

Reconocimiento de Gestos de la Mano Usando Señales Electromiográficas (EMG) e Inteligencia Artificial

Entregable 3 JONATHAN

Protocolo para la evaluación de la exactitud de reconocimiento, la exactitud de clasificación y el costo computacional en tiempo de modelos de reconocimiento de gestos.

Escuela Politécnica Nacional
Universidad de las Fuerzas Armadas ESPE
Universidad Técnica de Ambato

Ph.D. Benalcázar Palacios Marco Enrique
Director del Proyecto



Tabla de contenido

1. Introducción	3
2. Métricas de desempeño	3
Exactitud de clasificación	3
Exactitud de reconocimiento	5
Complejidad y funcionamiento en tiempo real	6
3. Protocolo de evaluación	7
Segmentación manual	7
Validez del vector de predicciones	8
Factor de solapamiento	10
Formato de respuesta para un modelo de reconocimiento de gestos	10
Ejemplo	11
4. Conclusiones	11
5. Bibliografía	12

1. Introducción

En este documento, el problema de clasificación y el problema de reconocimiento son definidos para el caso de reconocimiento de gestos con señales electromiográficas EMG. Basados en esta definición, se presenta un protocolo de evaluación de modelos de reconocimiento de gestos. Este protocolo fue implementado como una librería en Matlab. La librería permite evaluar la respuesta de un modelo de reconocimiento de gestos para una repetición de una señal EMG.

Existen varias investigaciones científicas enfocadas en mejorar modelos reconocimiento de gestos. El desempeño de estos modelos suele medirse considerando la exactitud y el tiempo. Algunas investigaciones [1][2], sin embargo, no presentan bien definidas sus métricas de evaluación o utilizan los términos clasificación y reconocimiento indistintamente. En esos casos, los resultados presentados no son comparables con los de otras investigaciones, pues cada publicación utiliza un método diferente de evaluación. En el área de visión por computadora no existe este problema, pues está definido explícitamente el problema de clasificación y el problema de reconocimiento[3][4]. El problema de clasificación en imágenes es, dada una imagen, retornar la etiqueta de la clase correspondiente. En cambio, el problema de reconocimiento es, dada una imagen, identificar las clases presentes junto a su localización. Es notoria la importancia de una definición explícita para clasificación y reconocimiento en el campo de reconocimiento de gestos. En este trabajo se define 3 métricas de evaluación para el problema de reconocimiento de gestos usando señales EMG. Estas métricas de evaluación son exactitud de clasificación, exactitud de reconocimiento, y complejidad computacional.

2. Métricas de desempeño

Exactitud de clasificación

El problema de clasificación consiste en determinar la categoría correspondiente de cierta observación, dado un conjunto de categorías. En el caso del reconocimiento de gestos usando señales EMG la observación es una señal EMG de cierto gesto. La clasificación involucra, además, el poder distinguir si la observación pertenece o no al conjunto de categorías inicial. Para esto se suele incluir una categoría *de fondo (background)* que representa al universo de clases posibles que no están representados en el conjunto de categorías inicial. En el caso del presente proyecto, a esta clase de fondo la denominamos *noGesture*.

La evaluación de la clasificación compara la clase predicha con la clase verdadera. El resultado de esta comparación es verdadero cuando ambas clases coinciden o falso en caso contrario. En la Figura 1 se muestra un ejemplo de la evaluación de clasificación de una señal EMG.

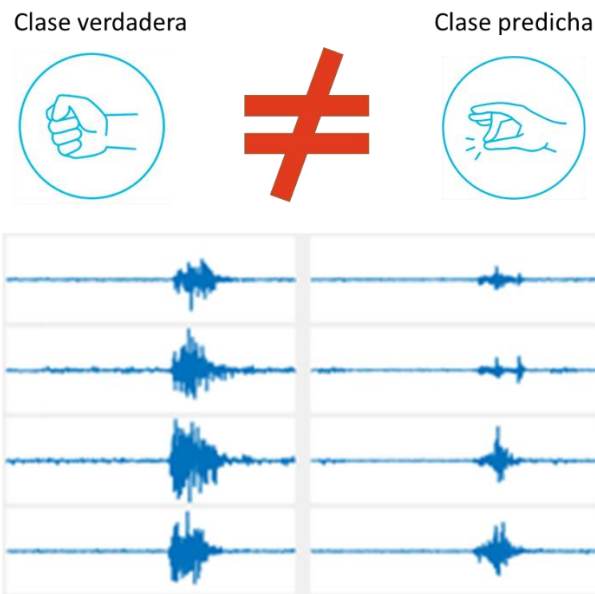


Figura 1. Ejemplo de la evaluación de clasificación para una señal EMG

Los resultados de la clasificación de varias repeticiones suelen ser presentados y analizados en matrices de confusión. Esta es una matriz que compara el conteo de las clases predichas con las clases verdaderas. Estas matrices pueden ser generadas para un usuario o para un conjunto de usuarios. En la Figura 2 se muestra la matriz de confusión del trabajo [5] como ejemplo.

	Targets					Predictions Count % Accuracy (Precision)
	wave in	wave out	fist	open	pinch	
wave in	1435	9	1	3	0	1448 99.1%
wave out	10	1460	1	13	10	1494 97.72%
fist	3	0	1465	2	2	1473 99.45%
open	5	15	2	1416	3	1441 98.27%
pinch	7	4	2	8	1408	1429 98.53%
relax	40	12	29	58	77	216 0%
Targets Count	1500	1500	1500	1500	1500	7500 95.79%
% Accuracy (Sensitivity)	95.67%	97.33%	97.67%	94.4%	93.87%	

Figura 2 Ejemplo de matriz de confusión para análisis de los resultados de clasificación (Matriz de confusión de [5])

En dicho trabajo, se puede observar que se evaluaron 7500 repeticiones (1500 repeticiones por categoría). En las columnas se encuentran las clases verdaderas, y en las filas las clases predichas. En las matrices de confusión, la última columna corresponde a la precisión, y la última fila a la sensibilidad. La **precisión** es el porcentaje del número de veces que la clase fue predicha correctamente del número total de veces predicha. La **sensibilidad** es el porcentaje de clasificaciones correctas del número de repeticiones evaluadas. Por ejemplo, en la matriz de confusión de la Figura 2, El gesto *open* tiene 94.4% de sensibilidad y 98.27% de precisión. Pues se tenía 1500 repeticiones de la clase, de las cuales 1416 fueron correctamente predichas. Sin embargo, esa misma clase fue predicha 1441 veces. Una mayor exactitud de clasificación significa un mejor modelo de reconocimiento de gestos.

Exactitud de reconocimiento

El problema de reconocimiento va más allá de la clasificación. Este requiere identificar los instantes de tiempo en que el gesto fue realizado. En el caso del reconocimiento, entonces, ya no se tiene una predicción sino un vector de predicciones. Cada elemento del vector de predicciones corresponde a la respuesta del modelo de reconocimiento de gestos en un instante de tiempo (sin conocer la información posterior de la señal EMG). En la Figura 3 se muestra un ejemplo de un vector de predicciones. Es importante notar que la predicción corresponde a los puntos desde la última predicción.

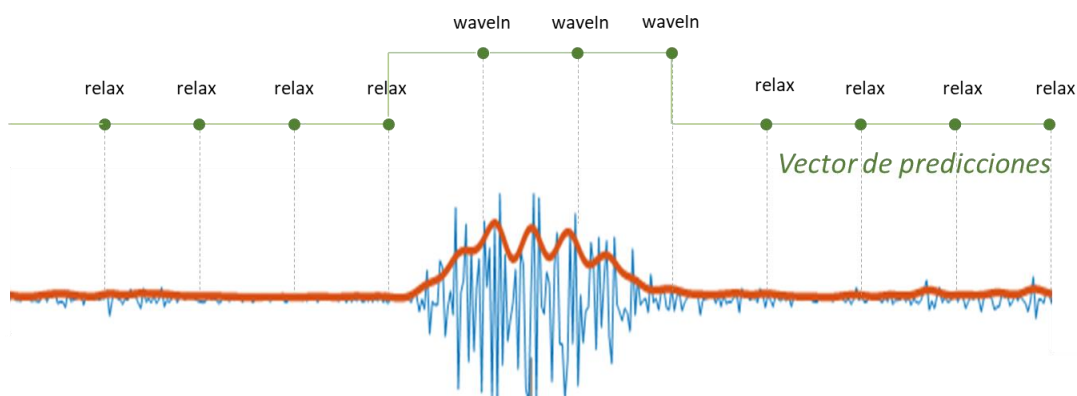


Figura 3 Ejemplo del vector de predicciones para una señal EMG

Para la evaluación del reconocimiento se necesita (además del vector de predicciones) del *ground truth*. El *ground truth* es un vector que etiqueta cada punto de la señal EMG, marcando los puntos en donde se encuentra la actividad muscular. En la Figura 4 se muestra cómo se relaciona el vector de predicciones con el *ground truth* y la señal EMG.

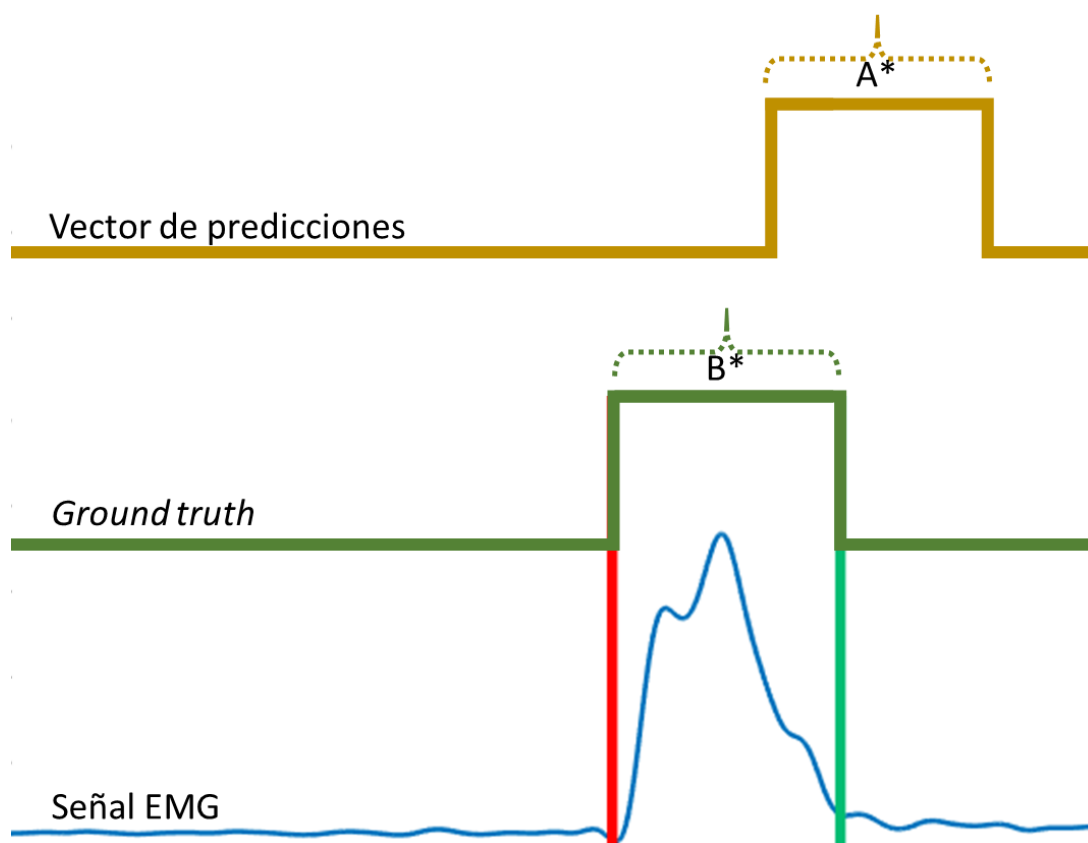


Figura 4 Ejemplo de análisis del reconocimiento

Complejidad y funcionamiento en tiempo real

La tercera métrica de desempeño es la complejidad. Esta métrica se refiere al costo computacional de ejecutar cierto modelo, es decir la memoria requerida y el tiempo de ejecución. Lo recomendable es diseñar modelos con una baja complejidad computacional. Ya que una baja complejidad implica que el modelo se ejecuta en corto tiempo, o que puede ser ejecutado en un computador de gama baja. También implica que el algoritmo tiene un menor consumo energético en sistemas embebidos[6], y que necesita menos datos para entrenamiento (i.e. aprendizaje).

La manera de reducir el costo computacional (i.e. mejorar su tiempo de respuesta) de un modelo es reduciendo su complejidad. En redes neuronales, por ejemplo, significaría utilizar menos neuronas en la capa oculta. Sin embargo, reducir la complejidad del modelo conlleva a limitar su capacidad de aprendizaje. Es decir, una reducción de la complejidad compromete el desempeño [7]. En el diseño de modelos de reconocimiento de gestos se debe llegar a un balance entre exactitud y complejidad del modelo que consiga un buen desempeño para un tiempo de ejecución razonable.

La manera ideal de medir la complejidad computacional es usando la notación asintótica $O(n)$. Esta notación describe matemáticamente el crecimiento (en memoria o en tiempo de ejecución) que tiene un algoritmo para entradas de tamaño cada vez mayor [8]. Esta notación, sin embargo,

requiere del conteo del número de operaciones que ejecuta cierto algoritmo (i.e. sumas, multiplicaciones, asignaciones, condicionales, etc.). Este requerimiento es la causa de que en modelos complejos con varias subetapas (i.e. modelos con filtros, redes neuronales de varias capas, etc) sea impráctica la representación usando la notación asintótica $O(n)$.

En lugar de usar la notación $O(n)$, en las investigaciones de reconocimiento de gestos se suele reportar el tiempo de ejecución del modelo implementado. Si bien el tiempo de ejecución depende directamente de la complejidad del algoritmo, este tiempo también depende del entorno en el que se ejecute. Claramente se obtendrá mejores tiempos al ejecutar el algoritmo en un supercomputador, que en un computador personal. Para remediar esta situación, en las investigaciones se incluye información del computador utilizado (i.e. memoria RAM, número de núcleos del procesador, velocidad del procesador, etc.).

3. Protocolo de evaluación

El protocolo de evaluación se define para una repetición de una señal EMG correspondiente a uno de los gestos definidos. La evaluación requiere de tanto la información verdadera de la señal, como de la respuesta del modelo de reconocimiento en cuestión. La información verdadera del gesto **replInfo** contiene la clase verdadera y el *ground truth* de dicha señal. La respuesta del sistema **response** contiene la clase predicha **class**, el vector de predicciones **vectorOfLabels**, el vector de tiempos correspondiente al vector de predicciones **vectorOfTimePoints**, y el vector de tiempos de procesamiento **vectorOfProcessingTimes**. La evaluación retorna el resultado de la clasificación **classResult**, el resultado del reconocimiento **recogResult** y el factor de solapamiento **overlappingFactor**. En la Sección Formato de respuesta para un modelo de reconocimiento de gestos se describe en profundidad estos campos.

Segmentación manual

Como se mencionó en la Sección de Exactitud de reconocimiento, la evaluación del reconocimiento requiere del *ground truth* de cada señal EMG. Esta información debe ser independiente del modelo implementado. El *ground truth* por ende debe ser incluido en el dataset, y es parte de la adquisición de datos. El proceso de encontrar el *ground truth* de cada señal EMG se denomina segmentación manual. Para este proceso se implementó una interfaz gráfica de segmentación. Cada segmentador debía marcar el punto de inicio y de fin de la actividad muscular por cada repetición de cada gesto de cada usuario. En la Figura 5 se muestra dicha interfaz. La segmentación se realiza por canal, pero se utiliza como referencia primaria a la señal rectificada y filtrada (Panel principal).

En el caso del *noGesto*, el reconocimiento no aplica, pues no existe *ground truth*. Para esas repeticiones, la librería de evaluación retorna vacío en factor de solapamiento y exactitud de reconocimiento.

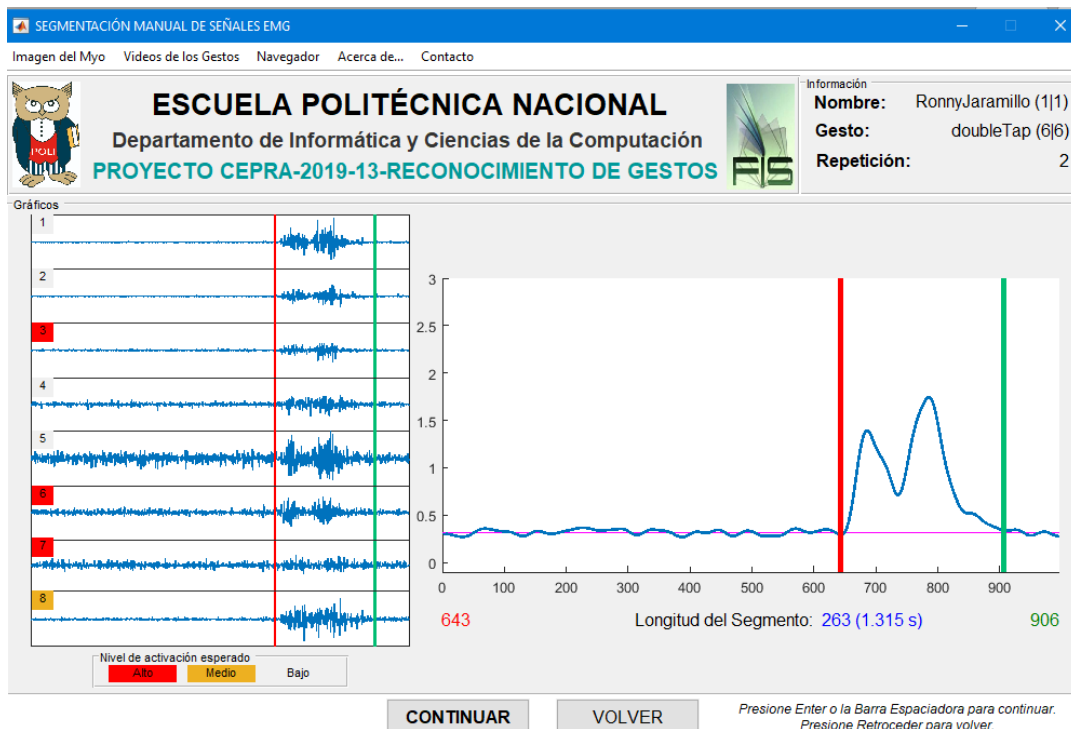


Figura 5 Interfaz de la segmentación manual

Validez del vector de predicciones

Para proseguir evaluando el reconocimiento, se necesita un vector de predicciones valedero. En general, se considera válido el vector de predicciones cuando solo cuando existe una clase predicha sin discontinuidades diferente del reposo. Es decir, todas las predicciones de la actividad muscular deben ser consecutivas. En el caso de una repetición con un inválido vector de predicciones falla en reconocimiento y tiene un factor de solapamiento de cero. En la Figura 6 se muestra un vector de predicciones que es invalido, pues contiene 6 bloques. En la Figura 7 se muestra ejemplos para varios casos de vectores de predicción válidos e inválidos.

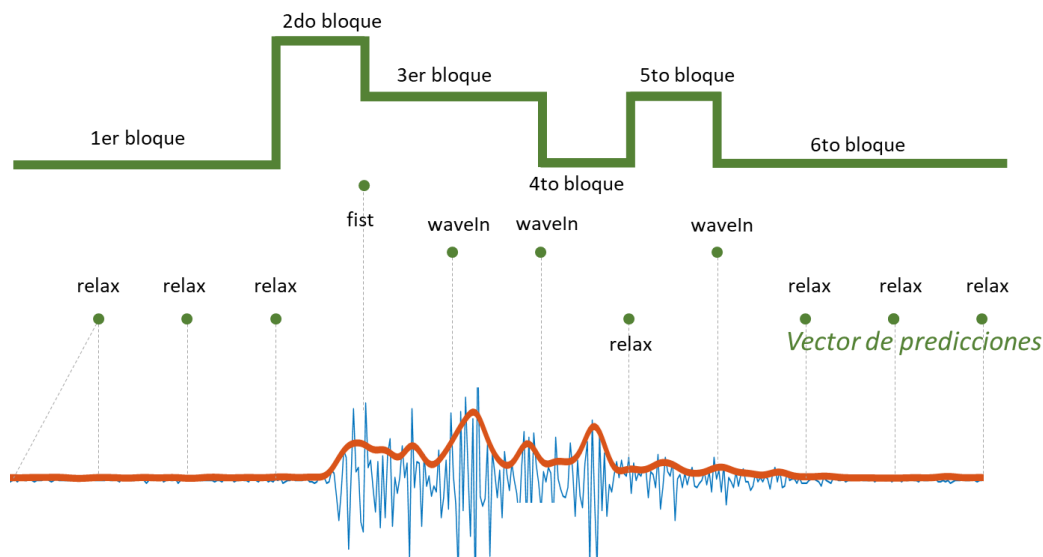


Figura 6 Ejemplo de un vector de predicciones no válido.

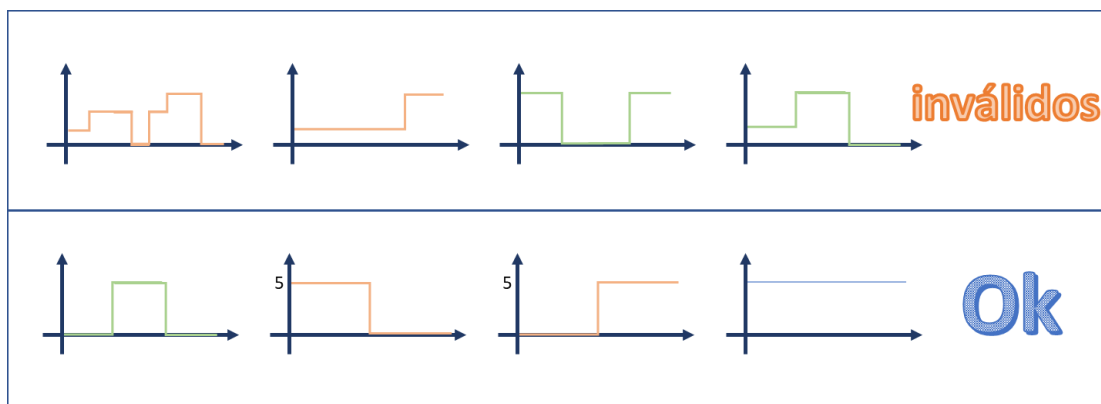


Figura 7 Validez del vector de predicciones

Para un vector de predicciones válido, la librería lo compara con el *ground truth* de dicha señal EMG, como se mostró en la Figura 4. Para esta comparación, la librería internamente genera, a partir del vector de predicciones, un vector equivalente que tiene el mismo tamaño que el *ground truth*. Este vector equivalente es generado iterativamente asignando a cada punto con una estampa de tiempo menor o igual a la predicción actual, caso contrario asigna la etiqueta de la predicción siguiente. Si la señal EMG posee puntos en instantes de tiempo mayores al de la última predicción en el vector de predicciones, estos puntos son asignados la última etiqueta del vector de predicciones. Para mayor claridad, este proceso se muestra en la Figura 8.

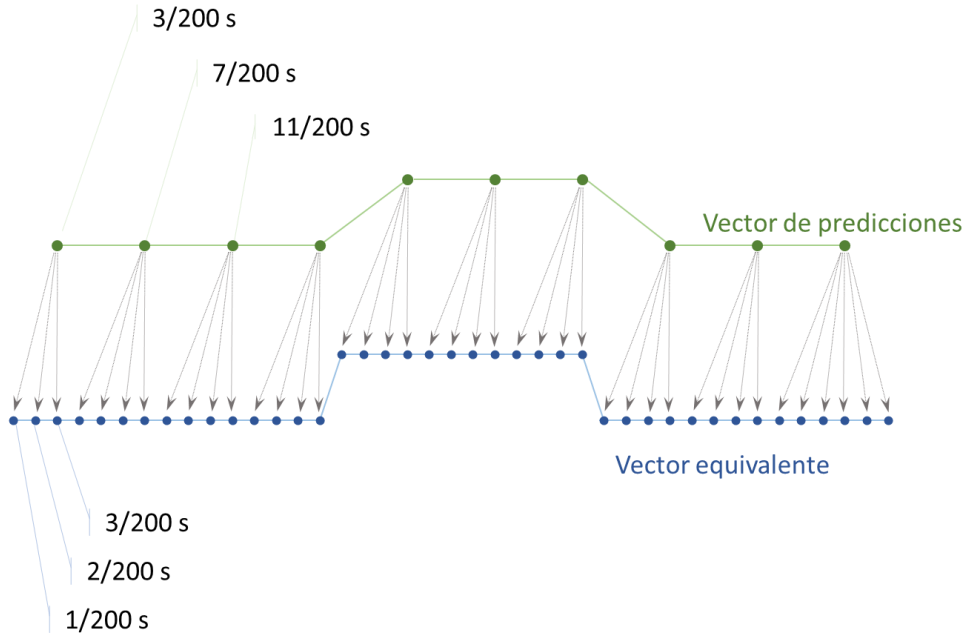


Figura 8 Generación del vector equivalente

Factor de solapamiento

El factor de solapamiento es una medida de la intersección entre el vector de predicciones y el *ground truth*. El objetivo de este factor es determinar el solapamiento desde el que el reconocimiento es correcto. Para el caso de señales electromiográficas, se determinó heurísticamente que el reconocimiento es correcto cuando se supera el 0.25 de factor de solapamiento.

El factor de solapamiento se calcula de la Ecuación (1).

$$overlappingFactor = 2 \frac{|A^* \cap B^*|}{|A| + |B^*|} \quad (1)$$

Donde:

A^* es el conjunto que se obtiene al remover del vector de predicciones aquellos elementos que contienen la etiqueta *noGesture*.

B^* es el conjunto que se obtiene de remover del *ground truth* aquellos elementos que contienen la etiqueta *noGesture*

El reconocimiento es correcto cuando el factor de solapamiento es mayor o igual a 0.25.

Formato de respuesta para un modelo de reconocimiento de gestos

Para aplicar el protocolo de evaluación descrito se implementó una librería en Matlab llamada **evalRecognition**. Esta librería se implementó como una clase que evalúa la clasificación y el reconocimiento de una repetición.

Tabla 1 Descripción de las entradas de la librería *evalRecognition*

Entrada	Tipo	Campos
<i>response</i>	<i>struct</i>	<i>vectorOfLabels</i> : 1-by-w categorical <i>vectorOfTimePoints</i> : 1-by-w double <i>vectorOfProcessingTimes</i> : 1-by-w double <i>class</i> : categorical predictedClass.
<i>replInfo</i>	<i>struct</i>	<i>gestureName</i> : categorical <i>groundTruth</i> : 1-by-w logical con unos en los puntos con actividad muscular. Esta información viene de la segmentación manual. Nota, el noGesture no tiene este campo.

Tabla 2 Propiedades del objeto *evalRecognition*

Propiedades	Tipo	Descripción
<i>classResult</i>	<i>bool</i>	Verdadero cuando la clase predicha coincide con el vector de predicciones.
<i>recogResult</i>	<i>bool</i>	Verdadero cuando el vector de predicciones es válido y el factor de solapamiento es mayor o igual que el umbral (0.25).
<i>overlappingFactor</i>	<i>double</i>	Factor de solapamiento.

Ejemplo

```
r1 = evalRecognition(repOrgInfo,response)
```

```
r1 =
```

```
evalRecognition with properties:
```

```
classResult: 1
```

```
recogResult: 1
```

```
overlappingFactor: 0.7470
```

```
thresholdRecognition: 0.2500
```

4. Conclusiones

En este informe se presentó la importancia de diferenciar entre el problema de clasificación y el problema de reconocimiento aplicado en reconocimiento de gestos con señales EMG. A partir

de esta diferenciación se plantea un protocolo de evaluación de modelos de reconocimiento de gestos.

El objetivo de este protocolo de evaluación es estandarizar el proceso de evaluación de modelos de reconocimiento de gestos. De esta forma, los resultados obtenidos en diferentes investigaciones pueden ser comparados.

El protocolo presentado fue implementado en Matlab por ser el lenguaje de desarrollo utilizado en el proyecto. Se recomienda realizar implementaciones en otros lenguajes de programación utilizados en el área (i.e. Python3). Así, se puede llegar a más grupos de investigación. De manera similar, se recomienda liberar el dataset y que sirva como punto de comparación entre diferentes investigaciones.

5. Bibliografía

- [1] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Trans. Syst. Man, Cybern. Part A Systems Humans*, vol. 41, no. 6, pp. 1064–1076, Nov. 2011, doi: 10.1109/TSMCA.2011.2116004.
- [2] X. Jiang, L.-K. Merhi, Z. G. Xiao, and C. Menon, "Exploration of Force Myography and surface Electromyography in hand gesture classification," *Med. Eng. Phys.*, vol. 41, pp. 63–73, Mar. 2017, doi: 10.1016/j.medengphy.2017.01.015.
- [3] T. De Souza Alves, C. S. De Oliveira, C. Sanin, and E. Szczerbicki, "From Knowledge based Vision Systems to Cognitive Vision Systems: A Review," in *Procedia Computer Science*, 2018, vol. 126, pp. 1855–1864, doi: 10.1016/j.procs.2018.08.077.
- [4] J. Owens and P. Osteen, "Ego-motion and Tracking for Continuous Object Learning: A Brief Survey," *US Army Res. Lab.*, 2017.
- [5] J. A. Zea and M. E. Benalcázar, "Real-Time Hand Gesture Recognition: A Long Short-Term Memory Approach with Electromyography," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1078, pp. 155–167, doi: 10.1007/978-3-030-33614-1_11.
- [6] J. Mantilla-Brito, D. Pozo-Espín, S. Solórzano, and L. Morales, "Embedded System for Hand Gesture Recognition Using EMG Signals: Effect of Size in the Analysis Windows," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1078, pp. 214–225, doi: 10.1007/978-3-030-33614-1_15.
- [7] R. Gilad-Bachrach, A. Navot, and N. Tishby, "An information theoretic tradeoff between complexity and accuracy," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 2003, vol. 2777, pp. 595–609, doi: 10.1007/978-3-540-45167-9_43.
- [8] T. H. Cormen, *Introduction to algorithms*. MIT Press, 2009.