

Sistema de reconocimiento de patrones de sustancias químicas cerebrales basado en minería de datos

Junior Altamiranda¹, José Aguilar¹ y Luis Hernandez²

¹ Universidad de Los Andes, Facultad de Ingeniería,
Departamento de Computación, CEMISID, Mérida,
Venezuela

² Universidad de Los Andes, Facultad de Medicina,
Laboratorio de Fisiología, Mérida,
Venezuela

{altamira, aguilar}@ula.ve, hernandenator@gmail.com

Resumen. En este trabajo se presenta un Sistema de Minería de Datos para interpretar los cambios bioquímicos que ocurren en los cerebros de los roedores. El análisis manual de tales experimentos es impráctico debido a la naturaleza voluminosa de los datos que son generados y la naturaleza tediosa del análisis de su significado, por lo que información importante se pierde. Por esta razón, se ha diseñado un Sistema de Minería de Datos el cual contiene varios pasos (pre-procesamiento de los datos, clasificación de la información recolectada, etc.), y utiliza la Red Neuronal Artificial basada en la Teoría de Resonancia Adaptativa en la realización de algunos de ellos. En este artículo se describe el Sistema, y se prueba su funcionamiento estudiando los neurotransmisores glutámato y aspartato de muestras extraídas del cerebro de roedores.

Palabras clave. Minería de Datos, Bioinformática, Redes Neuronales Artificiales, Teoría de Resonancia Adaptativa.

Pattern Recognition System Based on Data Mining for Analysis of Chemical Substances in Brain

Abstract. This paper presents a data mining system for analyzing biochemical changes in the brain of rodents. Manual analysis of such experiments is impractical due to a huge volume of generated data and tedious analytical procedures; as a result, important information is lost. Addressing this issue, our paper proposes a data mining system consisting of several steps (pre-processing, data classification, etc.). In some of the steps we apply the artificial neural network based on

the adaptive resonance theory. This paper describes the proposed system and experiments performed to validate it. In the experiments, glutamate and aspartate neurotransmitters in samples extracted from rodent brains were analyzed.

Keywords. Data mining, bioinformatics, neural network, adaptive resonance theory.

1. Introducción

Las investigaciones en ciencias biomédicas están generando un enorme volumen de información biológica, que también, cada vez es más compleja. El excesivo aumento, tanto en el número como en el tamaño de las bases de datos provenientes de la biología molecular, ha provocado que infinita cantidad de información de la que se dispone, exceda lo que puede ser procesado y entendido por el ser humano. Estas contienen grandes volúmenes de datos, que ocultan una enorme cantidad de información útil, difícil de descubrir [1]. Las herramientas computacionales, y en particular, la inteligencia artificial, son cruciales para interpretar grandes volúmenes de datos. En ese sentido, hay toda un área dentro de la biología, llamada biología computacional, la cual estudia el uso de técnicas de procesamiento de la información, como la minería de datos, para el análisis y la extracción de conocimiento desde fuentes de datos biológicos. Este campo interdisciplinario comprende la investigación y el desarrollo de

herramientas útiles para llegar a entender, entre otras cosas, el flujo de información de los genes a las estructuras moleculares, la función bioquímica de las proteínas, su papel en sistemas complejos como el cerebro, su influencia en las enfermedades y en la salud, entre otras cosas [2, 3, 4, 5, 6]. En particular, la minería de datos ha contribuido a recolectar, organizar y entender la información genética y de las proteínas, para emplearla en diferentes ámbitos (por ejemplo, en el análisis de las causas de ciertas enfermedades, el genoma humano, desarrollo de fármacos, etc.) [7].

La minería de datos usa entre sus técnicas, los sistemas inteligentes. Los sistemas inteligentes constituyen el campo de la computación en el que se estudian y desarrollan algoritmos que implementan o emulan distintos modelos de aprendizaje, comportamientos de ciertos sistemas biológicos, entre otras cosas, y su aplicación a la resolución de problemas complejos. Entre los problemas abordados en este campo está el de inducir conocimientos a partir de datos o ejemplos. Esto resulta una alternativa de solución a problemas que no pueden ser resueltos mediante algoritmos tradicionales, tales como métodos estadísticos, modelos matemáticos, entre otros. Todos estos métodos son esencialmente cuantitativos. En contraposición, los métodos basados en sistemas inteligentes están orientados, principalmente, hacia el desarrollo de descripciones simbólicas de los datos, que puedan caracterizar uno o más grupos de conceptos, diferenciar distintas clases, seleccionar los atributos más representativos de grupos de datos, ser capaces de predecir secuencias, etc. Estos métodos son esencialmente cualitativos, lo cual permite el descubrimiento de patrones en estructuras de información (en particular, en bases de datos de gran tamaño) [8].

Para realizar este trabajo, se usan las bases de datos del Departamento de Fisiología de la Facultad de Medicina de la Universidad de los Andes que contienen resultados de los experimentos sobre el cerebro de ratas, con el fin de entender las interacciones que suceden en el cerebro cuando una rata realiza una actividad específica (es decir, se alimenta, camina, corre,

duerme, etc.). En específico, el problema consiste en analizar las sustancias bioquímicas presentes en el cerebro, para descubrir los patrones que se derivan de ellas, los cuales pueden asociarse a comportamientos normales o anormales, estos últimos vinculados a enfermedades cerebrales. Al analizar las sustancias individualmente, se podrán asociar las mismas a estados normales o anormales del cerebro, estos últimos vinculados a enfermedades cerebrales. Por otro lado, al asociar los patrones a actividades de un roedor (comer, dormir, etc.), permitirá, en otro contexto, descubrir las actividades realizadas por los roedores. Esos son algunos de los posibles usos del sistema propuesto en este trabajo.

Los procesos de recolección o generación de información para el reconocimiento de las sustancias bioquímicas presentes en el cerebro, producen volúmenes de datos almacenados en bases de datos, que superan las capacidades humanas para analizarlos [9, 10]. Esta limitación se debe a varios factores, entre los cuales tenemos: la disponibilidad de tiempo, la incapacidad de relacionar grandes volúmenes de datos, entre otros. El valor real de los datos reside en la información que podamos extraer de ellos [31], en nuestro caso: información que ayude a reconocer patrones cerebrales, a mejorar la comprensión de lo que ocurre en el cerebro de un roedor, etc. Es eso lo que tratamos de explotar en este trabajo. Los experimentos extraen las sustancias químicas del cerebro del roedor a través de un capilar (tubo de diámetro interno fino - entre 2 μm y 150 μm) que es introducido en la cabeza de éste. Luego, ese líquido es procesado utilizando la electroforesis capilar, la cual, es una técnica de cromatografía que permite la separación rápida y eficiente de compuestos con carga iónica. La detección de las sustancias se realiza mediante la lectura de la absorbancia en la región del ultravioleta-visible. Con los datos obtenidos, se genera una gráfica que representa el conjunto de interacciones que ocurren en el cerebro según las sustancias que actúan, para lo cual se utiliza un sistema computacional llamado ONICE [11]. Así, las sustancias químicas se encuentran representadas en los electroferogramas obtenidos en los experimentos realizados a los roedores.

Esa metodología ha mostrado ser muy buena en la extracción de la información de las sustancias bioquímicas presentes en el cerebro cuando realiza una actividad, por ello su consideración como la fuente de información para la caracterización de los patrones presentes en el cerebro. Por lo tanto, es necesario para el análisis de los electroferogramas obtenidos en los experimentos realizados a los roedores, la automatización e implementación de técnicas de procesamiento de señales y de reconocimiento de patrones, para analizar de forma confiable y rápidamente el gran volumen de datos, lo cual representa el reto de este trabajo.

Por ello, se propone construir un sistema para la extracción de conocimiento desde esa gran cantidad de datos que se obtienen de los experimentos con las ratas, para encontrar relaciones y patrones de los flujos bioquímicos de las neuronas, que nos permitan entender el comportamiento de las interacciones moleculares que ocurren en el cerebro. Para realizar el sistema de extracción de conocimiento se utiliza la minería de datos basada en sistemas inteligentes, parecido al enfoque propuesto en [33] de minería de datos sobre series de tiempo, útil en este caso para resolver el problema de reconocimiento de patrones planteado en este artículo.

2. Problema

Por medio de la electroforesis capilar, se pueden extraer las sustancias químicas (neurotransmisores) que actúan en el cerebro en un momento dado, para su posterior análisis [9], [10]. En particular, en el Departamento de Fisiología de la Facultad de Medicina de la Universidad de los Andes se realizan estos experimentos en los roedores, para determinar los cambios bioquímicos en sus cerebros cuando realizan ciertas actividades [12].

La electroforesis capilar genera electroferogramas (gráficos) que describen las sustancias químicas que actúan en estos experimentos [13], que se deben aclarar, entender e interpretar. La electroforesis significa usar la electricidad para llevar o transportar algo

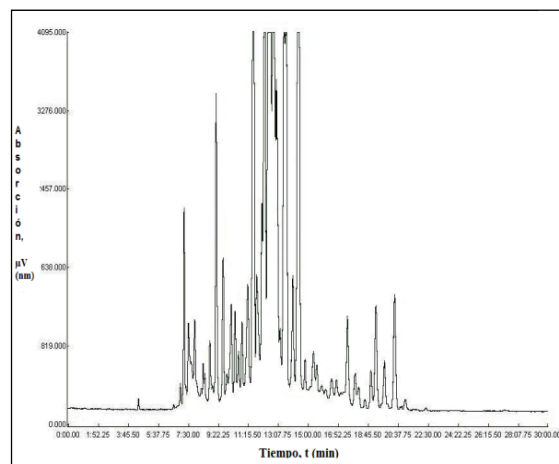


Fig. 1. Ejemplo de electroferograma

de un lado a otro [14]. La electroforesis capilar es una técnica de separación de moléculas, que utiliza un campo eléctrico para transportar sustancias químicas de un extremo al otro en el interior de un tubo de diámetro interno fino — entre 2 μm y 150 μm — llamado capilar [9], [10]. Así, la electroforesis capilar es una técnica de separación que, basada en la química analítica, proporciona un mapa detallado (señal en el tiempo) de las sustancias que conforman una muestra bioquímica, representando las sustancias por picos, donde la amplitud de los picos está relacionada con la concentración de las sustancias en la muestra. Las muestras extraídas son analizadas en un equipo de electroforesis capilar, obteniendo un conjunto de electroferogramas (ver figura 1), los cuales contienen un conjunto de picos que deben ser estudiados, cada uno de ellos representando una sustancia química que está actuando en un momento dado en el cerebro de un individuo mientras este realiza una actividad. Cada gráfica (electroferograma) representa un pequeño periodo de tiempo (5, 10, 20, 30 min.) de un experimento global, en el que el individuo está realizando una actividad.

La espectrofotometría es una de las técnicas experimentales más utilizadas para la detección específica de moléculas de distinta naturaleza (biomoléculas, contaminantes, etc.), y estados de agregación (sólido, líquido, gas). El fundamento

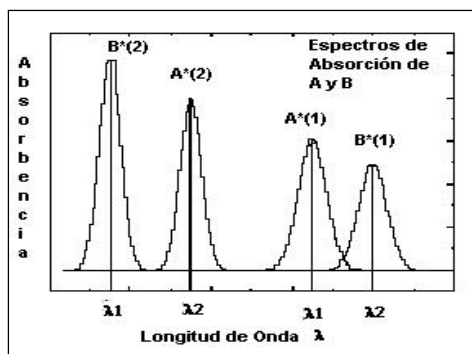


Fig. 2. Espectro de absorción de dos moléculas distintas A y B

físico-químico de la espectrofotometría está relacionado con la capacidad de las moléculas de absorber energía luminosa y almacenarla en forma de energía interna. Como consecuencia, el espectro de absorción, es decir, la luz absorbida en función de la longitud de onda, constituye una verdadera señal de identidad de la molécula (dos moléculas distintas presentarán espectros de absorción distintos, ver figura 2) [15, 28].

Actualmente, los datos de estos experimentos se almacenan en una base de datos, para su subsecuente interpretación. El volumen de estos datos, su complejidad, y el tiempo consumido en su análisis manual, causa inevitablemente que una cierta parte de la información que se encuentra en estos experimentos se pierda en el análisis.

De esta manera, en este trabajo se desea estudiar los comportamientos de los picos en los electroferogramas, y reconocer las sustancias bioquímicas que representan. Seguidamente, al descubrir las sustancias bioquímicas presentes en las muestras, y al asociar las muestras a actividades, se desea descubrir los patrones de sustancias bioquímicas presentes en cada actividad. Así, se descubre todo el conocimiento que se encuentra implícito en los electroferogramas.

El poder interpretar los gráficos generados por la electroforesis capilar, identificando las sustancias químicas que intervienen en las interacciones químicas, reconociendo patrones, entre otras cosas, es un área muy importante para comprender el comportamiento del cerebro,

y en particular, las enfermedades cerebrales [16, 17, 18, 19]. En ese sentido, las dos tareas que debe realizar el sistema de extracción de conocimiento son:

1. El descubrimiento de los picos presentes en las muestras, ya que cada uno de ellos representa una sustancia química (neurotransmisor). Esto permitirá conocer las sustancias químicas que actúan en una muestra.
2. El descubrimiento de los patrones de las actividades descritas en las muestras. Los patrones son expresiones que describen la secuencia de sustancias químicas que aparecen en una actividad dada del individuo.

Para realizar el sistema de extracción de conocimiento se utiliza la minería de datos basada en sistemas inteligentes, pero específicamente vinculada a temas de procesamiento de imágenes médicas [32]. Ella permite hacer la descripción simbólica de los picos en las bases de datos de los electroferogramas, para caracterizar las distintas clases de sustancias bioquímicas, como también el descubrimiento de los patrones de cada electroferograma.

3. Diseño del Sistema de Minería de Datos

El Sistema de Minería de Datos que se presenta realiza la clasificación y análisis automático de los picos de las gráficas, y la clasificación de las muestras de los roedores en una actividad dada: se alimenta, camina, corre, duerme, etc. El Sistema de Minería de Datos tiene tres fases: a) La primera fase genera una representación intermedia; b) la segunda fase almacena dichos datos en un repositorio de datos históricos; y c) la última fase realiza el reconocimiento y construcción de patrones (ver figura 3).

El Sistema de Minería de Datos requiere de una etapa previa de pre-procesamiento. En la etapa de pre-procesamiento, los experimentos y las gráficas se transforman, para que facilite su posterior análisis. Esa representación está

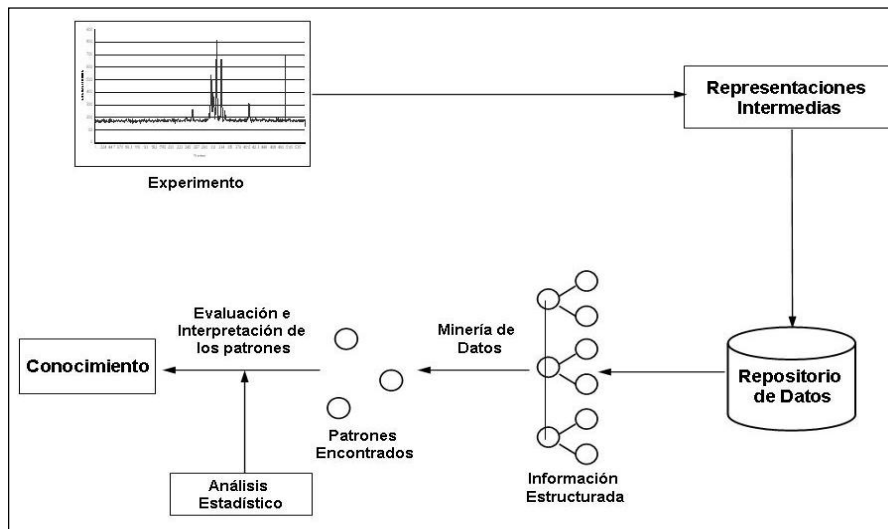


Fig. 3. Diseño del Sistema de Minería de Datos

compuesta para cada pico de la gráfica por: su área, altura, punto de inicio, punto de fin, y su ancho.

Esta estructura es la que se almacena en la segunda fase del sistema de minería de datos, en un repositorio de datos que contiene toda la información de los picos.

En la fase de reconocimiento y construcción de patrones, las representaciones intermedias de los picos se analizan con el objetivo de clasificarlos en sustancias químicas, para luego determinar el patrón de la gráfica y clasificarlo en una actividad específica. De esta manera, el sistema realiza las siguientes tareas:

1. Extraer del electroferograma los picos, con su área y altura.
2. Construir un repositorio de datos, para almacenar los valores extraídos de las gráficas que representan cada muestra.
3. Clasificar la información almacenada sobre los picos, en el repositorio de datos.
4. Construir los patrones de las actividades de las muestras, según el número de picos y las sustancias que representan.

A continuación hablaremos de las fases más relevantes.

3.1. Pre-procesamiento de los datos

En la etapa de pre-procesamiento los datos de los electroferogramas se filtran y se separan en sectores, para así obtener los picos los cuales representan las sustancias químicas presentes en la muestra, para ello se realizan dos pasos.

3.1.1. Filtrado de la señal de la electroforesis capilar

El filtro Savitzky-Golay, es utilizado en sistemas de datos que contienen ruido en la señal, permite suavizar la señal original sin perder precisión, esta es la razón de su escogencia. El objetivo del filtro es sustituir los datos originales por datos que permitan suavizar la señal original, sin producir distorsión de fase, preservando la anchura y la altura de los picos de las señales a las que se aplican. El filtro producirá una señal más suave que la señal original, pero con el mismo número de puntos [20]. El grado de suavización está controlado por el número específico de datos usados en el ajuste y por el orden del polinomio.

El procedimiento es el siguiente: para cada punto nuevo se toma el promedio ponderado usando el punto y sus vecinos (n) de la señal original, formando un polinomio de $2n+1$ puntos (incluyendo el punto a filtrar), donde n debe ser

mayor o igual al grado del polinomio que se va a aproximar el punto filtrado.

La descripción matemática del filtro Savitzky-Golay es la siguiente. Si tenemos t_i datos con ruido, los correspondientes datos suavizados $t_{s,i}$, deben ser expresados como sigue [30]:

$$t_{s,i} = \sum_{j=-n}^n c_j t_{i+j},$$

donde c_j son coeficientes. Vamos a mostrar como los coeficientes c_j pueden ser determinados para el caso específico de ajustes polinómicos de segundo orden. Supongamos que la función va a ser suavizada en los alrededores de un punto i , y puede ser aproximada por el siguiente polinomio:

$$a_0 + a_1(j-i) + a_2(j-i)^2, \quad j = -n+i, \dots, n+i,$$

donde los coeficientes a_0 , a_1 y a_2 son determinados por el ajuste de mínimos cuadrados. Por lo tanto, la siguiente función tiene que ser minimizada:

$$I = \sum [t_j - a_0 - a_1(j-i) - a_2(j-i)^2]^2.$$

Como ejemplo, el coeficiente a_0 es

$$a_0 = \sum_{j=-n}^n \frac{S_4 - j^2 S_2}{S_0 S_4 - S_2^2} t_{i+j}, \quad S_l = \sum_{j=-n}^n j^l.$$

En la figura 4 podemos ver, por ejemplo para $n = 2$, el cálculo de un punto filtrado, para el cual se toman los dos valores anteriores (t_{-2} , t_{-1}) a él y los dos valores posteriores (t_1 , t_2), con ello se obtiene el punto (t_0) sobre la curva utilizando una aproximación por polinomios. [20].

Para utilizar el Filtro Savitzky-Golay es necesario definir los siguientes parámetros:

1. El conjunto de datos de entrada.
2. El grado del polinomio al cual se van a aproximar los puntos filtrados.
3. El número de puntos vecinos al punto a filtrar, este valor debe ser impar.

En el sistema, para filtrar la muestra extraída del roedor se utilizan dos filtros de Savitzky-Golay, que van a actuar de manera secuencial sobre los datos, ya que es necesario suavizar lo

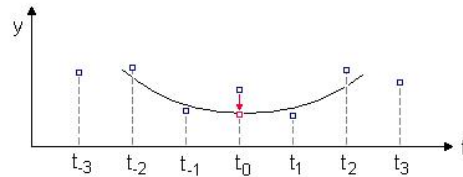


Fig. 4. Filtro Savitzky-Golay sobre un punto

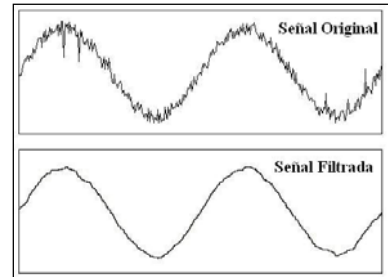


Fig. 5. Filtro de Savitzky-Golay aplicado al electroferograma

mayor posible la señal para poder extraer los picos de ésta (ver figura 5). Los parámetros utilizados son los siguientes:

Filtro 1: Vector de datos originales, orden del polinomio 0 (función constante), número de puntos vecinos 15.

Filtro 2: Vector de datos formado por la salida del filtro 1, orden del polinomio 0 (función constante), número de puntos vecinos 9.

Los coeficientes utilizados fueron obtenidos a través de pruebas realizadas a los electroferogramas mediante aproximaciones de mínimos cuadrados en las inmediaciones del punto. Por otro lado, el número de puntos vecinos y el grado del polinomio son los factores que condicionan el proceso de suavización.

3.1.2. Extracción de los picos de la muestra

Permite dividir la gráfica en sectores, donde cada uno de ellos representa el comienzo y el final de un pico, los sectores importantes son los que poseen mayor ancho y altura ya que debido a la electrofotometría es donde se concentra un electrotransmisor específico, además se reflejan en un espacio específico de la gráfica. Para ello

se construyó un algoritmo que recorre el conjunto de datos filtrados del electroferograma y coloca el valor cero (0) al comienzo y fin de un sector, para así dividir la gráfica en varios sectores y luego extraer los picos. Los sectores representan la sustancia química presente en ese instante de tiempo en la muestra.

Este proceso se realiza en dos pasos:

Paso 1: Extracción de los puntos de los picos.

Paso 2: Colocar el valor cero (0) para dividir los sectores del electroferograma.

Los pasos en detallado se presentan en Algoritmo 1, donde A es un vector vacío de tamaño igual a los datos; p es un vector que contiene los datos del electroferograma, n es un contador inicializado en cero (0); i es un contador inicializado en uno (1).

A través del algoritmo anterior obtenemos los picos de la señal, los cuales representan la presencia de una sustancia química en ese instante de tiempo. En la figura 6 se observa como la señal es dividida en sectores por medio de una barra vertical (representa el valor 0 en el conjunto de datos), donde cada sector representa el comienzo y el fin de un pico.

Después de realizar estos procedimientos se procede a procesar la muestra y los picos para obtener la representación intermedia. Este es un punto crítico del Sistema de Minería de Datos, ya que si no se define adecuadamente la representación intermedia los resultados generados en el proceso de Minería de Datos pueden ser incoherentes. La extracción de las características de la gráfica, consiste en sacar algunas medidas numéricas de los datos originales. Particularmente, la representación intermedia debe contener las características extraídas de los picos de la gráfica, ellas son: el área, la altura, el punto de inicio del pico, el punto de fin del pico y el ancho del mismo (ver figura 7).

Las representaciones intermedias obtenidas de los picos de las muestras, son almacenadas en un repositorio de datos (ver estructura del mismo en la figura 8).

Por otro lado, sobre los picos obtenidos se hace un análisis estadístico para caracterizar las propiedades generales de los mismos (media, varianza, desviación estandar, etc.); esto va a

```

Repita mientras (n ≠ tamaño([p] - 3))
Si((p(i + 1) > p(i)) Y (p(i + 2) ≥ p(i + 1)))
  A(i) = p(i)
Fin Si
Si((p(i + 1) < p(i)) Y (p(i + 2) ≤ p(i + 1)))
  A(i) = p(i)
Fin Si
Si(i ≠ 1)
  Si((p(i-1) < p(i)) Y (p(i-1) < p(i)))
    A(i) = p(i)
  Fin Si
  Si((p(i-1) < p(i)) Y (p(i) ≤ p(i + 1)))
    A(i) = p(i)
  Fin Si
  Si((p(i-1) > p(i)) Y (p(i) < p(i + 1)))
    A(i) = p(i)
  Fin Si
Fin Si

Si ((i > 2) Y (p(i-2) ≤ p(i-1))
  Y (p(i-1) >= p(i))
  Y (p(i) >= p(i + 1))
  Y (p(i + 1) ≤ p(i + 2))
  Y (p(i + 2) ≥ p(i + 3)))
  A(i) = p(i)
Fin Si

//División de los sectores por medio de cero (0)
A(i) = 0
n = n + 1
i = i + 1
fin repita mientras
repita para j = i : tamaño (p)
  A(j) = 0
fin repita para
Regresar A
Fin

```

Algoritmo 1. Extracción de picos de muestra

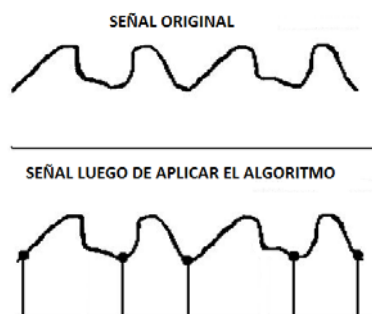


Fig. 6. Extracción de los picos

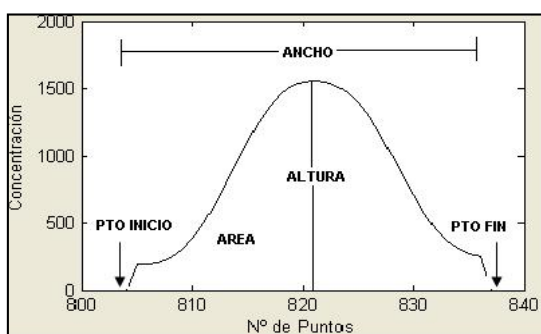


Fig. 7. Datos que se obtienen de los picos para la representación intermedia

permitir analizar las tendencias que tienen las sustancias químicas presentes en una muestra.

Las variables a estudiar son la altura y el área de los picos (cada pico representa una clase de sustancia química), de tal manera de determinar la concentración de estas en las muestras extraídas del cerebro del roedor. Así, cada valor estadístico de área y altura de cada clase de sustancia química puede ser usado para compararlo con los obtenidos en cada muestra, para analizarlos y llegar a conclusiones.

3.2. Reconocimiento y construcción de patrones

En esta fase las representaciones intermedias de los picos son analizadas, con el objetivo de clasificarlas en sustancias químicas. Además, una vez identificados los picos, trata de definir los patrones de sustancias químicas presentes en el cerebro para cada actividad realizada por el roedor.

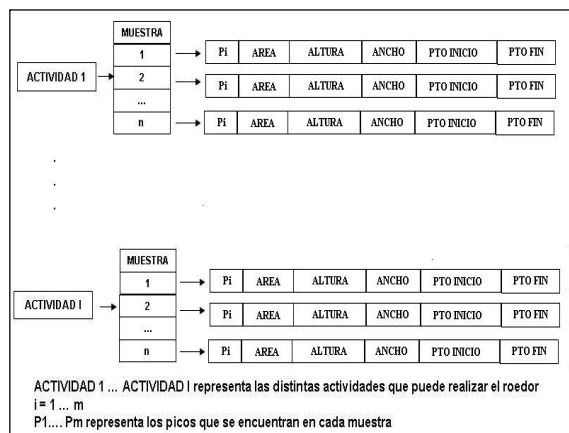


Fig. 8. Estructura del repositorio de datos

Esta fase automatiza el proceso de identificación de los picos en las muestras, el análisis de los parámetros de área y altura obtenidos de éstos, y clasificar las muestras según las sustancias químicas presentes en una actividad. Se utiliza como técnica inteligente las Redes Neuronales Artificiales [21, 22].

La teoría de las Redes Neuronales Artificiales emulan las redes neuronales biológicas. Pueden ser consideradas como un sistema de procesamiento de información con características como aprendizaje a través de ejemplos, adaptabilidad, robustez, capacidad de generación y tolerancia a fallas. Para el desarrollo de este trabajo se usa el modelo neuronal ART (Adaptive Resonance Theory). La red ART utiliza los dilemas de estabilidad y de plasticidad del aprendizaje. Estos dilemas plantean los siguientes interrogantes: ¿Cómo una red puede aprender nuevos patrones? (plasticidad del aprendizaje); ¿Cómo una red puede retener los patrones previamente aprendidos? (estabilidad del aprendizaje).

En respuesta a estos dilemas, Grossberg y Carpenter desarrollaron la denominada Teoría de Resonancia Adaptativa. Esta teoría se aplica a sistemas competitivos (redes con aprendizaje competitivo), en los cuales, cuando se presenta cierta información de entrada sólo una de las neuronas de salida de la red (o un grupo de ellas) se activa. Esta neurona recibe el nombre de vencedora.

La Teoría de Resonancia Adaptativa se basa en la idea de hacer resonar la información de entrada con los representantes o prototipos de las categorías que reconoce la red. Si entra en resonancia con alguno, es suficientemente similar, la red considera que pertenece a dicha categoría y únicamente realiza una pequeña adaptación del prototipo almacenado representante de la categoría, para que incorpore algunas características del dato presentado. Cuando no resuena con ninguno (no se parece a ninguno de los existentes, recordado por la red hasta ese momento), la red se encarga de crear una nueva categoría con el dato de entrada como prototipo de la misma. Así, las clases o categorías deben ser creadas por la propia red (aprendizaje no supervisado) a través de las correlaciones entre los datos de entrada. Estas redes suelen denominarse ART y ART2. Ambas difieren en la naturaleza de la información que se presenta de entrada. La red ART trabaja con vectores de entrada binarios, mientras que ART2 es capaz de procesar información continua o analógica [22, 23, 24, 25]. Particularmente, el Sistema de Minería de Datos consta de dos redes neuronales ART2 para estas tareas.

3.2.1. Red ART2 para clasificar los picos de una muestra

Esta red neuronal permitirá la clasificación y reconocimiento de los picos en la muestra. En este caso, un patrón es una curva que representa una sustancia química que actúa en el cerebro del roedor. La red neuronal ART2 debe ser capaz de reconocer una sustancia química por la posición que ocupa el pico en la gráfica, por lo tanto se utiliza como entrada a la red el punto de inicio y punto de fin de cada pico; y se obtiene como salida de la red la clase a la que pertenece el pico (ver figura 9). En esta Figura P1, P2, P3, P4, P5, P6, P7, P8, P9 representan la salida de la red neuronal ART2 y, por lo tanto, las clases a los que pertenecen los picos. Cada uno de ellas representa una sustancia química diferente presente en la muestra.

Los valores de las neuronas de entrada deben ser normalizados (ya que como se dijo anteriormente, esta red neuronal ART2 acepta entradas entre 0 y 1). Para ello se divide el punto

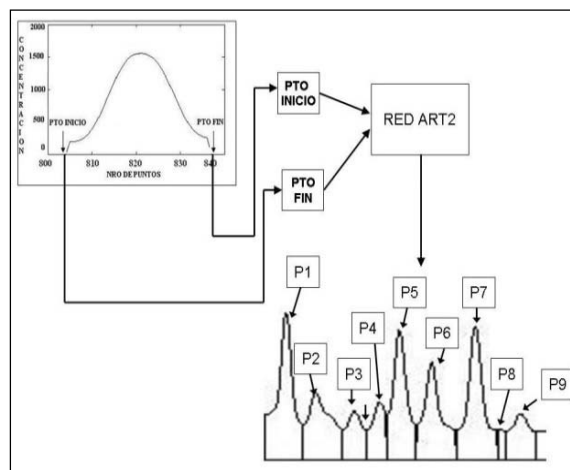


Fig. 9. Diagrama de la red neuronal ART2 para el reconocimiento de los picos de la muestra

de inicio y punto final entre número de datos presentes en la muestra, de esta manera se tiene un valor absoluto entre 0 y 1 para cada uno de ellos; y así poderlos presentar como entradas a la red neuronal ART2.

La arquitectura de la red neuronal ART2 será la siguiente:

1. Número de Neuronas de Entrada = 2 (punto inicio pico, punto fin pico).
2. Número de Neuronas de Salida = m (número de clases reconocidas en un momento dado).
3. Número iteraciones máximo = 2000.
4. Tasa de Aprendizaje (β) = 0.9.
5. Parámetro de Vigilancia (ρ) = 0.99.
6. Bias (θ) = 0.1.

Cada neurona S de la capa de salida compete con las demás de esta capa hasta que sólo una pertenece activa toma el valor (1) y el resto de las neuronas toman el valor (0). La neurona vencedora es aquella S_j que verifica una mínima diferencia (distancia euclidiana) entre el patrón de entrada y los pesos de las conexiones entre esta neurona y las de la capa de entrada. La función de activación de una neurona de salida S_j estará dada por la ecuación (1):

$$S_j = \begin{cases} 1 & \text{Min } |E_k - W_j|, \\ 0 & \text{Resto} \end{cases} \quad (1)$$

donde S es una neurona de la capa de salida, E_k son pesos de las neuronas capa de entrada (valores de punto de inicio y punto de fin del pico) y W_j pesos de la neurona j

La función de activación trabaja de la siguiente manera:

1. Se calcula la diferencia absoluta entre los datos de entrada E_k (punto de inicio, punto de fin) y los valores de los pesos de cada una de las neuronas de salida (representan las clases almacenadas) de la red neuronal.
2. El valor mínimo obtenido en (1), hace que se active la clase a la que pertenece los datos almacenados en ésta.
3. A S_j se le asigna el valor uno.

Para que una entrada E_k a la red neuronal ART2 pertenezca a una clase se debe cumplir con la siguiente ecuación (2):

$$|\text{Punto Inicio} - w_{j1} + \text{Punto Fin} - w_{j2}| < 1 - \rho \quad (2)$$

donde w_{1j} es el peso 1 de la neurona j; w_{2j} es el peso 2 de la neurona j; ρ representa el parámetro de vigilancia. Luego de realizar pruebas se asignó el valor de 0,99 a ρ ; esto implica que la clasificación es muy exigente, los valores de inicio y fin de dos picos deben ser muy cercanos para que estos sean agrupados en una misma clase.

Como podemos ver, para que una neurona de salida se active el valor de la diferencia en la ecuación (2) debe ser menor 0.01, según el valor dado a ρ esto se debe a que es necesario que cada pico de la gráfica se parezca lo más posible a las clases almacenadas por la red, por lo tanto, la red neuronal ART2 va a tener un gran número de neuronas de salida. Así, si el valor resultante es mayor que 0.010 se formará una clase nueva que almacenará los valores de la entrada (punto Inicio y punto de fin de un pico).

Si varias neuronas se activan para una misma entrada a la red neuronal, mediante el algoritmo

(3) se escoge cual debe quedar activa y cual(es) debe(n) inhibir(se):

```

repita para j=1 : numClases
    matchVector = min (Ek, wj)
    SumaPeso = suma(wj)
    ActivaciónClase(j) =  $\frac{\text{suma}(\text{match Vector})}{(\theta + \text{SumaPeso})}$ 
fin
  
```

donde w_j son los pesos de la neurona j, Activación Clase (j) almacena el resultado de la división en un vector para cada una de las j clases activadas, numClases es el número de clases activadas para una entrada dada a la red neuronal ART2. De esta manera, la clase que obtenga el mayor valor de "ActivaciónClase" activa a la neurona de salida a la que pertenece, las demás clases se inhiben.

Si una entrada dada a la red neuronal ART2 pertenece a una clase, el peso de la neurona de entrada será modificado para añadir las características del dato de entrada a las almacenadas anteriormente, para lo cual se utiliza la ecuación (4):

$$w_j(t+1) = (\beta * E_k) + ((1 - \beta) * w_j(t)) \quad (4)$$

donde $w_j(t+1)$ son pesos de la neurona j en el momento $t+1$, $w_j(t)$ son pesos de la neurona j en el momento t

Al finalizar la clasificación cada uno de los picos de la muestra va a estar representado por su nombre, sus características, y la clase a la que pertenece.

3.2.2. Red ART2 para clasificar una actividad determinada de un roedor en un patrón

Clasifica una actividad en comer, dormir, bajo los efectos de una droga, etc. Para realizar esta tarea se va a utilizar una red neuronal ART2 que permitirá clasificar la muestra. Se usa como variables de entrada una cadena de clases (sustancias químicas) que va a representar el conjunto de picos presentes en la muestra a ser clasificada, y como salida va a estar la actividad a la que pertenece (ver figura 10).

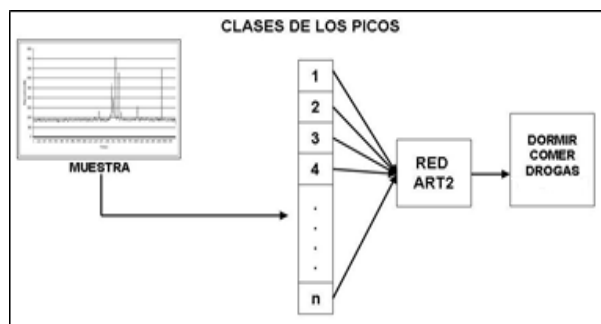


Fig. 10. Diagrama de la red neuronal ART2 para el reconocimiento de la muestra

La clasificación de una muestra en una clase de actividad sigue los siguientes pasos: a) se tiene una muestra extraída del cerebro de un roedor; b) se clasifican los picos de la muestra utilizando la red neuronal ART2 de la sección 3.2.1; c) se construye una cadena de clases de picos, obtenida del paso anterior; d) esta cadena es la entrada a la red neuronal ART2 de esta sección; e) la salida de la red neuronal ART2 es una clase de comportamiento de la muestra.

El número de picos en las muestras no es constante, y esto representa un problema ya que es necesario tener un número de entradas fijas para la red neuronal ART2. Como se hace imposible dinámicamente cambiar el número de neuronas, la red neuronal ART2 está formada por un número n máximo de neuronas de entrada (picos clasificados en la red neuronal ART2 de la sección 3.2.1), y las neuronas de salida representan las clases de actividades del roedor. Para la clasificación de las muestras se van a tener clases desde 1 hasta k , es decir, el sistema podrá reconocer k actividades distintas. La arquitectura de la red neuronal ART2 será la siguiente:

1. Número de Neuronas de Entrada = m (número de clases de la muestra es un valor fijo).
2. Número de Neuronas de Salida = k (variable dependiendo de las clases que se formen, crece dinámicamente).
3. Bias (θ) = 0.01.
4. Número de épocas o iteraciones para clasificar un patrón = 1000.

5. Tasa de Aprendizaje (β) = 0.9.
6. Parámetro de Vigilancia (ρ) = 0.80.

La función de activación de una neurona de salida estará dada por la ecuación (5):

$$S_j = \left\{ \begin{array}{l} \text{repita para } t = 1: \text{nro datos entrada} \\ \text{Si } (e_t \neq 0) \\ \quad elem_t = e_t \\ \text{Fin Si} \\ \text{Fin repita para} \\ \text{suma} = 0 \\ \text{repita para } i = 1: \text{nro datos entrada} \\ \quad \text{repita para } j = 1: \text{nro datos clase} \\ \quad \quad \text{Si}(e_i = w_j) \\ \quad \quad \quad \sum_{t=1}^n I + \text{suma} \\ \quad \quad \text{Fin Si} \\ \quad \text{Fin repita para} \\ \quad \text{Fin repita para} \end{array} \right. \quad (5)$$

0 $Resto$

donde S_j es una neurona de la capa de salida, $elem_t$ son valores de la entrada de la red neuronal, e_t son valores de la entrada de la red neuronal distintos a cero (0), w_j son pesos de la neurona j , suma es un número entero.

La función de activación trabajará de la siguiente manera:

1. Se toma la entrada de la red neuronal.
2. Se seleccionan los datos de entrada que son distintos de cero (0) (ya que la red tiene n neuronas de entrada, es posible que algunas neuronas tengan valor cero (0) y no serán tomadas en cuenta).
3. Se comparan los datos de entrada con los datos de las clases almacenadas en la red neuronal.
4. Si los datos de entrada son iguales a los datos de una clase almacenada se aumenta en uno (1) el valor de suma.
5. El máximo valor de suma hace que se active la neurona de la clase de salida que contiene

los datos almacenados que generan este valor.

6. A S_j se le asigna el valor uno.

Para que una entrada a la red neuronal ART2 pertenezca a una clase se debe cumplir la ecuación (6):

$$\frac{\text{Suma}}{\text{nro datos clase}} < 1 - \rho \quad (6)$$

Como podemos ver, para que los datos de entrada activen una neurona de salida el resultado de la división debe ser menor 0.2, sino se formará una nueva clase con los datos de entrada.

Si varias neuronas se activan para una misma clase, se utilizará el algoritmo que contiene la ecuación (3) presentado antes para escoger la clase ganadora.

Si una entrada dada a la red neuronal ART2 pertenece a una clase, el peso de la neurona de entrada será modificado para añadir las características del dato de entrada a las almacenadas anteriormente para ello se utiliza la ecuación (4).

Al finalizar el proceso, la muestra formará parte de una clase que representa una actividad.

4. Caso de estudio

En nuestro caso, vamos a estudiar dos neurotransmisores excitadores que se encuentran concentradas en el sistema nervioso. Los neurotransmisores a estudiar son:

1. *Glutámato (Acido Glutámico)*: es el principal neurotransmisor excitador del sistema nervioso central de los mamíferos y actúa a través de receptores ionotrópicos (que actúan sobre el paso de iones). La activación de estos receptores es responsable de muchas formas de plasticidad sináptica, que se cree están implicadas en los procesos del aprendizaje y de la memoria. En las lesiones agudas del sistema nervioso central, los receptores para glutámato están implicados en desórdenes neurodegenerativos y neurotóxicos, epilepsia e isquemia cerebral. Además, juega un papel fundamental en la

progresión de los infartos cerebrales, por lo que la medición de los niveles de este neurotransmisor son una prueba que puede servir a los médicos para tratar las enfermedades cerebros vasculares. En específico, es importante su estudio ya que los investigadores la ven como una interesante clase de molécula polivalente, que en definitiva podría llevar a la fabricación de diversas drogas útiles en el tratamiento de muchas enfermedades mentales, como la esquizofrenia, la epilepsia, la ansiedad o las dependencias de drogas. Por lo tanto, el estudio de los receptores de glutámato permitirá un mayor conocimiento del funcionamiento del sistema nervioso, y abre las puertas para el desarrollo de estrategias terapéuticas más eficaces contra enfermedades producidas por la alteración de la neurotransmisión glutámica [9].

2. *Aspartato*: Es un neurotransmisor que se encuentra concentrado en el sistema nervioso central, y ejerce potentes efectos excitadores sobre la actividad neuronal. Si aumenta puede destruir neuronas, incluso antes de que se presenten síntomas clínicos de alguna enfermedad crónica, por lo tanto, parece relacionarse también con los mecanismos de muerte celular, dado que al activarse elevan las concentraciones intracelulares de calcio. El Aspartato, junto con el Glutámato, se encuentran involucrados en procesos tan diversos como la epilepsia, las lesiones cerebrales isquémicas y el aprendizaje, influyendo en el desarrollo de las conexiones sinápticas normales del cerebro. Su actuación es tan patente que podría ser que las grandes degeneraciones neurológicas, como el Alzheimer, pudieran deberse a una hiperactividad de los mismos [9, 10].

El espectro del Glutámato y Aspartato en la electroforesis capilar se puede observar en el siguiente electroferograma (ver figura 11), los cuales son reconocidos por los expertos de forma visual contando solo con la experiencia.

Nuestro sistema de extracción de conocimiento cerebral analiza las muestras extraídas del cerebro de roedores, y en particular, determina las sustancias presentes, para

posteriormente detectar el patrón de las muestras.

4.1. Pruebas

Para la construcción del Sistema de Minería de Datos se utilizó Matlab [26] bajo Linux. Para realizar la parte experimental se tomaron 6 muestras a un roedor mediante electroforesis capilar. Con ellas se va a estudiar los neurotransmisores Glutámato y Aspartato, analizando las muestras extraídas utilizando el Sistema de Minería de Datos propuesto. Estas muestras son experimentos realizados en el Departamento de Fisiología en la Universidad de Los Andes.

Las muestras fueron almacenadas de manera numerada en archivos: Grupo_A_01.elf, Grupo_A_2.elf, Grupo_A_3.elf, Grupo_A_4.elf, Grupo_A_5.elf, Grupo_A_6.elf. La extensión .elf define a un archivo del tipo electroforesis capilar. Para cada muestra, mediante la electroforesis capilar se generan entre 3500 a 5000 datos que representa el electroferograma.

A continuación, se observa el panel principal del Sistema (ver figura 12). Para analizar la muestra es necesario filtrar los datos.

El sistema de Minería de Datos clasificara los picos a la clase correspondiente de la sustancia química que representa. Lo mostrara al usuario a través de una ventana, donde para cada pico tendremos: número del pico, altura, área, punto de inicio, punto de fin, ancho y clase a la que pertenece. Todo esto compone la representación intermedia.

El Sistema clasifica los picos por números, desde 1 hasta n, ya que es una vía fácil para representar la clasificación. Cada número representa una sustancia química diferente. Los patrones son expresiones que describen la secuencia de sustancias químicas que aparecen para una actividad dada del roedor. Una sustancia química cuando está presente en la gráfica aparece siempre en la misma posición o en una posición cercana dentro de la muestra, por lo tanto, la red neuronal ART2 debe ser capaz de reconocer una sustancia química por la posición que ocupa el pico en la gráfica, por lo tanto, distintas posiciones de los picos en la

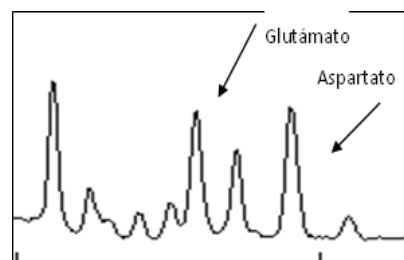


Fig. 11. Espectro del Glutámato y Aspartato, visto por los expertos [9]

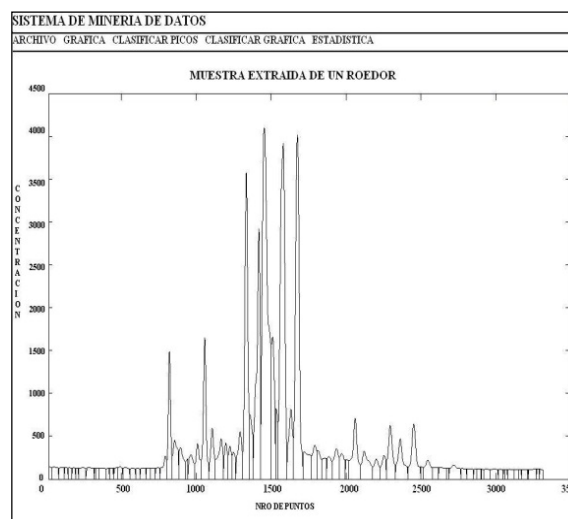


Fig. 12. Panel Principal del Sistema de Minería de Datos y filtrado de una muestra

gráfica representará distintas sustancias químicas.

Por otro lado, el sistema permite asociar las muestras a una actividad. Al igual que los picos, utilizaremos la nomenclatura 1 hasta m para clasificar a dichas actividades, donde este número m representa el número de actividades consideradas/estudiadas. Esto le da una gran versatilidad a nuestro enfoque, ya que el número de actividades se define según el interés del estudio.

Para realizar esta tarea se va a utilizar la red neuronal ART2 que permite clasificar la muestra. Se utiliza como variables de entrada a la red neuronal una cadena de clases (sustancias

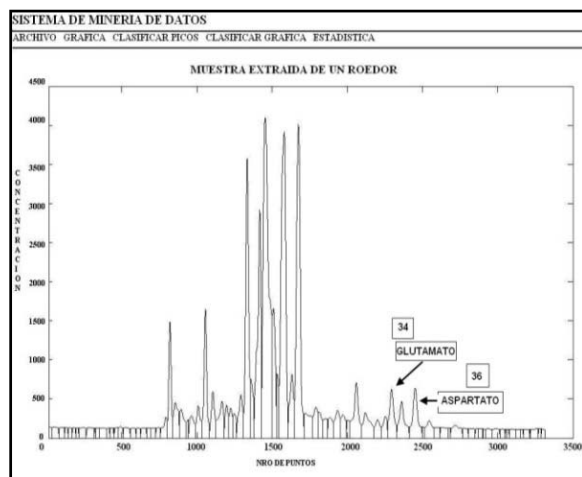


Fig. 13. Localización de los neurotransmisores Glutámato y Aspartato en una muestra

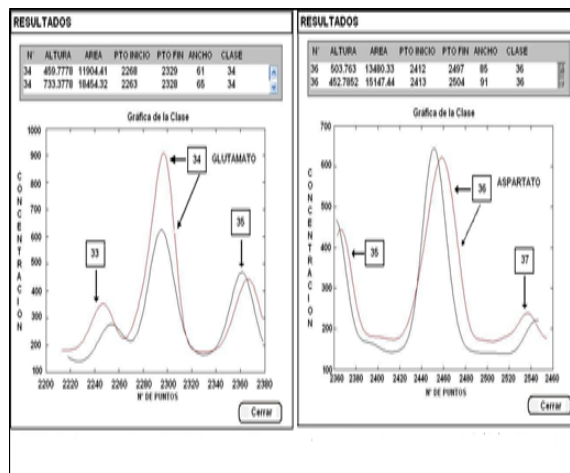


Fig. 15. Comparación del Glutámato y Aspartato entre dos muestras

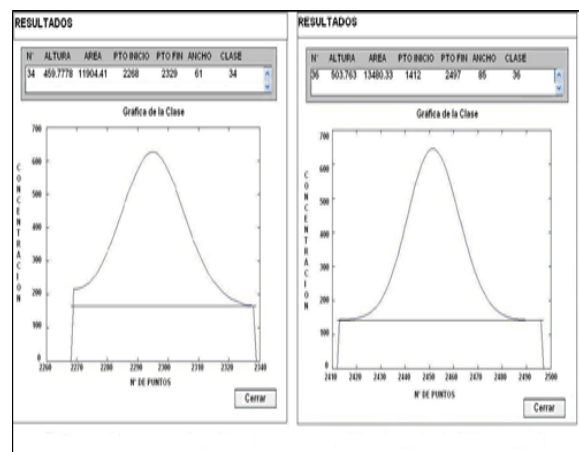


Fig. 14. Resultado de la búsqueda de Glutámato y Aspartato en una muestra

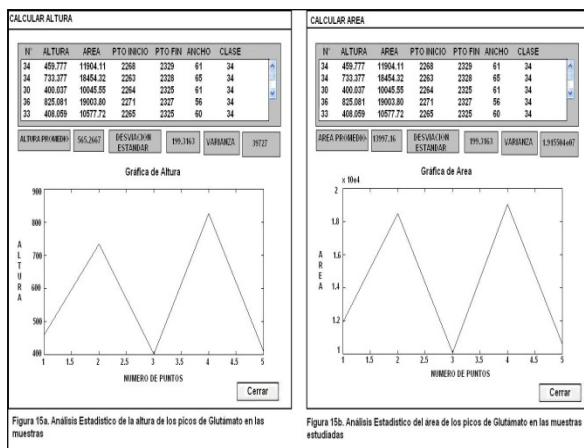


Figura 15a. Análisis Estadístico de la altura de los picos de Glutámato en las muestras

Figura 15b. Análisis Estadístico del área de los picos de Glutámato en las muestras estudiadas

Fig. 16. Análisis estadístico de la altura y el área de los picos de Glutámato de las muestras

químicas) que van a representar el conjunto de picos presentes en la muestra a ser clasificada, y como salida de la red va a estar la actividad a la que pertenece.

Veamos el caso de los neurotransmisores Glutámato y Aspartato, aparecen según los expertos en los picos 34 y 36 [14]. En la clasificación hecha por el sistema en una muestra

(ver figura 13), usando la red ART clasificadora de picos, para los picos de Glutámato y Aspartato en una muestra, el sistema calcula las características de altura, área, punto de inicio, punto de fin, ancho y clase asociados a cada uno (ver figura 14).

A partir de esa clasificación hecha, el sistema puede comparar picos de diferentes muestras

Tabla 1. Características de los picos que representan el Glutámato

Pico Nº	Altura	Area	Pto Inicio	Pto Fin	Ancho	Clase
34	460	11905	2268	2329	61	34
34	733	18454	2263	2328	65	34
30	400	10045	2264	2325	61	34
36	825	19004	2271	2327	56	34
33	408	10578	2265	2325	60	34

(ver figura 15). Esto permite comprender la aparición de cambios en las células del sistema nervioso central (SNC). Los cambios en el SNC pueden mostrar, por ejemplo, tendencias degenerativas en los neurotransmisores entre muestras, lo que puede ayudar en el diagnóstico temprano de enfermedades [27]. Además, para el caso específico de la figura 15, se puede estudiar la función de las moléculas transportadoras de Glutámato en los fenómenos de plasticidad, tanto en la regulación de la función neuronal normal como en eventos patológicos.

Como se dijo antes, el sistema también realiza un análisis estadístico para observar las tendencias de los picos de las sustancias químicas en diferentes muestras, usando como variables la altura y el área. En nuestro caso estudiamos al neurotransmisor Glutámato (figura 16) en 5 muestras. Como para esta ilustración es difícil de discernir los valores de esta figura, se presentan en la tabla 1.

Así, podemos ver que existen variaciones en los valores del Glutámato en las muestras, esto es estudiado por los expertos para llegar a conclusiones del funcionamiento del cerebro del roedor (por ejemplo, determinar una tendencia anormal).

En particular, para nuestro caso de estudio determinar el incremento en las concentraciones de Glutámato es un caso de daño neuronal agudo o crónico excitotóxico. Hay indicios que vinculan esos cambios con diversos daños agudos y crónicos, entre éstos el trauma craneoencefálico, la isquemia cerebral, los síndromes convulsivos, la enfermedad de Alzheimer, la esclerosis lateral amiotrófica y la neuropatía por VIH [27].

4.2. Comparación con otros trabajos

En la tabla II comparamos nuestra propuesta con trabajos similares. En [16] se especifica un método para el análisis de datos provenientes de la Electroforesis Capilar basado en un algoritmo de pre-procesamiento para eliminar el ruido, que incluye la transformada wavelet discreta (reducir la dimensión de los datos sin perder la información relevante), detección de regiones interesantes y la corrección de la línea base. La señal resultante es mapeada en una secuencia, explotando la primera derivada y la cuantificación multi-nivel en la altura de los picos. A continuación, se aplican algoritmos de alineación locales para el reconocimiento de los picos. Finalmente, se lleva a cabo una aproximación gaussiana para asegurar mediciones precisas de la altura de los picos.

En [29] se presenta un algoritmo para la detección de picos en un espectro, se construyó un framework donde se introduce la señal, y se tienen varias etapas de suavizado, corrección de la línea base y recolección de los picos. En nuestro trabajo se presenta un sistema para la extracción de conocimiento a partir de los grandes volúmenes de datos que se obtienen de experimentos, para determinar las sustancias químicas que actúan en el cerebro, y entender las interacciones que suceden en él cuando un roedor realiza una actividad. Dicho sistema se basa en el uso de la red neuronal artificial basada en la Teoría de Resonancia Adaptativa.

En todos los trabajos anteriores se hace un suavizado de la señal por métodos matemáticos, para hacer la clasificación de los picos. La clasificación es hecha en nuestro sistema usando técnicas inteligentes, a diferencia de los otros dos

Tabla 2. Comparación con otros trabajos

Característica	[16]	[29]	Nuestro Enfoque
Suavizado de la Señal	Procesamiento Wavelet	Filtro de Savitzky-Golay Filtro Gaussiano Ventana de Keiser Filtros basados en Wavelet	Filtro de Savitzky-Golay
Corrección de la Línea Base	Interpolación cúbica utilizando coeficientes preseleccionados para Wavelet	Mínimo monótono Interpolación Lineal Loess Transformada Continua de Wavelet Media Móvil de Mínimos	Línea Recta en el valor mínimo del pico
Herramienta para la Clasificación de los picos	Alineamiento Local	SRN Detección/Intensidad Umbral Pendiente de los picos Máximo Local Relación de forma Líneas de Crestas Criterio basado en el modelo Ancho de los picos	Red Neuronal ART2
Herramienta para la Clasificación del electroferograma en una actividad	-	-	Red Neuronal ART2
Datos sobre los picos (Área, Altura)	-	-	Medidas Estadísticas
Comparación de Picos de diferentes muestras	-	-	Si

métodos que usan técnicas de procesamiento de señales.

Finalmente, nuestro trabajo es el único que tiene una segunda fase (ver cuarta fila de la tabla II), en la cual, una vez identificado los picos, se clásica cada muestra en una actividad, según las concentraciones de las sustancias químicas presentes en ella. Los resultados de las clasificaciones de los picos en cada trabajo son exactos (siempre logran identificar cada pico correctamente).

Por otro lado, nuestra herramienta provee otras opciones para ayudar en el proceso de análisis de los picos, como las medidas estadísticas y el mecanismo de comparación

antes comentados (ver últimas dos filas de la tabla II).

5. Conclusiones

La extracción de patrones y conocimiento desde fuentes de información es posible porque existen herramientas computacionales que se han podido conjugar para: a) la recolección masiva de datos a través de grandes bases de datos; b) el procesamiento de una gran cantidad de datos en poco tiempo; c) el descubrimiento de conocimiento.

En este trabajo se desarrollan e implementan algoritmos novedosos, efectivos y eficientes en el

tiempo, para realizar el análisis automático de grandes cantidades de datos provenientes de la electroforesis capilar. La solución tiene la capacidad de eliminar componentes ruidosas de la señal adquirida y permitir la compresión de datos, incrementar la capacidad de identificar sustancias solapadas, reconocer patrones con variaciones en tiempo y amplitud, realizar mediciones precisas, extraer información no explícita e incorporar herramientas de visualización de fácil comprensión de la información.

Nuestro sistema permite representar de forma más específica y natural el conocimiento que emerge, por medio de patrones que representan sustancias químicas extraídas del cerebro de roedores. En específico, el sistema descubre las sustancias químicas presentes en las muestras extraídas del cerebro del roedor, y clasifica las muestras a una actividad de manera comprensible para los usuarios.

En el caso de estudio, se analizaron los neurotransmisores Glutamato y Aspartato utilizando nuestro Sistema, permitiendo un mayor conocimiento del funcionamiento del sistema nervioso. Además, abre las puertas para el desarrollo de estrategias terapéuticas más eficientes contra enfermedades producidas por la alteración de la neurotransmisión glutámica (esquizofrenia, epilepsia, la ansiedad o las dependencias de drogas). Todo esto gracias al análisis en los electroferogramas, ya que se obtuvo el reconocimiento y clasificación de todos los picos presentes en las muestras y la clasificación de éstas en una actividad. De esta manera, los patrones encontrados por el sistema proporcionan una manera más rápida de analizar las muestras de los experimentos.

Los resultados de las pruebas realizadas pueden ser considerados muy prometedores, ya que el Sistema descubre patrones de las sustancias químicas presentes en el cerebro. Para lograr esto fue necesario analizar al conjunto enorme de datos, descubrir promedios, desviaciones, tendencias, dependencias, asociaciones entre las sustancias químicas presentes en los fluidos biológicos de los roedores, entre otras cosas, y ofrecer la

capacidad de comparar y relacionar la información descubierta.

Algo relevante final a señalar es que la Minería de Datos es una herramienta explorativa y no explicativa, es decir, explora los datos para sugerir hipótesis útiles para la toma de decisiones. Eso es básicamente lo que permite hacer nuestra herramienta, para ayudar a un experto en el análisis del comportamiento del cerebro.

Agradecimiento

Al Proyecto CDCHTA I-1407-14-02-B de la Universidad de Los Andes por su apoyo financiero.

Referencias

1. **Pérez, L., Arencibia, R., Conill, C., Achón, G., & Araujo, J. (2003).** Impacto de la Bioinformática en las ciencias biomédicas. *ACIMED*, Vol. 11, No. 4, Cuba.
2. **Altamiranda, J., Aguilar, J., & Delamarche, C. (2013).** Comparison and fusion model in protein motifs. *Proceedings XXXIX Latin America Computing Conference (CLEI 2013)*, pp. 1–12.
3. **Altamiranda, J., Aguilar, J., & Delamarche, C. (2011).** Similarity of Amyloid Protein Motif using an Hybrid Intelligent System. *Latin America Transactions IEEE*, Vol. 9, No. 5, pp. 700–710. doi:10.1109/TLA.2011.6030978
4. **Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J., & Zhang, J. (2004).** Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, Vol. 5, No. 10. doi:10.1186/gb-2004-5-10-r80
5. **Cock, P., Antao, T., Chang, J., Chapman, B., Cox, C., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., & Hoon, M. (2009).** Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, Vol. 25, No. 11, pp 1422–1423. doi: 10.1093/bioinformatics/btp163

6. **Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B., & Rätsch, G. (2008).** Support Vector Machines and Kernels for Computational Biology. *PLoS Comput Biol*, Vol. 4, No. 10. doi:10.1371/journal.pcbi.1000173
7. **Sánchez, F., López, G., & García, V. (1999).** Bioinformática y salud: Impactos de la aplicación de las nuevas tecnologías para el tratamiento de la información genética en la investigación biomédica y la práctica clínica. *Informática y Salud*, No. 19.
8. **Febles, J., & González, A. (2002).** Aplicación de la minería de datos en la Bioinformática. *ACIMED*, Vol.10, No. 2, Cuba.
9. **Hernandez, L., Joshi, N., Murzi, E., Verdeguer, P., Mifsud, J., & Guzman, N. (1993).** Collinear laser-induced fluorescence detector for capillary electrophoresis. Analysis of Glutamic acid in brain dialysates. *Journal of Chromatography A*, Vol. 652, pp. 399–405. doi:10.1016/0021-9673(93)83259-U
10. **Guzman, N., Park, S., Schaufelberger, D., Hernandez, L., Paez, X., Rada, P., Tomlinson, A., & Naylor, S. (1997).** New approaches in clinical chemistry: on-line analyte concentration and microreaction capillary electrophoresis for the determination of drugs, metabolic intermediates, and biopolymers in biological fluids. *J. Chromatogr. B.*, Vol. 697, pp. 7–66. doi:10.1016/S0378-4347(97)00275-2
11. **León, V. (1998).** *Manual de usuario del ONICE*. Universidad de los Andes, Facultad de Medicina, Departamento de Fisiología, Mérida, Venezuela.
12. **Aguilar, J., & Hernández, L. (2009).** Design and Implementation of a Patterns Recognition System for Analysis of Biological Liquids. *IEEE Trans. Latin America*, Vol. 7, No. 1, pp. 12–26. doi:10.1109/CEC.2008.4631056
13. **Cazes, J. (2004).** *Encyclopedia of Chromatography*. Marcel Dekker.
14. Electroforesis [online]. <http://www.javeriana.edu.co/Facultades/Ciencias/neurobioquimica/libros/celular/electroforesis.html>
15. Practica 4. Espectrofotometría [online]. <http://www.upo.es/depa/webdex/quimfis/docencia/quimbiotec/FQpractica4.pdf>
16. **Ceballos, G., Hernández, L., & Paredes, J. (2009).** Data Processing and Pattern Recognition in High – Throughput Capillary Electrophoresis. *17th European Signal Processing Conference (EUSIPCO 2009)*, pp.1592–1596.
17. **Ceballos, G., Hernández, L., & Paredes, J. (2007).** A novel approach for Pattern Recognition in Capillary Electrophoresis Data. *CLAIB 2007, IFMBE Proceedings*, Springer, Vol. 18, pp.150–153.
18. **Ceballos, G., Cuadros, J., & Paredes, J. (2008).** Herramienta de software para el análisis de datos. *Mecánica Computacional*, Vol. 27. pp.3299–3315.
19. **Paredes, J. & Sosa, E. (2009).** Corrección de línea base en datos electroforéticos usando optimización local del algoritmo legend en el dominio wavelet. *Revista Interciencia*, Vol. 34, No 8.
20. **Savitzky, A. & Golay, M. (1964).** Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, Vol. 36, No. 8, pp. 1627–1639. doi: 10.1021/ac60214a047
21. **Aguilar, J. & Rivas, F. (2001).** *Introducción a la Computación Inteligente*. MERITEC. Venezuela.
22. **Hilera, J. & Martinez, V. (1995).** *Redes Neuronales Artificiales: Fundamentos, Modelos y Aplicaciones*. Addison-Wesley.
23. **Carpenter, G. & Grossberg, S. (1987).** ART 2: self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, Vol. 26. No. 23, pp. 4919–4930. doi: 10.1364/AO.26.004919.
24. **Carpenter, G. & Grossberg, S. (1988).** The ART of Adaptive Resonance Theory by a Self-Organizing Neural Network. *IEEE Computer*, Vol. 21, No. 3, pp. 77–88. doi:10.1109/2.33
25. **Carpenter, G. & Grossberg, S. (1991).** *Pattern Recognition by Self-Organizing Neural Network*. MIT Press.
26. **Gobbert, M. ([online])** *How to Get Started with Matlab at UMBC*. <http://www.math.umbc.edu/~gobbert/matlab.html>
27. **Medina, A. & Escobar, M. (2004).** Plasticidad Neural y su relación con el sistema de transportadores de Glutámico. *Revista Colombiana de Psiquiatría*, Vol. 33, No. 1, pp. 155–164.
28. **Stewart, R., Gideoni, I., & Zhu, Y. (2011).** Signal Processing Methods for Capillary Electrophoresis Systems and Computational Biology. *Bioinformatics and Computational Modeling*, Ning-Sun Yang (Ed.).
29. **Yang, C, He, Z, & Yu, W. (2009).** Comparison of public peak detection algorithms for MALDI mass

spectrometry data analysis. *BMC Bioinformatics*, Vol. 10, No. 4.

30. **Díaz, J. (2008).** Análisis de Comparativa de Métodos de Adquisición de Datos en los Sistemas Automáticos del Currículo de Tecnología Industrial. *Revista Digital CSI*, No. 5.
31. **Feilmayr, C. (2013).** Optimizing Selection of Assessment Solutions for Completing Information Extraction Results. *Computación y Sistemas*, Vol. 17, No. 2, pp. 169–178.
32. **Morales, L., Vázquez, R., Morales, E., & Semenovich, Y. (2012).** Speckle Noise Reduction in Ultrasound Imaging using the Key Points in Low Degree Unbiased FIR Filters. *Computación y Sistemas*, Vol. 16, No. 3, pp. 287–295.
33. **Batyrshin, I. & Sheremetov, L. (2008).** Perception-based approach to time series data mining. *Applied Soft Computing*, Vol. 8, No. 3, pp. 1211–1221.

Junior Altamiranda se graduó de Ingeniero de Sistemas en 2002 en la Universidad de Los Andes Mérida-Venezuela. En 2006 obtuvo el título de MSc. en Computación en la Universidad de Los Andes Mérida-Venezuela. En el 2007 realizó una Estadía de Investigación en el Master 2 de Bionformática en la Universidad de Rennes I en Francia. Formo parte del Plan II: Programa de Generación de Relevo de la Universidad de Los Andes del 2008 al 2012. En 2012 obtuvo el título de PhD en Ciencias Aplicadas en las Universidad de Los Andes Mérida-Venezuela. Es profesor Asistente en el Departamento de Computación de la Universidad de Los Andes en Venezuela. Ha publicado artículos y capítulos de libros en Minería de Datos, Computación Inteligente, Sistemas Multiagentes, Inteligencia Artificial, Bioinformática.

José Aguilar se graduó de Ingeniero de Sistemas en 1987 en la Universidad de Los Andes Mérida-Venezuela. En 1991 obtuvo el título de MSc. en Ciencias Computacionales en la

Universidad Paul Salbatier en Francia. En 1995 obtuvo el título de PhD en Ciencias Computacionales en las Universidad Rene Descartes en Francia. Fue Investigador Post-Doctoral en el Departamento de Computer Sciences en la Universidad de Houston (1999-2000). Es profesor Titular en el Departamento de Computación de la Universidad de Los Andes en Venezuela. Ha publicado más de 200 artículos y 5 libros, en Sistemas Paralelos y Distribuidos, Computación Inteligente, Computación Emergente, Inteligencia Artificial y Sistemas Multiagentes, etc. Ha sido profesor/investigador visitante en diferentes universidades y laboratorios, además ha sido coordinador de más de 20 proyectos industriales y tutor de tesis de más de 20 estudiantes entre Doctorado y Maestría.

Luis Hernandez se graduó de Médico en 1966 en la Universidad de Los Andes Mérida-Venezuela, Coordinador General del Consejo de Desarrollo Científico, Humanístico, Tecnológico y de las Artes (CDCHTA) de la Universidad de Los Andes (1980-1984). Científico Visitante del Programa de Neurociencia en la Universidad de Princeton USA (1986-1988). Es Profesor Titular en el Departamento de Fisiología en la Universidad de Los Andes. Fue Co-fundador de la Compañía Europhor S.A. en Toulouse Francia. Ha obtenido los siguientes premios y distinciones: Premio Nacional de Innovación Tecnológica 1993, Orden Andrés Bello 2da Clase 1993, Premio Francisco Vananzi 1994 y Miembro de la Academia de Mérida. Ha publicado más de 126 artículos, 33 capítulos de libros, realizando contribuciones significativas contribuciones para el estudio de los procesos bioquímicos en el cerebro.

Artículo recibido el 24/09/2012, aceptado el 09/12/2014. Autor de correspondencia es Junior Altamiranda.