

Modelo Cliente-Servidor sin Control de Estado para Aprendizaje Profundo de Datos en Dispositivos IoT Aplicados a Parámetros Ambientales

Ing. Javier A. Ouret y Luciano Parodi

UCA - Facultad de Ingeniería y Ciencias Agrarias.
Ingeniería en Informática - Cátedra Protocolos de Internet
javierouret@uca.edu.ar
luciano.parodi97@gmail.com

Resumen. El crecimiento exponencial de dispositivos IoT requiere de la investigación y desarrollo de nuevas arquitecturas para la gestión de protocolos de acceso a sensores, operaciones cliente servidor y el análisis de grandes volúmenes de datos con múltiples parámetros relacionados. El objetivo de este trabajo es investigar y proponer un modelo cliente-servidor, sin control de estado, para el acceso a sensores IoT, con brokers MQTT y arquitectura REST. El modelo agrupa por medio de análisis profundo los valores de concentración de CO₂ (variable objetivo) de un lugar determinado, para luego correlacionar los resultados con los posibles efectos sobre la salud de las personas, a lo largo del tiempo. Los sensores son accesibles en tiempo real por medio de gateways GNSS (con acceso a redes celulares LTE-M1, WiFi mesh o Lorawan), monitoreados y gestionados con protocolos SNMP/Netconf [11]. La normalización de la variable se hace con datos ambientales externos obtenidos por geolocalización. Comparamos los resultados de K-NN, K-Means y GMM para el aprendizaje automático (supervisado y no supervisado) y asignación del grupo de riesgo del lugar para la variable concentración de CO₂, en rangos de tiempo. Con la información obtenida se pueden realizar acciones de corrección (o alarma) sobre otros dispositivos controlados por IoT para regular la ventilación del lugar y su capacidad operativa.

Palabras Clave: Internet of Things, IoT Sensor, Machine Learning, MQTT, RestAPI.

1 Introducción

La Internet de las Cosas (*IoT, Internet of Things) es un concepto que engloba a la interconexión digital de objetos cotidianos, sensores, software y otras tecnologías con internet, con objetos en lugar de con personas, permitiendo la gestión entre dispositivos y el ambiente para mejorar la calidad de vida de las personas. Se estima que hay más de 14 mil millones de dispositivos IoT* conectados en la actualidad, con una proyección de 27 mil millones para el 2025 [1]. En la actualidad son limitados los mecanismos normalizados para acceder y configurar cualquier dispositivo de IoT en Internet y hay muchos temas a mejorar, como la falta de visibilidad y de integración a la seguridad de las redes o el uso de protocolos abiertos que no están totalmente probados, la inyección

en la red de grandes volúmenes de datos o la vulnerabilidad de las APIs. Los protocolos existentes, como SNMP, sólo están orientados a darnos información sobre lo que ocurre en cada equipo y modificar parámetros, Protocolos como MQTT [2] de la norma OASIS [3] son usados para el transporte ligero de mensajes bajo el esquema de publicación/suscripción, utilizando poco código y un mínimo de ancho de banda. La automatización de la captura de los datos de los dispositivos IoT puede usar técnicas de aprendizaje profundo supervisado y no supervisado para asignar grupos de pertenencia a los datos obtenidos (clustering).

2 Descripción del trabajo

Se desarrolló una maqueta que implementa un modelo operativo para ser replicado en múltiples localizaciones. Entre los objetivos destacamos:

- a) Obtener estados ambientales en sitios de trabajo, esparcimiento, estudio, etc. que permitan evaluar la concentración de CO y CO₂ en distintos períodos y condiciones climáticas, y reportar su estado en tiempo real para realizar acciones correctivas inmediatas. Para ello agrupamos las mediciones de las variables de los sensores para clasificarlas y establecer el estado ambiental del lugar, utilizando: a) K-NN (K-Nearest-Neighbor, algoritmo del tipo supervisado), b) K-means y c) GMM (Gaussian Mixture Model, versión probabilística de K-means), como algoritmos de clasificación no supervisados. El modelo agrupa por medio de análisis profundo los valores de concentración de CO₂ (variable objetivo) de un lugar determinado, para luego correlacionar los resultados con los posibles efectos sobre la salud de las personas, a lo largo del tiempo. En esta etapa del trabajo nos focalizamos en la concentración de CO₂ corregido para lo cual se mide simultáneamente temperatura, humedad y presión atmosférica.
- b) Analizar diferentes algoritmos para etiquetar los datos extraídos con el sensor. No se intenta predecir valores de CO₂, sino a cuál grupo debería pertenecer cada valor de CO₂ medido y colocarle una etiqueta adecuada para su futuro estudio y visualización. Como extensión descubrir grupos de etiquetas que se puedan adaptar a los distintos ambientes en donde se extrajeron los datos, mostrando las tendencias en diferentes horarios y épocas del año.
- c) Utilizar protocolos SNMP/MQTT para la gestión de las configuraciones y los datos. Gestionar los datos en los dispositivos IoT utilizando un modelo cliente-servidor sin control de estado, con arquitectura REST [4] que busca aumentar la capacidad de HTTP para operar sobre datos en forma escalable. La mensajería de los datos se resuelve con un protocolo liviano como MQTT y el acceso a los recursos de procesamiento y visualización de aprendizaje por medio de RestAPI. Para la operación de los sensores IoT, la captura de datos, la geolocalización, la tunelización de tráfico en forma segura y el almacenamiento de datos en SQL se utiliza el lenguaje Python 3.8 con las librerías REST API con Flask y SQLAlchemy.
- d) Lograr un consumo energético mínimo (módulos de sensado alimentados por paneles solares o generadores eólicos), usando el mínimo ancho de banda posible por

medio de LTE Cat M1, una tecnología de la red celular para IoT del tipo LPWA (Low Power Wide Area).

3 Modelo de Laboratorio

Como plataforma de desarrollo e implementación, a nivel de hardware, se trabaja con Sensores IoT SCD30 Sensirion y un Gateway IoT Lantronix E210 para la conecti-

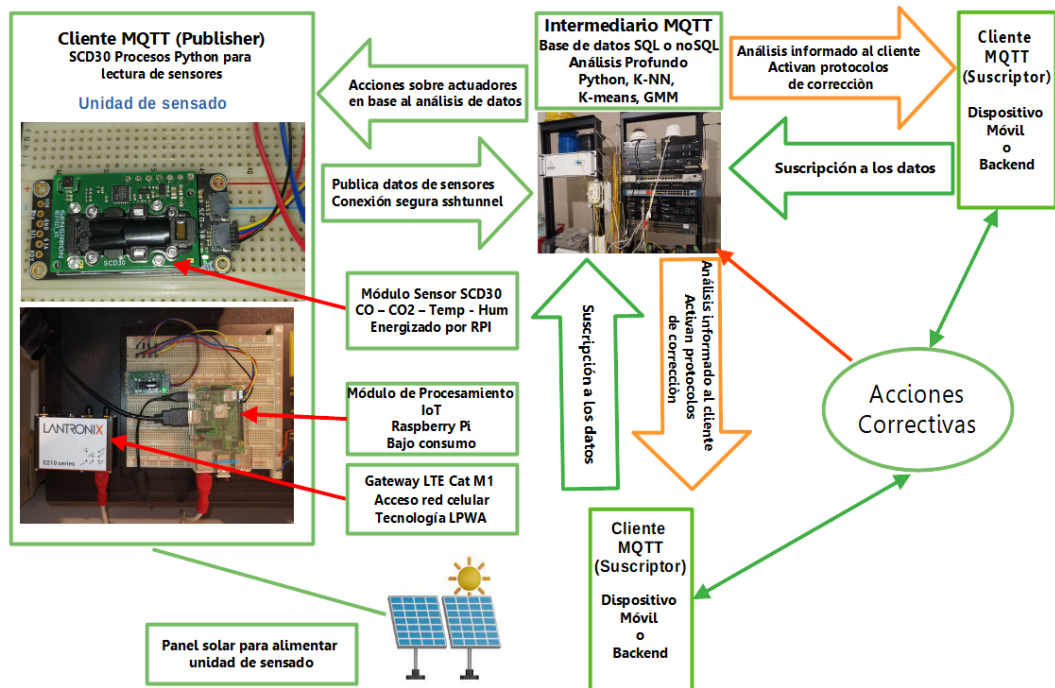


Fig. 1. Maqueta primaria utilizada para modelo funcional

vidad a internet por medio de LTE Cat M1. Los sensores SCD30 realizan la medición de CO2 por el principio de NDIR de 2 mediciones comparadas, en el rango de 0-1% (0-10.000 ppm), integrando sensores de humedad y temperatura en el mismo módulo. Para el procesamiento de la información y la conectividad el sensor se conecta a una Raspberry Pi con el sistema operativo Linux Raspbian por medio de un bus I2C ya integrado al PCB.

4 Ejecución del modelo cliente-servidor y análisis de mediciones

En esta etapa se trabajó con clasificadores de concentración de CO2 “compensado o normalizado” en condiciones ambientales variables de temperatura, humedad y

presión, para una muestra de entrenamiento de 21100 registros de CO2, temperatura, humedad y presión ingresados a una tabla SQL para su análisis. Los programas en Python utilizan scikit-learn [8] para el cálculo y gráficos. En las concentraciones a las que se encuentra habitualmente se considera como indicador de calidad del aire, ya que la principal fuente de emisión en interiores son las propias personas. En la [Tabla 1] observamos los niveles de clasificación elegidos. Se etiquetan los valores de CO2 extraídos, se divide el conjunto de datos en “Datos de Prueba” y “Datos Medidos” para entrenar y poner a prueba el algoritmo KNN respectivamente y se mide la precisión del algoritmo comparando las etiquetas predichas con las reales. Para los datos de prueba se utiliza un $K = 5$ que presenta un equilibrio entre un modelo ruidoso con valores atípicos y posibles datos mal etiquetados (número bajo $K=1$ o $K=2$), o un K grande, que provoca un ajuste insuficiente. Números pares pueden darnos resultados aleatorios si el valor está a la misma distancia de 2 clases conocidas. La precisión del modelo para clasificar nuevos datos da valores del orden de 83% para $K=5$ en K-NN. Para realizar el agrupamiento con K-means, se aplica la normalización de la variable CO2 ($ppm\ CO2\ corregido = ppm\ CO2\ medido \cdot ((T_{medido} \cdot Pref) / (P_{medido} \cdot Tref))$). Si los datos no se normalizan, los grupos resultantes se verán influenciados por la magnitud de los valores contenidos en la variable, en lugar de su importancia relativa [9]. En el caso de K-means se definen aleatoriamente los centroides, que son el centro de las clasificaciones, luego los puntos de los datos se asignan a grupos según el centroide al que se encuentran más cerca por la distancia euclidiana. Estos pasos se repiten hasta que los centroides convergen (“no se mueven”) y se reubican en el promedio de los puntos dentro de cada clasificación. Observar las “estrellas” en el gráfico K-means de la Fig.4. Finalmente se aplica GMM donde observamos una clasificación resultante en grupos más cercanos a los esperados, con un menor error de asignación de nuevos datos a cada clase y agregamos un cierto grado de incertidumbre a la clasificación, lo que genera límites más suaves entre las clases encontradas, con respecto a lo que se obtiene con K-means.

CONCENTRACIÓN	EFECTO
200 - 450 ppm	Concentración atmosférica típica
600 - 800 ppm	Calidad del aire interno aceptable
900-1200 ppm	Calidad del aire interno tolerable
2000- 5000 ppm	Límite promedio de exposición en un período de ocho horas
6000 - 30000 ppm	Preocupación, solo exposición breve
3 - 8%	Incremento de la frecuencia respiratoria, dolor de cabeza
> 10%	Náuseas, vómitos, pérdida de conocimiento
> 20%	Pérdida de conocimiento repentina, muerte

Tabla 1. Efectos de la concentración de CO2 Una elevada concentración conduce a la asfixia por desplazamiento de oxígeno. [5][10]

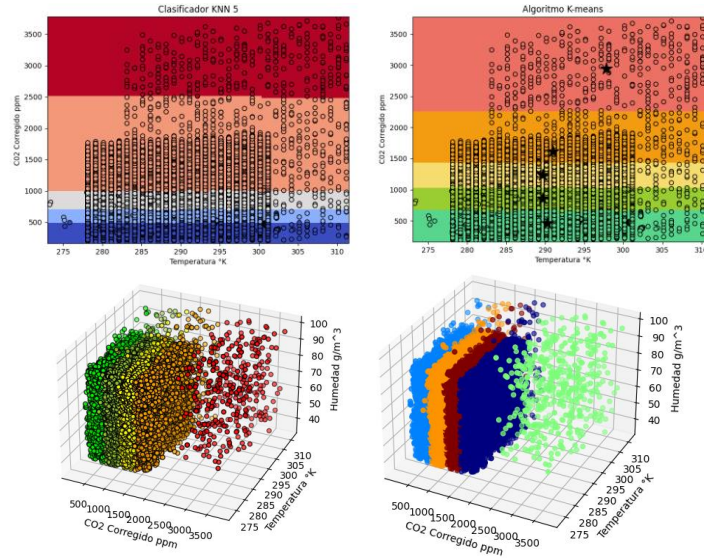


Fig. 3. Gráficos para K=5, K-NN, K-means, K-means 3D y GMM 3D.

5 Conclusiones

La efectividad de los modelos supervisados y no supervisados se midió en base a diferentes factores como su complejidad computacional, su velocidad para operar los volúmenes de datos requeridos y su capacidad para adaptarse a nuestro conjunto de datos ya sea en 2 o 3 dimensiones. Si bien se observan similitudes en los resultados obtenidos (con los diferentes algoritmos y configuraciones), K.Means y GMM (como era de esperarse) fueron más precisos para generar los grupos de clasificación (Fig.3), con GMM presentando menos dispersión en los límites de cada grupo. Este trabajo concluye la etapa de prototipo del proyecto y se ha logrado la integración de sensores a una arquitectura REST para IoT sobre la red de telefonía celular, utilizando protocolos livianos como MQTT y LTE Cat M para transferir los datos a un gestor intermediario MQTT con un motor SQL, para el análisis de profundo de datos, y el uso de clientes RestAPI con rutas a los clasificadores para luego tomar acciones de corrección de los niveles de CO2 sobre otros actuadores que permitan corregir las dispersiones o, generar los niveles de alarma correspondientes. Los procesos de clasificación en esta etapa se limitaron a verificar los conjuntos de datos de aprendizaje y el nivel de error para clasificar nuevos datos. Con respecto a los algoritmos de clasificación se obtuvieron resultados similares con $K=5$ y $K=3$, una menor tasa de error usando $K=5$, y tanto con K-mean como con GMM se obtienen grupos de clasificación menos dispersos que con K-NN. La clasificación obtenida coincide con la esperada [Tabla 1] en base a los datos de prueba.

6 Siguientes pasos. Trabajos Futuros.

Pasar de la etapa de prototipo a la fase de producción para el modelo desarrollado distribuyendo el módulo de sensado en múltiples localizaciones con diferentes condiciones ambientales. Se adaptará el modelo para ambientes mixtos calefaccionados con existencia de CO y CO₂, agregando un Gravity MEMS Gas Sensor al mismo módulo de sensado. A diferencia del CO₂, el CO no se genera en forma natural, sino por combustión incompleta por falta de aire. Se buscará aplicar el modelo para realizar el análisis de datos por K-NN, K-Means y GMM, contemplando la relación estadística entre los 2 valores haciendo uso del coeficiente de correlación de Pearson [6]. Aplicado a una cantidad de sensores ubicados en diferentes localizaciones para analizar los resultados. Para volúmenes de datos masivos se analizará el uso de base de datos tipo NoSQL para la etapa de clasificación. El código desarrollado y detalle de los prototipos está disponible en: <https://github.com/jaouret/IoT01> . Enviar email a javierouret@uca.edu.ar para recibir el permiso de lectura.

Referencias

1. IOT Analytics. Disponible: <https://iot-analytics.com/number-connected-iot-devices>.
2. MQTT Message Queuing Telemetry Transport. Disponible; <https://mqtt.org/>. MQTT Versión 5.0 Estándar OASIS. (Marzo 2019).
3. OASIS. Organization for the Advancement of Structured Information Standards. Disponible: <https://www.oasis-open.org/>
4. Roy Thomas Fielding. Representational State Transfer (REST). Architectural Styles and the Design of Network-based Software Architectures dissertation. 2000. University of California, Irvine.
5. Vaisala. <https://es.vaisala.com>. Cómo medir el dióxido de carbono. Nota de Aplicación. 2013.
6. Fridama. Uso de la correlación estadística para la gestión energética en locales de oficina empleando técnicas de la Industria 4.0. Revista Ciencia de los Ordenadores. (Nov-Dic 2018).
7. Theresa Henle. (Abril 2017). Analyzing Development through K-means Clustering and Matching.
8. Disponible: <https://scikit-learn.org/stable/>
9. Padhraic Smyth, “Mixture Models and the EM Algorithm”, Department of Computer Science, University of California, Irvine, 2017
10. T. Teleszewski1 · K. Gładyszewska-Fiedoruk1. “The concentration of carbon dioxide in conference rooms: a simplified model and experimental verification”. Published online: 27 May 2019.
11. J. Ouret, I. Parravicini. Reconfiguración Automática de Parámetros de Calidad de Servicio Dispositivos por medio del Protocolo NETCONF. 48 JAIIO. 2019.