**Aprendizaje profundo para datos multimedia en IoT**

**Previo**

El Internet de las Cosas (IoT, Internet of Things) está configurando el mundo industrial y de los consumidores. Las tecnologías inteligentes siguen permeando todas las capas de los sectores empresarial y de consumo, desde los comercios hasta la salud, pasando por las finanzas o la logística. Los 4 componentes del modelo del IoT: Sensores, Redes (Comunicaciones), Analítica (la nube) y Aplicaciones.

Un dispositivo IoT consiste en un objeto al que se le ha dotado de conexión a Internet y cierta inteligencia software, sobre el que se pueden medir parámetros físicos o actuar remotamente y que por tanto permite generar un ecosistema de servicios alrededor del mismo.

**Artículo**

**Resumen (Diapositiva 1)**

Con la llegada de Internet que ha llevado a la proliferación de grandes cantidades de datos multimedia, el análisis de los datos multimedia agregados ha demostrado ser una de las áreas activas de investigación y estudio. Los datos multimedia incluyen audio, video, imágenes asociadas con aplicaciones como búsquedas de similitudes, resolución de entidades y clasificación.

La minería de datos multimedia incluye muchos desafíos como el volumen de datos, la variedad y la naturaleza no estructurada, no estacionaria y en tiempo real. Necesita capacidades de procesamiento avanzadas para tomar decisiones casi en tiempo real.

Para procesar cantidades tan grandes de datos, se pueden utilizar técnicas avanzadas como el aprendizaje automático, métodos de aprendizaje profundo.

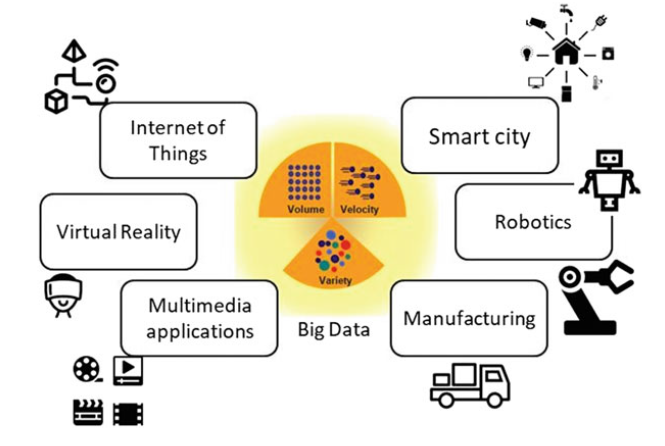
El avance de los teléfonos inteligentes ha dado como resultado aplicaciones basadas en el fitness basadas en el número de pasos caminados, el recuento de calorías, los kilómetros recorridos, etc. Todos estos tipos de datos pueden clasificarse como datos multimedia para Internet de las cosas (IoT).

**Capítulo 1**: Introducción a multimedia e IoT

**(Diapositiva 2)** La minería de datos ha ganado una mayor importancia en los últimos días debido a la gran cantidad de datos que se recopilan y su disponibilidad a través de Internet. Los importantes avances en la tecnología y las herramientas de big data han llevado al desarrollo de análisis e investigación en esta área a lo largo de los años. Hay un cambio de paradigma de la minería de datos hacia el análisis de big data que involucra varias etapas de procesamiento y análisis. La analítica de big data consiste en la exploración de datos, identificando las relaciones entre las diferentes características de los datos y visualizándolas. 

Los datos multimedia incluyen imágenes, video, audio y texto. Debido a la llegada de los teléfonos inteligentes e Internet, este tipo de datos multimedia debe conducir a varias aplicaciones y plataformas para compartir. Las fuentes valiosas de datos multimedia son las redes sociales como Instagram, YouTube, Twitter y Facebook. Esta rica fuente de información se denomina "Big data". Los macrodatos poseen tres características esenciales: variedad, velocidad y volumen.

**(Diapositiva 3)** Internet de las cosas (IoT) es una red de cosas que están conectadas entre sí y que es compatible con Internet para la comunicación. La mayoría de los dispositivos de hoy en día están basados en sensores y normalmente están conectados a uno o más. Con el aumento en la cantidad de dispositivos que están conectados entre sí, también aumenta la generación de datos multimedia. El alcance de IoT no se limita solo a los datos del sensor, sino que también se relaciona con multimedia. Por ejemplo, la cámara CCTV captura los datos de video con fines de vigilancia que se pueden clasificar como datos multimedia. Los datos de huellas dactilares también se pueden considerar como datos multimedia ya que están en forma de imagen.



**(Diapositiva 4)** La analítica multimedia se puede llevar a cabo utilizando enfoques analíticos convencionales, como agrupamiento, métodos de árbol de decisiones. Sin embargo, un desafío importante al que se enfrentan los enfoques convencionales es la

escalabilidad de los métodos y el tiempo de ejecución necesario para producir los resultados. En este sentido, los métodos de computación GPU y Deep Learning se utilizan para análisis multimedia. Los métodos de aprendizaje profundo que involucran redes de avance profundo, las redes recurrentes se utilizan la mayoría de las veces para análisis multimedia. El principal desafío del aprendizaje profundo es que el proceso de extracción de características involucra etapas de reconocimiento de objetos, detección de objetos desde los bordes identificados en la imagen.

**Capítulo 2**: Avances en datos multimedia

**(Diapositiva 5)** En el mundo de la analítica de datos, los datos se han multiplicado de diferentes formas. Los datos multimedia son uno de esos datos que se han generado recientemente en grandes volúmenes en Internet. Está siendo capturado por varios dispositivos informáticos multimedia como computadoras, tabletas, teléfonos móviles y cámaras.

|  |  |
| --- | --- |
|  | **Datos visuales**: La forma más común de datos multimedia que se encuentran son los datos de video. Un dato visual consiste en una secuencia de secuencias de imágenes que deben analizarse una por una. La mayor parte de la información no estructurada existe en forma de datos visuales y contiene información muy rica. El proceso de análisis de datos visuales implica extraer información significativa de las diferentes secuencias de imágenes que están presentes en los datos visuales. |
|  | **Datos de audio**: Otro tipo de datos multimedia que se utiliza principalmente es el audio / voz. datos. Se necesitan aplicaciones analíticas de audio en tiempo real en las redes sociales, la atención médica y la industria. El análisis de audio implica extraer información útil de las diferentes partes de la información presente en los datos de audio. Las plataformas de big data como Spark, Hadoop y diferentes bibliotecas se utilizan ampliamente para este tipo de análisis de audio. |
|  | **Datos de texto**: Los datos multimedia se pueden incrustar en el contexto textual en la forma de páginas web, encuestas, feeds y metadatos. El análisis de dichos datos ayuda a obtener información interesante. Los datos multimedia en forma de texto pueden estar estructurados o no estructurados. El tipo de datos estructurados se puede analizar con la ayuda de las técnicas tradicionales de recuperación de consultas de bases de datos relacionales. Sin embargo, los datos multimedia en forma de feeds no están estructurados y deben transformarse en un formato estructurado para su posterior análisis |
|  | **Datos del sensor**: IoT está desempeñando un papel importante hoy en día y los sensores están presentes en casi todas partes. Los sensores están equipados no solo para capturar los datos, sino también para aplicar análisis en tiempo real. Es muy difícil analizar estos datos y desarrollar una aplicación analítica basada en ellos. Las aplicaciones de datos de sensores son muy apreciadas en las ciencias astronómicas para patrones meteorológicos, acondicionamiento de satélites y monitoreo. |
|  | **Redes sociales**: La principal fuente de datos multimedia son las redes sociales. los los avances en las redes sociales y el intercambio que comenzaron con un texto normal ahora han llegado a imágenes, videos, videos en vivo, grupos públicos, etc.  La principal característica de los datos multimedia es la variedad. Existe en diferentes formas y en diversas fuentes. Estos avances en datos y análisis multimedia han permitido varios desafíos y tecnologías para desarrollar diferentes aplicaciones. |

**Capítulo 3**: Desafíos en datos multimedia

**(Diapositiva 6)** Los datos multimedia involucran los datos de diversas fuentes como cámaras, redes sociales, sensores y otras de naturaleza heterogénea. Con el fin de llevar a cabo análisis de dichos datos, la naturaleza de heterogeneidad de dichos datos debe transformarse con fines de análisis. La transformación de los datos implica convertir los

datos de diferentes formatos a un formato singular para el análisis.

FALTA

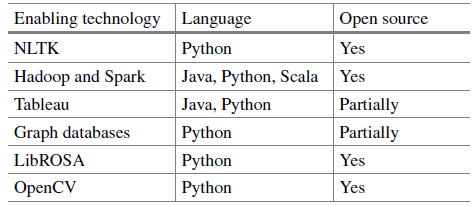
**Capítulo 4**: Tecnologías habilitadoras

**(Diapositiva)** El análisis multimedia consiste en analizar datos multimedia como audio, video y texto para identificar patrones en los datos y tomar decisiones basadas en los patrones encontrados. Hay varias etapas de análisis multimedia como preprocesamiento de datos, extracción de características y análisis. En cada etapa, se utilizan diferentes tipos de

tecnologías para aprovechar el proceso de análisis y toma de decisiones.

Los datos multimedia se clasifican en tres partes, a saber, análisis de texto, análisis de video y análisis de audio. Las tecnologías habilitadoras o disponibles en cada una de las categorías se resumen a continuación.

**(Diapositiva)**



**(Diapositiva) Análisis de texto**

La forma más común de datos multimedia es el texto. Es de naturaleza muy volátil ya que el texto escrito por humanos no está estructurado y difiere de una persona a otra. Hay varias etapas de análisis de texto, como la eliminación de palabras vacías como "es", "era", "esto". Por ejemplo, en el caso del análisis sentimental, las frases clave del texto son importantes para el análisis en lugar de las palabras vacías en el texto.

|  |  |
| --- | --- |
|  | **NLTK**  Es una de las famosas herramientas basadas en Python que se utilizan para el procesamiento del lenguaje natural.Por ejemplo, en el escenario del análisis sentimental se necesita un corpus de información que identifique los sentimientos positivos y negativos. |
|  | **Deep Learning**  Los textos que no se pueden capturar a través de la World Wide Web se analizan utilizando técnicas de aprendizaje profundo como las arquitecturas CNN y RNN. Algunos de los ejemplos incluyen la detección de acciones en video, la identificación de desastres en los conjuntos de datos rastreados en la web. Las plataformas de aprendizaje profundo como Pytorch y Tensorflow se utilizan ampliamente para el análisis de texto en sistemas multimedia. |
|  | **Scikit learn**  Es uno de los kits de herramientas científicas disponibles en la plataforma Python. En el análisis de texto, se requiere una bolsa de palabras para identificar las palabras más significativas que están presentes en el conjunto de datos de texto considerado. Scikit learn proporciona funciones como "bolsa de palabras", "tf-idf" que son necesarias para el análisis de texto. |
|  | **Hadoop and Spark**  Hadoop es una de las plataformas de código abierto que se puede utilizar para analizar cualquier formato de datos, puede extraer los datos en forma de texto presentes en el sistema de archivos.  Apache Spark es un marco de código abierto para el análisis de datos en tiempo real en un entorno informático distribuido. Se basa en Hadoop MapReduce, diseñado para un cálculo rápido. |

**(Diapositiva) Análisis de video**

La mayor parte de los datos multimedia está en forma de video. El análisis de datos visuales es un proceso complicado, ya que implica una serie de secuencias de imágenes. Aparte de las secuencias de imágenes, los datos visuales pueden estar en forma de gráficos y redes de texto.

|  |  |
| --- | --- |
|  | **Redes neuronales convolucionales (CNN) mediante Depp Learning**  CNN es la técnica más utilizada para el análisis de video. Las aplicaciones de CNN al análisis de video incluyen detección de objetos, detección de delitos, etc. En CNN, la secuencia de imágenes se divide primero en varios píxeles y luego se reduce a escala para la detección de objetos. La imagen se divide primero en varios filtros que representan una matriz de valores de la imagen. |
|  | **Tableau**  Los datos multimedia involucran datos de redes sociales que pueden estar en forma de feeds o tweets. La visualización de dichos datos ayuda a revelar los patrones interesantes sobre los datos. Tableau es una de esas herramientas que ayuda a comprender los datos al visualizar primero los datos y luego realizar el análisis |
|  | **OpenCV**  OpenCV es una de las plataformas más utilizadas para análisis de video. incluyen sistema de reconocimiento facial, reconocimiento de gestos, seguimiento de movimiento, realidad aumentada, robótica móvil. |

**(Diapositiva) Análisis de audio**

Los datos multimedia que se encuentran en forma de audio necesitan una gran cantidad de computación y tiempo para llegar a resultados. Los datos presentes en el formato de audio deben analizarse etapa por etapa, ya que difieren de vez en cuando.

|  |  |
| --- | --- |
|  | **LibROSA**  Es una de las bibliotecas que está disponible en Python para el análisis de varios archivos de audio. Ayuda en el desarrollo de varias aplicaciones de audio basadas en diferentes formatos. Ayuda a extraer las características de los archivos de música como ritmo, tempo, tiempos, series de tiempo de audio, espectrograma de potencia del audio, rollo de frecuencia, planitud espectral, etc. |
|  | **Deep learning**  El aprendizaje profundo se utiliza para el análisis de audio utilizando las redes neuronales artificiales, como las redes neuronales recurrentes (RNN), las redes neuronales convolucionales (CNN) y las redes de retroalimentación |

**Capítulo 5**: métodos de Deep Learning

**(Diapositiva)** En cualquier conjunto de datos, las características / atributos juegan un papel importante en el desarrollo de algoritmos de aprendizaje. En los métodos de aprendizaje profundo, estas características ayudan en el desarrollo de algoritmos de aprendizaje adecuados.

Una característica se puede definir como la parte interesante del conjunto de datos que actúa como punto de partida para el aprendizaje. Hay tres etapas en cualquier algoritmo de detección de características: extracción, selección y clasificación.

Una característica es una función que representa el valor cuantificado de un objeto o conjunto de datos. Representa las características importantes del objeto, como el color, la textura y los píxeles.

**(Diapositiva)** Métodos de extracción de características locales

Las características locales de un conjunto de datos o imagen se refieren a las características de la imagen que se adhiere a la detección de bordes, la segmentación de la imagen y las características que se calculan sobre los resultados de la subdivisión de la imagen.

|  |  |
| --- | --- |
|  | **Detección de esquinas Harris**  La detección de esquinas se utiliza en los sistemas de visión por computadora para inferir algunos de los datos interesantes de la imagen y sus esquinas. La idea básica es observar un gran cambio en la apariencia cuando hay un desplazamiento en la ventana en cualquier dirección. |
|  | **Transformación de características invariantes de escala (SIFT)**  Se utiliza para detectar características locales en la imagen para aplicaciones tales como reconocimiento de gestos, seguimiento de video, modelado 3D, etc. Un conjunto de características clave en la imagen se extraen y almacenan en una base de datos. Para una nueva imagen, se comparan las características contra la base de datos existente para determinar la mejor coincidencia. |
|  | **BRIEF orientado rápido y girado (ORB)**  ORB es básicamente una fusión del detector de puntos clave FAST y el descriptor BREVE con muchas modificaciones para mejorar el rendimiento. Primero usa FAST para encontrar puntos clave, luego aplica la medida de esquina de Harris para encontrar los N puntos principales entre ellos. |

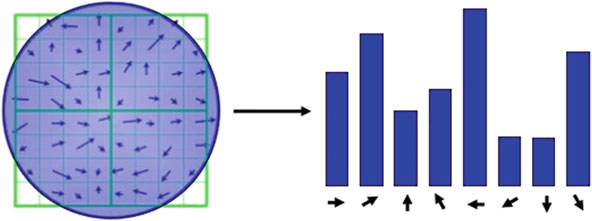
**(Diapositiva)** Métodos de extracción de características globales

Una vez que las características locales, como los bordes de las imágenes o conjunto de datos, se extraen de la imagen, las características globales se calculan para toda la imagen. Estas características globales ayudan en la clasificación final de la imagen utilizando los elementos calculados de las características locales.

**Histograma de gradientes orientados (HOG)**

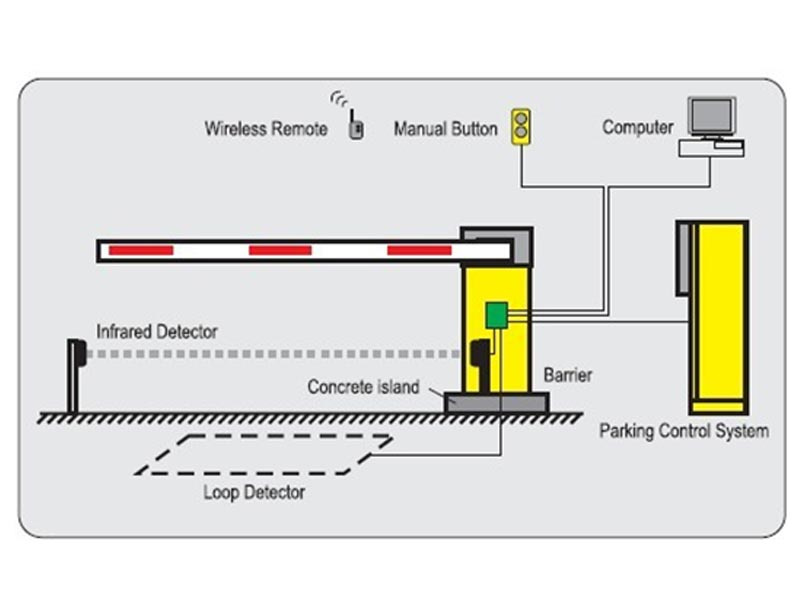
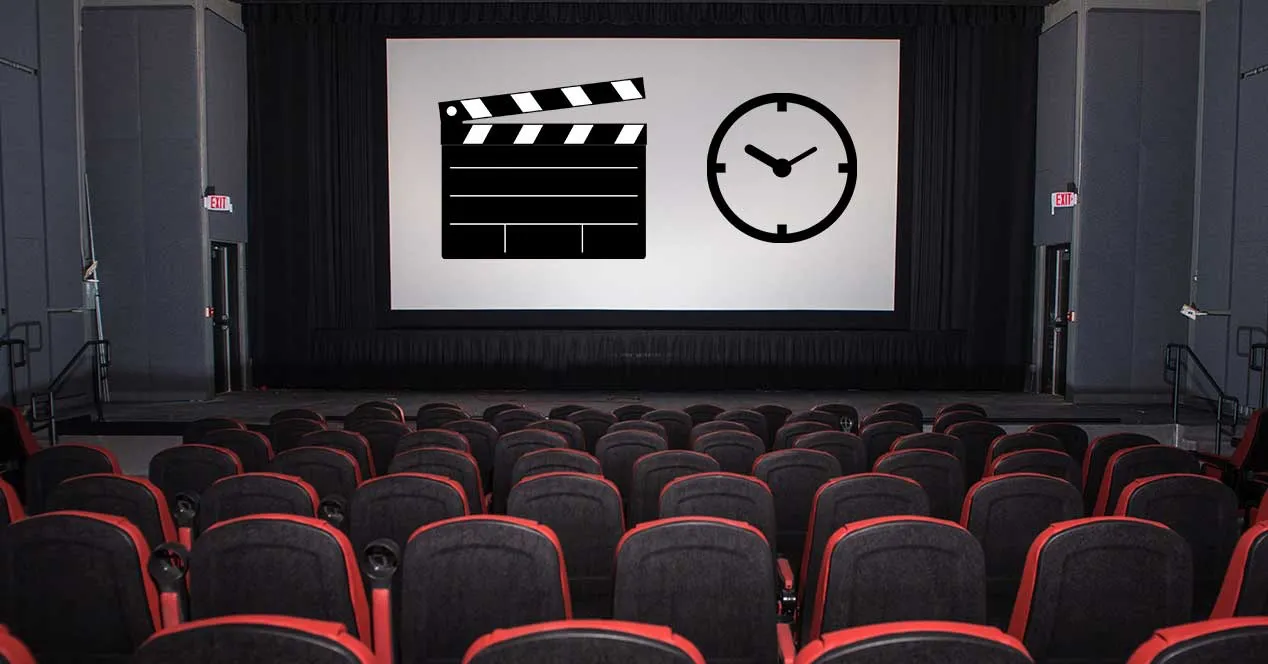
HOG se utiliza como descriptor de funciones en el procesamiento de imágenes y la visión por computadora para la detección de objetos. Se basa en el histograma de gradientes que actúa como una característica global para la detección de objetos. En partes locales de la imagen, cuenta las ocurrencias de orientación de gradiente para producir el histograma. La intensidad de la distribución de los histogramas presentes en las apariencias de los objetos locales guía la producción de los histogramas.



****

**Capítulo 6**:

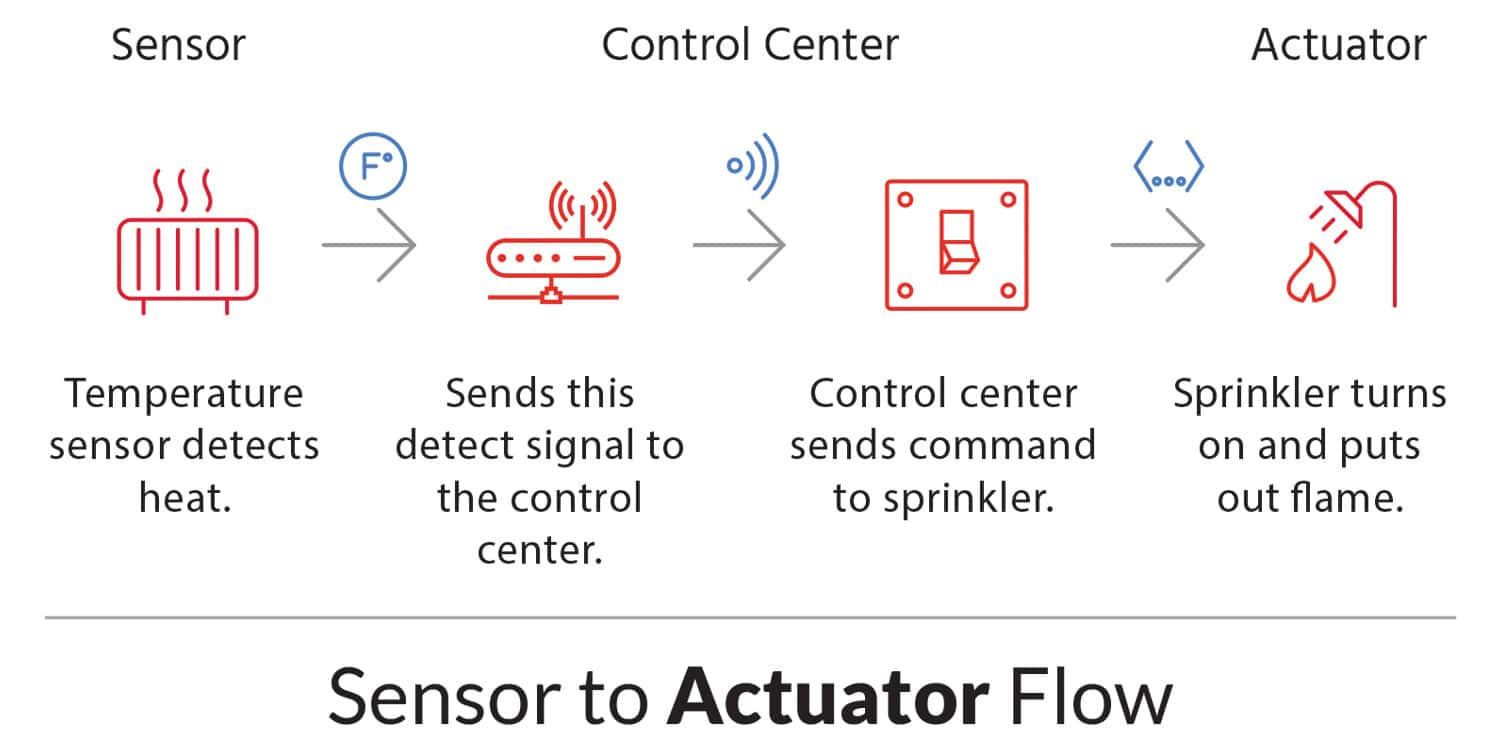
Precisamente como ya se ha venido comentando la magnitud de dispositivos interconectados va en aumento, y no solo aumentan las conexiones también se diversifican. Ahora, a partir de las conexiones que se generan entre los dispositivos se empiezan a constituir diferencias entre ellos dada su relación con los datos. Por ejemplo, el sistema de lectura de lectura de tarjetas en un parqueadero; uno acerca una tarjeta a un lector el cual genera un “aprobado/rechazado” y a partir de esa información se obtiene una acción.

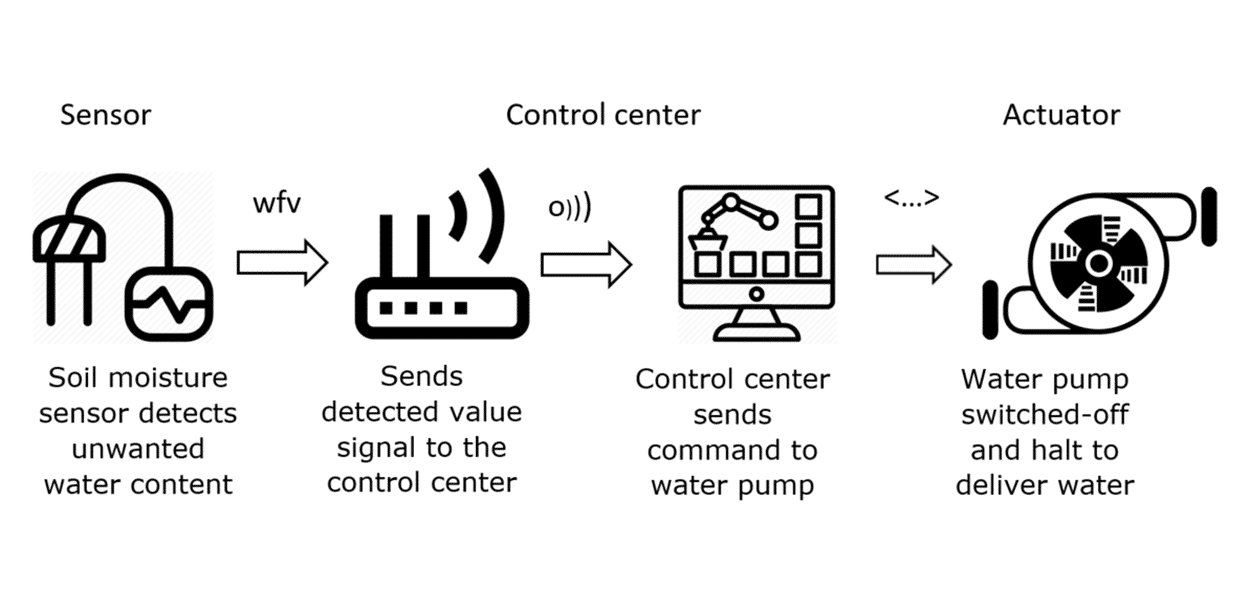
A veces la relación entre los dispositivos puede perderse, pues asumimos que un conjunto de dispositivos es un todo sin considerar sus partes. Entonces, pensemos que la diferencia que surge entre los dispositivos es algo tan natural como nuestro cuerpo y su forma de procesar estímulo. Por ejemplo, cuando estamos observando algo, nosotros somos concientes que son nuestros ojos y sus componentes quienes son capaces de decodificar la luz de forma tal que luego nuestro cerebro sea capaz de procesar esa información y convertirla en una imagen. 

Bueno, los ejemplos anteriores buscan evidenciar cual es la diferencia entre los dispositivos. A grandes rasgos, la diferencia está en que un grupo de dispositivos se centran en capturar información del ambiente y otros en tomar acciones a partir de la información del ambiente.

(Diapositiva sensores/ejecutores)

En la siguiente imagen se muestra como se distinguen los dispositivos a partir de su relación con la información



La diferenciación entre los dispositivos se genera a partir de su relación con los datos. Un grupo de dispositivos, lo sensores, se encargan de reunir información mientras que los ejecutores requieren de datos del ambiente para realizar un análisis y tomar una decisión.

Sensor:

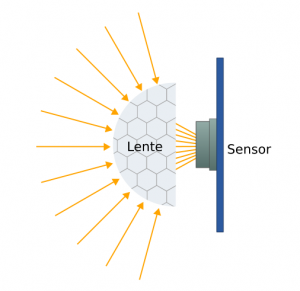


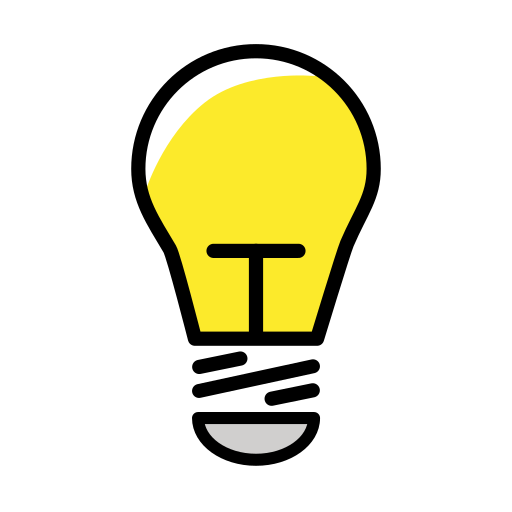
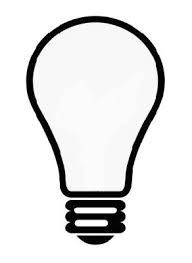
Ejecutor:



**(diapositiva)** Diagrama 2 tipos de dispositivos -> 2 tipos de datos

**6.1.1 Analysis Data:**

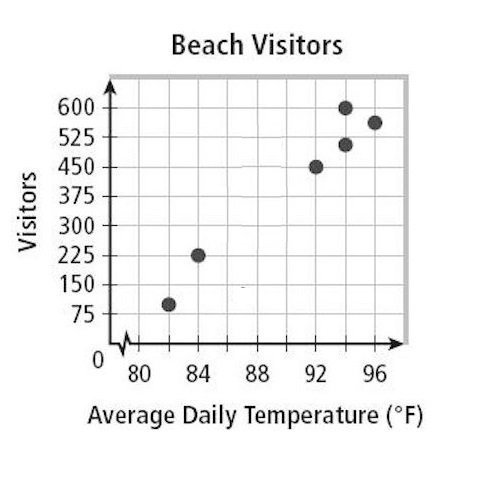
Los datos de análisis es un término que refiere a toda la información que los dispositivos capturan del ambiente, es decir son los datos provenientes de las lecturas de sensores. Estas lecturas pueden ser tan simples como respuestas a movimiento o tan complejas como geolocalización.





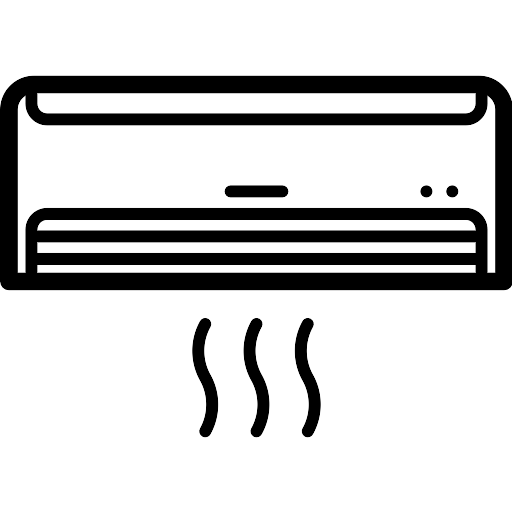
**(diapositiva)** Los datos de análisis son precisamente esos de los que podemos generar estadísticas que nos ayuden a comprender una situación. Es decir, son datos de los cuales se puede inferir información que abra paso para la automatización.

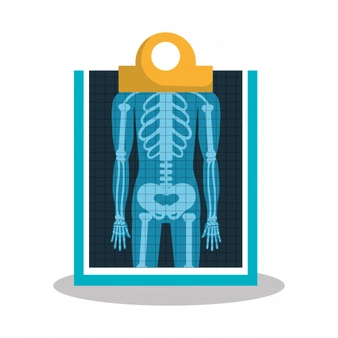




**(diapositiva)**  Los dispositivos que recolectan información para análisis son de un amplio espectro. Algunos reciben mayor protagonismo que otros, sin embargo están presentes en nuestro día a día facilitando actividades. Tenemos interacción con ellos desde lo más sencillo como un botón hasta lo más complejo como la membrana táctil de nuestros celulares, tabletas o computadores.







**6.1.2 Actuation Data**

Un paso adelante frente al análisis estadístico