

Desarrollo de un prototipo para detección de sustancias volátiles haciendo uso de distancia euclidiana

ANGELA HARUMI FORT ISHIWARA, JOSÉ MIGUEL VADILLO PRO, ALEJANDRO GONZÁLEZ MARTÍNEZ, JOSÉ AMBROSIO BASTIÁN, VÍCTOR RAMOS VITERBO

Resumen— En la actualidad se han desarrollado distintos sensores capaces de detectar olores, lo que ha generado una gran cantidad de aplicaciones potenciales para un sistema de reconocimiento, como aplicaciones médicas, alimenticias, detección de sustancias peligrosas, farmacéuticas, etc. Debido a esto, se ha planteado realizar un dispositivo inteligente capaz de emular el sentido del olfato para identificar sustancias volátiles, con base en un entrenamiento previo. Dichas sustancias pueden encontrarse dentro de un compuesto y en diferentes concentraciones. Como primera aproximación al problema, se desarrolló un prototipo que es capaz de identificar sustancias volátiles, generar bases de datos y comunicarse con una computadora, además de un algoritmo basado en el método de reconocimiento de patrones de distancia euclidiana. Este método se utilizó debido a su simplicidad algorítmica, lo que lo hace perfecto para integrarlo en un microcontrolador por el poco almacenamiento que poseen y como primera iteración al problema con los requerimientos mínimos necesarios.

I. INTRODUCCIÓN

Gracias al avance que ha tenido la electrónica, es posible implementar encapsulados útiles para distintas aplicaciones. Entre ellos se encuentran los sensores, los cuales ayudan a detectar magnitudes químicas o físicas, pudiendo describir el entorno que vivimos.

El uso de sensores se ha ido extendiendo hacia distintas ramas, con la llegada de los sensores químicos, en la segunda mitad de la década de los noventa, se desarrollaron los sistemas de detección de olores, teniendo aplicaciones como son: clasificación de aceite comestible, control de alimentos, detección de narcóticos, detección de explosivos, identificación por sistemas biométricos, aplicaciones biomédicas, diagnósticos médicos, como cetoacidosis en diabéticos [1], identificación de licor, industrias farmacéuticas, aplicaciones de seguridad, y otras [3-11].

En el caso de las aplicaciones médicas, en los últimos años se han desarrollado varios experimentos que muestran la relación del aliento de los pacientes con ciertas enfermedades,

ANGELA HARUMI FORT ISHIWARA, JOSÉ MIGUEL VADILLO PRO y ALEJANDRO GONZÁLEZ MARTÍNEZ, son egresados de la carrera INGENIERÍA MECATRÓNICA de la Facultad de Ingeniería y realizaron el proyecto en el GIDi de Sistemas Embebidos.

El proyecto fue asesorado por el Dr. JOSÉ AMBROSIO BASTIÁN y el Ing. VÍCTOR RAMOS VITERBO.

Los autores agradecen a: German Stadelmann por el desarrollo de la cámara de gases y la matriz de sensores, así como la facilitación de su documentación. También agradecemos a Víctor Ramos Viterbo, por realizar la estructura principal del software del microcontrolador y a José Ambrosio Bastián por la asesoría en los algoritmos de reconocimiento.

lo que podría generar a futuro, análisis clínicos, más rápidos, económicos y no invasivos.

En los diagnósticos médicos, es de gran utilidad detectar concentraciones de compuestos en las sustancias, y es necesario ser específicos en las sustancias a identificar.

Para identificar correctamente una sustancia, es necesario obtener información relevante de ésta por medio de sensores y con ello, realizar modelos matemáticos que describan a dicha sustancia, generando su perfil. Para ello, existen diversos métodos de reconocimiento de patrones.

A. Reconocimiento de patrones

El reconocimiento de patrones [2] se encarga de la descripción y clasificación (reconocimiento) de objetos, personas, representaciones, etc. Esta ciencia trabaja con base en un conjunto previamente establecido de todos los posibles objetos (patrones) individuales a reconocer. En un sistema de reconocimiento de patrones, el sensor tiene como propósito proporcionar una representación factible de los elementos del universo a ser clasificados, además que existen varias etapas que ajustan el sistema para que sea capaz de clasificar señales u objetos de entrada en una de las *clases* predefinidas. Para ello deberá analizar un cierto número de características y para poder clasificar satisfactoriamente señales de entrada, es necesario un proceso de aprendizaje en el cual el sistema crea un modelo de cada una de las clases a partir de una secuencia de *entrenamiento* o conjunto de vectores de características de cada una de las clases. La secuencia de muestras (elementos) de entrenamiento debe contener para cada una de las clases un mínimo de *elementos* para su análisis adecuado. El *reconocimiento* es la etapa de toma de decisiones en el sistema. Su rol es asignar los patrones de clase desconocida a la categoría apropiada.

Si para cada clase se grafica cada una de sus características (valor de los sensores), la dimensión n sería el número de sensores ocupados.

Es pertinente obtener los centros para cada clase dependiendo de los resultados de sus características (sensores). Esto se realiza obteniendo el promedio de valores para cada sensor extraído en cada uno de los *elementos*. Posteriormente se aplica la distancia euclidiana.

B. Distancia euclidiana

A partir de conocido teorema de Pitágoras, se estableció una técnica para determinar la distancia entre dos puntos [3]. Sean los puntos $p_1(x_1, y_1)$ y $p_2(x_2, y_2)$ y d la distancia entre p_1 y p_2 , entonces:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

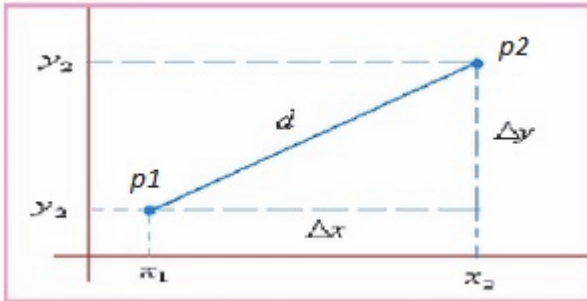


Figura 1. Distancia entre dos puntos

Gracias a este método se puede clasificar una clase, midiendo desde el punto de las coordenadas de la clase desconocida a la clase guardada en la fase de entrenamiento. Siendo cada punto el valor de cada sensor.

Este proceso se realiza con cada clase que se encuentre guardada en la fase de entrenamiento, la clase en la que se tenga la distancia menor, será donde pertenezca la clase desconocida.

La fórmula para n dimensiones de la distancia euclidiana entre dos puntos P y Q en \mathbb{R}^n es:

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (2)$$

II. HIPÓTESIS

Si el prototipo se entrena mediante una base de datos conformada por muestras de olores (con rasgos característicos), de determinadas sustancias o clases, entonces podrá identificar posterior al entrenamiento del sistema de reconocimiento, elementos que pertenezcan a estas clases, mediante la detección de forma indirecta, a partir de sensores de gas no especializados en cada clase original.

III. ETAPAS Y DIVISIÓN DEL PROTOTIPO

El prototipo consta de los siguientes módulos:

1. Tarjeta de Principal
2. Interfaz USB
3. Matriz de Sensores
4. Cámara de Gases

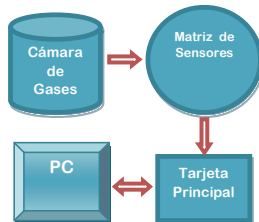


Figura 2. Diagrama de bloques del prototipo

A. Tarjeta principal e interfaz USB

Es una tarjeta que permite el control de los periféricos por medio del microcontrolador PIC18F4620. Los periféricos son:

1. Matriz de Sensores – Identifica concentraciones de 11 diferentes tipos de gases.
2. Interfaz USB – Comunica al prototipo con la PC utilizando un puerto USB.

3. Teclado – Ingresar datos al software por medio de la elección del usuario.
4. Pantalla LCD – Interfaz visual para el usuario.
5. Memoria EEPROM – Proporciona más espacio de almacenamiento.

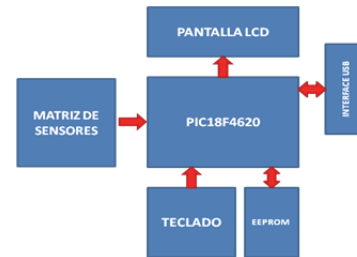


Figura 3. Esquemático de la tarjeta principal



Figura 4. Resultado de la manufactura de la tarjeta principal

B. Matriz de sensores

La tarjeta consta de un conjunto de sensores que detectan diversos gases, los cuales se muestran a continuación:

1. TGS 821 → hidrógeno
2. TGS 822 → solventes orgánicos
3. TGS 823 → solventes orgánicos
4. TGS 2444 → amoníaco
5. TGS 2600 → contaminantes
6. TGS 2611 → metano
7. TGS 2612 → metano y gas LP
8. TGS 2620 → vapores de solventes
9. TGS 3870-A04 → monóxido de carbono
10. TGS 5042 → monóxido de carbono
11. TGS 2610 → propano
12. LM35 → temperatura
13. hih4000 → sensor de humedad



Figura 5. Resultado de la manufactura de la matriz de sensores

C. Cámara de gases

Este dispositivo sirve para cerrar herméticamente las sustancias, permitiendo no perder datos importantes de ésta, ya que los compuestos que se utilizan son volátiles.

En la tapa superior, se fija la matriz de sensores, permitiendo capturar las muestras adecuadamente. Internamente, en la parte inferior se coloca la muestra.

La Figura 7 muestra cómo luce el ensamblaje del prototipo.

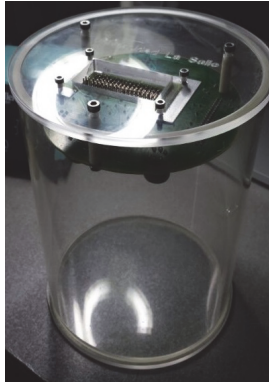


Figura 6. Resultado de la manufactura de la cámara de gases

IV. BLOQUES QUE COMPONEN EL SOFTWARE DEL PROTOTIPO

El software del prototipo se realizó en lenguaje C. La estructura de este programa proporciona tanto la interfaz de usuario como el manejo de periféricos del PIC, se divide en:

1. Bienvenida
2. Elección del procesamiento de datos
3. Muestreo
4. Entrenamiento
5. Reconocimiento
6. Interfaz con la computadora
 - a. Muestreo por lote
 - b. Muestreo en línea



Figura 7. Ensamble de prototipo

A. Bienvenida

Al iniciar el programa, se muestra en la pantalla una breve bienvenida, así como la versión del programa.

B. Elección del procesamiento de datos

Se debe seleccionar si se quiere trabajar con la tarjeta principal o de forma paralela con la computadora.

C. Muestreo

Obtiene el valor de los sensores, guardándolos en una variable que posteriormente se utiliza en los algoritmos de entrenamiento y reconocimiento.

Es importante destacar que se necesita estabilidad térmica en los sensores, por lo que el algoritmo espera por cada clase 40 segundos antes de iniciar con el primer muestreo de los 11 sensores.

D. Entrenamiento

Es una función que tiene como objetivo obtener y guardar en la memoria interna de la tarjeta los centros para cada clase, mediante una base de datos obtenida por la matriz de sensores.

E. Reconocimiento

Es una función encargada de identificar las sustancias. Se obtiene una nueva lectura de la matriz de sensores, con la cual, mediante el algoritmo de reconocimiento de patrones de distancia euclidiana se obtiene la clase a la cual pertenece.

F. Interfaz con la computadora

El prototipo cuenta con una interfaz que permite la comunicación con la computadora. Esta interfaz es a partir de un UB232R, que es una herramienta que se usa para comunicación de USB a protocolo serial, por medio de un miniUSB.

En cuanto a lenguajes de programación en la PC se escogió Java por facilidad de uso, rapidez de desarrollo y mejor manejo de interfaz gráfica. Se desarrolló un programa con dos tipos de procesamiento de datos: “recibir por lote” y “recibir en línea”.

El módulo “recibir por lote” se implementa cuando no se tiene acceso a la computadora, con el fin de hacer las mediciones de campo y después se extrae la información formando una base de datos en la PC. Para poder acceder al este modo, es necesario elegir en el menú del prototipo la opción “muestreo por lote”.

El módulo “recibir en línea” se desarrolla cuando se tiene acceso a la computadora con el fin de hacer la base de datos en el momento automáticamente. En el menú del prototipo se elige la opción muestreo por línea.

Para tener una adecuada intercomunicación entre el dispositivo y la computadora, se estableció un formato de envío de información, en el cual se asignan caracteres predefinidos para distinguir qué tipo de captura se está realizando y la confirmación de término de cada envío.

La base de datos se extrae de manera automática. El formato también se realiza de manera automática. Con el fin de complementar la base de datos, se agregaron distintos campos de información, como son: nombre de clase, nombre de proyecto, nombre del equipo, etc.

La base de datos proporciona un procesamiento y manejo de datos conveniente para el usuario, capaz de tabular los datos en Excel, por lo que facilita su manipulación y su forma de almacenamiento, ya que una computadora tiene más memoria que la proporcionada para la tarjeta principal.

V. EXPERIMENTO PARA COMPROBAR LA FUNCIONALIDAD DEL PROTOTIPO

Se propone una prueba, la cual consiste en tomar 50 elementos, por cada sensor, en 4 clases, las cuales se muestran a continuación:

1. Nada (cámara de gases vacía, únicamente contiene aire)
2. Mezcal
3. Cerveza
4. Alcohol etílico

Se ha optado en experimentar con sustancias alcohólicas debido a su similitud química y su grado alcohólico proporcionada en la botella. Como éstas contienen distinta concentración de etanol, son útiles para poder acercarse o alejar la distribución de las muestras y observar el comportamiento del algoritmo. Si las sustancias fueran demasiado diferentes entre sí, o no contáramos con información de su contenido, no nos proporcionarían información valiosa para probar el dispositivo.

A. Requerimientos

Cada sustancia se coloca en un recipiente pequeño. El volumen del líquido contenido en todos los recipientes, debe ser el mismo.

Un recipiente se coloca dentro de la cámara de gases, que será la clase a entrenar en ese momento. También, hay que tapar la cámara para comenzar a realizar el muestreo.

B. Entrenamiento

Se capturan los 50 elementos de la clase elegida, de los cuales 45 fueron tomados para el entrenamiento.

Es necesario ventilar la cámara de gases para evitar combinación de gases de diferentes sustancias.

Se repite el procedimiento para cada una de las clases (nada, mezcal, cerveza, alcohol etílico).

Dentro del entrenamiento, se obtiene el promedio de los valores de cada sensor para cada clase, es decir, los centros de las clases.

TABLA I
CENTROS DE CADA CLASE

Número de sensor	Nada	Cerveza	Mezcal	Alcohol
1	39.62222	37.48889	37.08889	36.62222
2	50.88889	91.06667	124.51111	146.77778
3	89.04444	157.46667	187.15556	198.93333
4	54.84444	52	52.4	63.2
5	63.91111	91.48889	121.6889	137.2
6	43.35822	43.48222	42.61044	43.64911
7	0	0	0	0
8	55.17778	145.0222	189.3111	207.2667
9	0	0	0	0
10	111.5556	172.8222	198.8889	208.6
11	89.11111	151.7556	179.5778	188.0889
12	0	0	0	0
13	66.53333	66.46667	65.91111	65.97778
14	0	0	0	0

Los números de sensor corresponden a la numeración previamente descrita en la descripción de "Matriz de Sensores".

C. Reconocimiento

De los 50 muestreos que se generaron a partir de la base de datos, se utilizaron los últimos 5 para utilizarlos en este bloque de Reconocimiento de la sustancia.

Se calcula la distancia euclidiana entre los centros entrenados y las muestras de las clases a identificar, para conocer la clase a la que pertenece cada sustancia.

Las Tablas II, III, IV y V muestran las distancias euclidianas para cada sustancia a reconocer, obtenidas a partir de la base de datos. En cada una de estas tablas, la columna con una distancia euclidiana menor representa a la clase a la que pertenece la sustancia a reconocer.

TABLA II
DISTANCIA EUCLIDIANA PARA CADA ELEMENTO DEL RECONOCIMIENTO DE LA CLASE NADA

Número de muestra	Nada	Cerveza	Mezcal	Alcohol
1	2.861409	151.4458	228.8093	263.1178
2	0.976434	151.0171	228.4189	262.7385
3	1.921124	151.6954	229.1278	263.4742
4	2.785208	150.7742	228.1215	262.3919
5	4.685448	150.4741	227.8799	262.2415

TABLA III
DISTANCIA EUCLIDIANA PARA CADA ELEMENTO DEL RECONOCIMIENTO DE LA CLASE CERVEZA

Número de muestra	Nada	Cerveza	Mezcal	Alcohol
1	152.1486	1.99699	78.50402	115.2375
2	152.4305	3.032229	78.10998	114.7705
3	152.1347	2.14119	78.52715	115.2612
4	152.1469	2.876953	78.56472	115.2805
5	152.4085	1.709182	78.07712	114.7526

TABLA IV
DISTANCIA EUCLIDIANA PARA CADA ELEMENTO DEL RECONOCIMIENTO DE LA CLASE MEZCAL

Número de muestra	Nada	Cerveza	Mezcal	Alcohol
1	229.7084	80.79945	2.379967	37.57698
2	230.9161	82.02174	3.013501	36.41557
3	230.6567	81.66033	2.702482	36.52403
4	230.6563	81.66156	3.291921	36.5669
5	230.3298	81.50526	3.201741	36.67869

TABLA V
DISTANCIA EUCLIDIANA PARA CADA ELEMENTO DEL RECONOCIMIENTO DE LA CLASE ALCOHOL

Número de muestra	Nada	Cerveza	Mezcal	Alcohol
1	263.7019	117.2694	39.67693	2.508752
2	265.1113	118.718	40.93507	3.448672
3	263.4997	117.0013	39.22773	3.327404
4	264.5652	118.2791	40.72253	2.856573
5	263.7141	117.277	39.64724	1.845251

Con los resultados obtenidos en las tablas se comprueba que la sustancia reconocida sea la misma que la que se encuentre dentro de la cámara de gases.

D. Observaciones

Con estos resultados, podemos observar que el algoritmo identificó adecuadamente las sustancias. Resulta que la distancia menor a los centros, corresponde a la clase que se quiere reconocer. Esto sucedió en todos los casos.

Cabe mencionar que no se tomaron en cuenta los sensores 6, 12, 13 y 14, debido que no son relevantes para la base de datos; se trata del sensor de *humedad*, *temperatura*, uno no

asignado físicamente hablando (no existe nada conectado a esa entrada) y de *monóxido de carbono* del cual no se ha completado su configuración.

En cuanto a los sensores 7 y 9 (metano/gas LP y monóxido de carbono) fueron igualmente descartados porque existían inconsistencias en la lectura arrojada por ellos, por lo que les fue atribuido un mal funcionamiento interno.

La *temperatura* y *humedad* son únicamente para conocer las características del entorno en el que se desempeñan los sensores, pues cambia su comportamiento dependiendo de estos dos factores.

Para el experimento estos sensores se igualaron a cero, lo que genera que estén desfasados los centros de cada clase.

A pesar de que sólo algunos sensores fueron útiles, se integraron al prototipo varios para identificar distintos compuestos (no sólo alcoholes). Algunos compuestos pueden ser identificados de forma directa y otros mediante la información que arrojen los sensores en conjunto (aunque no exista algún sensor específico para dicho compuesto).

REFERENCIAS

- [1] Bayer Diabetes Care. Cetonas. [en línea]. Disponible: http://www.bayerdiabetes.com.mx/scripts/pages/es/diabetes_info/pruebas/cetonas/index.php (s.f.).
- [2] UNAM Facultad de Ingeniería Biometría Informática. Bases Teóricas y Sistemas Biométricos. [en línea]. Disponible: <http://redyseguridad.fi-p.unam.mx/proyectos/biometria/basesteoricas/reconocimiento.html> (s.f.).
- [3] Herrera Alejandro. Clasificación numérica y su aplicación en la ecología. INTEC. 2000.
- [4] J. W. Gardner and J. Yinon, *Electronic Noses & Sensors for the Detection of Explosives*, Kluwer Academic Publishers, 2004.
- [5] Peris Miguel, Escuder-Gilabert Laura. A 21st century technique for food control: Electronic noses. *Analytica Chimica Acta* 638, 1–15 (2009).
- [6] Z. Haddia, A. Amaria, H. Alamia, N. ElBarib, E. Llobetç, B. Bouchikhia. A portable electronic nose system for the identification of cannabis-based drugs. *Sensors and Actuators B* (2010).
- [7] Mahmoud Z. Iskandarani. A novel odor key technique for security applications using electronic nose system. *American Journal of Applied Sciences* 7 (8): 1118-1122, (2010).
- [8] Silvano Dragonieria, et al. An electronic nose distinguishes exhaled breath of patients with Malignant Pleural Mesothelioma from controls. *Lung Cancer* 75 326–331, (2012).
- [9] Eun Hae Oh, Hyun Seok Song, Tai Hyun Park. Recent advances in electronic and bioelectronic noses and their biomedical applications. *Enzyme and Microbial Technology* 48 427–437, (2011).
- [10] Elizabeth A. Baldwin, Jinhe Bai, Anne Plotto and Sharon Dea. Electronic noses and tongues: applications for the food and pharmaceutical industries. *Sensors* 2011, 11, 4744–4766.
- [11] James, D. et al. 'Classification of fresh edible oils using a coated piezoelectric sensor array-based electronic nose with soft computing approach for pattern recognition', *Transactions of the Institute of Measurement and Control*, 26 (1), pp.3–18 (2004).
- [12] Zhi-biao Shi, et al. Comparison of Algorithms for an Electronic Nose in Identifying Liquors. *Journal of Bionic Engineering* 5 253–257, (2008).