpo humano. Aprovechando esta información, varias aplicaciones en la literatura han efectuado reconstrucción de movimientos de la escritura, modelado, simulación de movimientos, reconocimiento de movimientos. Con respecto al reconocimiento de caracteres existen muy pocas bases de datos. Sin embargo, muchos de los estudios imponen restricciones de diferentes tipos: colocación de los electrodos sobre músculos del antebrazo, número y complejidad de los trazos manuscritos.

En nuestro caso, proponemos abordar el problema del reconocimiento de caracteres manuscritos desde un punto de vista más general. En lugar de imponer restricciones, se utiliza una pulsera Myo en el antebrazo y el usuario puede usar su estilo de la escritura. Para contrarrestar las variaciones de las señales EMG, se propone el uso de redes neuronales convolucionales (CNN), para el procesamiento y extracción de características. Esto es debido a que se ha demostrado que las CNN son excelentes para extraer características directamente de los datos en bruto [3,12]. Para el procesamiento de las secuencias se propone el uso de redes recurrentes (RNN) con funciones de activación ReLU. Estas redes son menos pesadas computacionalmente y tienen igual o mejor desempeño que otras redes más recientes tales como las LSTM.

En la siguiente sección se presenta una descripción de los materiales y métodos empleados durante este estudio. Finalmente, se presentan los resultados y conclusiones.

2. Métodos y materiales

2.1. Materiales

A fin de validar nuestra metodología, se escogió una gran base de datos que contiene señales EMG. Las señales EMG fueron capturadas por medio de un brazalete Myo colocado en el antebrazo derecho de tres personas. El brazalete Myo es un dispositivo comercial inalámbrico de bajo costo, no invasivo, de 8 canales, y tiene una frecuencia de muestreo de 200 Hz. La Fig. 1 muestra cómo se usa el dispositivo para capturar las señales EMG generadas al momento de escribir con el dedo sobre una pantalla.

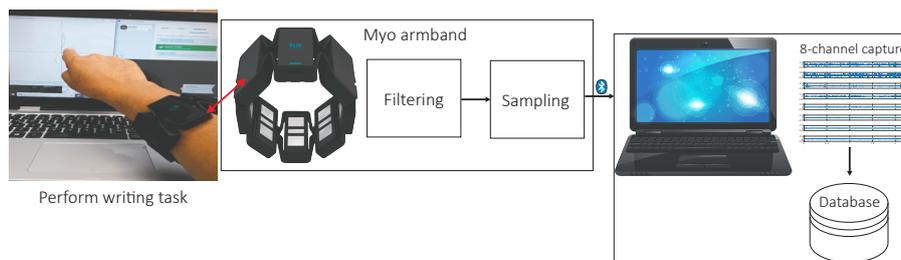


Fig. 1. Diagrama del sistema de captura de las señales EMG.

Por lo tanto, contiene 3,000 capturas por clase. Cada captura comprende 8 señales correspondientes a los 8 sensores del dispositivo. La Fig. 2 muestra una captura para la letra 'b'. En este caso particular, se puede observar más actividad en los 4 primeros canales. En el caso general, el brazalete puede colocarse en cualquier posición, y los canales activos suelen ser diferentes.

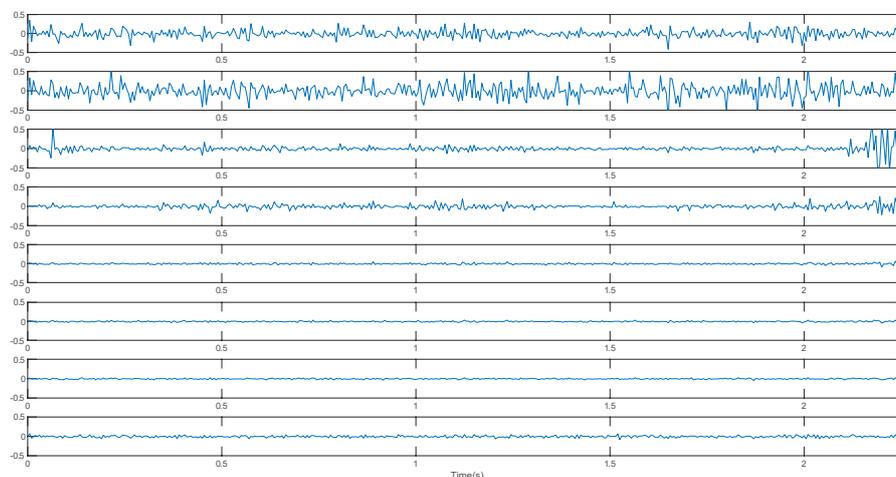


Fig. 2. Ejemplo de una captura correspondiente a la letra 'b'. La parte superior corresponde al canal 0 y la parte inferior al canal 7.

2.2. Métodos

Redes Neuronales Convolucionales Una red neuronal convolucional (CNN), es una red neuronal estática inspirada originalmente de hallazgos en neurociencias. En 1959, Hubel & Wiesel [7], propusieron un modelo jerárquico de red para explicar el comportamiento de las células simples del gato. A partir de ese modelo se propusieron varios otros modelos como el Neocognitrón [2] que quizás constituye una de las primeras CNN, ya que usaba capas convolucionales y capas de submuestreo. Otro modelo, el Cresceptrón [21], utilizó por primera vez el mecanismo de max-pooling.

El modelo LeNet fue el primero en utilizar el algoritmo de retropropagación [12]. Este modelo demostró que una CNN puede procesar la imagen directamente sin ningún o poco preprocesamiento. Lo que abrió el camino para el desarrollo de las arquitecturas actuales de Deep Learning. Aunque el modelo LeNet-5 tuvo gran éxito en el reconocimiento de dígitos, su uso para tareas más complejas estaba muy limitado debido, por un lado, al poco poder de cómputo disponible en su época, y, por otro lado, a algunos problemas que presentaban las CNNs de mayor profundidad (DCNN) [3].

Después de varios años, las CNN resurgieron exitosamente con el modelo AlexNet al ganar por primera vez un concurso internacional (ILVRC2012). El modelo AlexNet [9], tenía una arquitectura más profunda que el modelo LeNet-5, con capas más grandes y nuevos avances como: función de activación ReLU, métodos de inicialización adHoc, mecanismo de regularización llamado Dropout, y mecanismo max-pooling.

El modelo ResNet [5], quien ganó algunos años después el mismo concurso (ILSVRC2015), era mucho más profunda que el modelo AlexNet, e incorporó innovaciones como el bloque residual, y capas batch-normalization.

Redes Neuronales Recurrentes Las redes Neuronales recurrentes (RNN) son un tipo de red neuronal dinámica capaz de procesar secuencias [17]. Las redes de Elman y Jordan constituyen algunas de las primeras redes recurrentes. Su característica principal es que tienen la capacidad de memorizar patrones por medio de la retroalimentación de las neuronas. Su entrenamiento utiliza un algoritmo llamado retropropagación en el tiempo (BPTT). En realidad, la red recurrente es desdoblada en el tiempo dando lugar a una variante de red estática con pesos compartidos que se entrena usando el algoritmo de retropropagación. Estas primeras redes recurrentes tuvieron un éxito limitado debido a que presentaban problemas para aprender secuencias largas (desvanecimiento del gradiente). Se podría decir que estas redes sólo eran capaces de memoria de corto plazo. La Fig. 3 muestra una RNN y su desenvolvimiento en el tiempo.

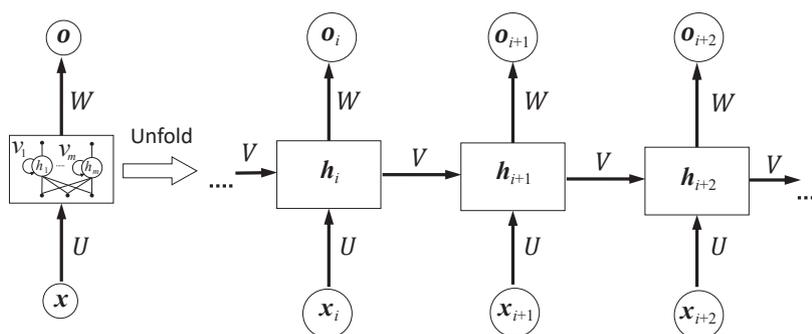


Fig. 3. Arquitectura de una RNN y su desenvolvimiento en el tiempo.

Las redes recurrentes LSTM (Long Sort-Term Memory) fueron las primeras en incorporar la capacidad de memoria de largo plazo ya que no presentaban más el problema del desvanecimiento del gradiente. Una red LSTM está compuesta de unidades de memoria, con la capacidad de decidir cuándo recordar, olvidar y adquirir información. La operación de una unidad LSTM está controlada por tres compuertas multiplicadoras: una para olvidar, otra para adquirir información y otra para dar salida al estado de la celda de memoria. El problema de las redes LSTM es que contienen mucho más parámetros que sus predecesoras.

3. Metodología propuesta

En general, el reconocimiento de caracteres a partir de señales EMG se ha basado en el pre-procesamiento, extracción de características, y clasificación [14].

Sin embargo, existen métodos que se basan en el preprocesamiento, medida de similitud usando DTW y clasificación usando distancia Euclidiana. Por otro lado, la mayoría de los estudios anteriores han sido realizados bajo condiciones controladas de tal manera que se facilita la clasificación de las señales EMG. Por ejemplo, colocar cuidadosamente cada electrodo sobre cada músculo de la mano o antebrazo. Otras restricciones son que cada letra conste de un solo trazo y que las letras sean de la misma forma.

En este trabajo, se propone una metodología más general para el reconocimiento de caracteres a partir de señales EMG. En particular, la escritura es multi-usuario, sin restricciones en la forma del caracter y el número de trazos. Además, en lugar de colocar electrodos sobre el brazo de los sujetos, se utiliza un brazaletes Myo en el antebrazo. Para hacer frente a la problemática de la variabilidad de las señales EMG, se usan capas convolucionales y max-pooling para la extracción de características, IRNN para el procesamiento de la secuencia y una capa totalmente conectada para la clasificación. La razón por la que se escogió una RNN es que son menos pesadas computacionalmente que otras redes más recientes. Además, el problema del desvanecimiento gradiente se debe al uso sigmoides como función de activación. Por lo tanto, las RNN que utilizan ReLU como función de activación, llamadas IRNN, no presentan el problema del desvanecimiento del gradiente [11]. Se ha observado que en muchos casos hasta han llegado a superar a otras redes tales como las LSTM [11,20].

Sin embargo, las IRNN aún pueden presentar problemas de convergencia debido al uso de ReLU. Afortunadamente, eso se puede resolver fácilmente inicializando los pesos recurrentes a 1 y los demás pesos a valores pequeños generados de una distribución normal [11].

La Fig. 4 muestra la arquitectura de la red propuesta. Se utiliza la función de activación ReLU en todas las capas convolucionales. Para combatir el sobreaprendizaje y mejorar la generalización de la red, se aplicó la técnica de Dropout en cada capa convolucional. La capa totalmente conectada utiliza la función de activación SoftMax para que las 36 salidas puedan corresponder a probabilidades. La longitud máxima de secuencia se puso en $L = 900$.

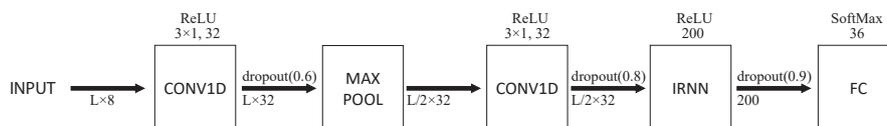


Fig. 4. Arquitectura de la red propuesta. La longitud de la secuencia de entrada está indicada con L .

4. Implementación

Para estimar el desempeño del sistema se utilizó validación cruzada estratificada k-fold. Por lo que la base de datos se dividió al azar en 10 partes disjuntas

de igual tamaño. Es importante mencionar que se tuvo cuidado de preservar la proporciones de cada clase en cada parte. Cada parte se tomó como conjunto de prueba y el resto se utilizó para entrenar la red neuronal. El desempeño de la red se calcula como el promedio de los resultados de las 10 pruebas.

La implementación se efectuó bajo Python 3.6 utilizando el paquete Keras con Tensorflow. El entrenamiento se efectuó en una estación de trabajo con Windows 10, procesador INTEL Core i7, 16 GB de RAM, y GPU nVidia RTX 2080. Para entrenar la red, se utilizó el algoritmo de momento adaptivo acelerado con Nesterov. A fin de acelerar aún más el entrenamiento y además evitar mínimos locales, se utilizó un algoritmo de tasa de aprendizaje cíclico [19] (tasas mínima y máxima de 0.000001 y 0.0001, respectivamente, tamaño de paso igual a 285 o tres veces el número de iteraciones por época). El tamaño de batch se fijó en 1024. El entrenamiento se repitió varias veces con diferentes condiciones iniciales. Al principio se observó que la red presentaba algo de sobre-entrenamiento. Por lo tanto, se optó por regularizar la red usando dropout en las dos capas convolucionales.

5. Resultados

Después de entrenar la red durante 3,500 épocas se encontró que la tasa de aciertos promedio en el conjunto prueba es de 94.92 %.

La tabla 1 muestra una comparación con el estado del arte. A partir de esto, podemos concluir que, en general, la metodología propuesta se compara bien con otros trabajos del estado del arte. Es de resaltar que todos los trabajos anteriores utilizaron bases de datos pequeñas con menos clases y menos variabilidad en las señales EMG.

Tabla 1. Comparación con otros trabajos del estado del arte.

Trabajo	Muestras	Clases	%
Template Matching [14]	350	10	97
DTW [6]	520	26	84.29
DTW [13]	780	26	92.42
Propuesto	108,000	36	94.92

La Fig. 5 muestra la matriz de confusión obtenida para la arquitectura propuesta en el conjunto de prueba. Se puede observar que, en general, todas las clases son bien reconocidas. Sin embargo, hay ciertas clases que presentan más errores, por ejemplo, '0', '3', 'v' y 'x' son confundidas la mayoría de las veces con '6', '5', 'u' y 'y', respectivamente.

Trabajos a futuro podrían abordar el problema del reconocimiento de palabras manuscritas a partir de señales EMG. Igualmente, aún cuando las IRNN no son muy pesadas computacionalmente, se considera que es necesario estudiar su implementación en tiempo real.

Agradecimientos. Los autores agradecen a la Universidad de Guanajuato, en particular a la División de Ingenierías Campus Irapuato-Salamanca por el apoyo brindado.

Referencias

1. De Luca, C.J.: Physiology and Mathematics of Myoelectric Signals. *IEEE Transactions on Biomedical Engineering BME-26*(6), 313–325 (jun 1979)
2. Fukushima, K.: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36(4), 193–202 (apr 1980)
3. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T.: Recent advances in convolutional neural networks. *Pattern Recognition* 77, 354–377 (may 2018)
4. Guigon, E., Baraduc, P., Desmurget, M.: Computational Motor Control: Redundancy and Invariance. *Journal of Neurophysiology* 97(1), 331–347 (jan 2007)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (jun 2016)
6. Huang, G., Zhang, D., Zheng, X., Zhu, X.: An EMG-based handwriting recognition through dynamic time warping. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. pp. 4902–4905. IEEE (aug 2010)
7. Hubel, D.H., Wiesel, T.N.: Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology* 195(1), 215–243 (mar 1968)
8. Ines Chihi, Afef, A., Mohamed, B.: Analysis of Handwriting Velocity to Identify Handwriting Process from Electromyographic Signals. *American Journal of Applied Sciences* 9(10), 1742–1756 (oct 2012)
9. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems* 25, pp. 1097–1105. Curran Associates, Inc. (2012)
10. Lansari, A., Bouslama, F., Khasawneh, M., Al-Rawi, A.: A novel electromyography (EMG) based classification approach for Arabic handwriting. In: *Proceedings of the International Joint Conference on Neural Networks, 2003*. vol. 3, pp. 2193–2196. IEEE (2003)
11. Le, Q.V., Jaitly, N., Hinton, G.E.: A simple way to initialize recurrent networks of rectified linear units (2015)
12. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1(4), 541–551 (dec 1989)
13. Li, C., Ma, Z., Yao, L., Zhang, D.: Improvements on EMG-based handwriting recognition with DTW algorithm. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 2144–2147. IEEE (jul 2013)

14. Linderman, M., Lebedev, M.A., Erlichman, J.S.: Recognition of Handwriting from Electromyography. *PLoS ONE* 4(8), e6791 (aug 2009)
15. Maldarelli, J.E., Kahrs, B.A., Hunt, S.C., Lockman, J.J.: Development of early handwriting: Visual-motor control during letter copying. *Developmental Psychology* 51(7), 879–888 (jul 2015)
16. Okorokova, E., Lebedev, M., Linderman, M., Ossadtchi, A.: A dynamical model improves reconstruction of handwriting from multichannel electromyographic recordings. *Frontiers in Neuroscience* 9, 1–15 (oct 2015)
17. Sinha, N.K., Gupta, M.M., Rao, D.H.: Dynamic neural networks: An overview. In: *Proceedings of the IEEE International Conference on Industrial Technology*. vol. 1, pp. 491–496. IEEE (2000)
18. Slim, M.A., Abdelkrim, A., Benrejeb, M.: An efficient handwriting velocity modelling for electromyographic signals reconstruction using Radial Basis Function neural networks. In: *2015 7th International Conference on Modelling, Identification and Control (ICMIC)*. pp. 1–6. IEEE (dec 2015)
19. Smith, L.N.: Cyclical Learning Rates for Training Neural Networks. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. pp. 464–472. IEEE (mar 2017)
20. Talathi, S.S., Vartak, A.: Improving performance of recurrent neural network with relu nonlinearity (2015)
21. Weng, J., Ahuja, N., Huang, T.S.: Cresceptron: a self-organizing neural network which grows adaptively. In: *International Joint Conference on Neural Networks (IJCNN)*. vol. 1, pp. 576–581 (1992)