

diciembre, 2023

Informe Trabajo Inteligencia Ambiental

Alberto García Martín



VNiVERSiDAD
D SALAMANCA

Máster en Sistemas Inteligentes
Universidad de Salamanca

Índice

1. Introducción	1
2. Preprocesamiento de datos	2
3. Importancia de características	5
4. Entrenamiento de modelos	7
5. Evaluación de modelos	9
6. Conclusión	11
Referencias	12

1. Introducción

En este informe se va a describir el trabajo realizado para la asignatura de Inteligencia Ambiental del máster de Sistemas Inteligentes. En este caso se ha desarrollado un algoritmo de tratamiento de datos de inteligencia artificial para la predicción del tipo de actividades físicas realizadas por una persona según los datos de un *smartwatch*. Este caso de estudio se enmarcaría dentro del campo de la inteligencia ambiental al ser un sistema que se encarga de recoger datos del entorno mediante un dispositivo *wearable* con el objetivo de procesarlos para obtener información relevante para el usuario o un profesional médico.

Para el desarrollo de este trabajo se ha utilizado un conjunto de datos publicado en 2020 por Fuller et al.[1] En este conjunto de datos se recogen datos de 49 participantes (23 hombres y 26 mujeres) utilizando dos tipos de *smartwatch* distintos, un Apple Watch Series 2 y un Fitbit Charge HR2. Los participantes completaron un circuito de 65 minutos que consistía en 40 minutos en total en una máquina de correr y 25 minutos descansando sentado o tumbado. El estudio clasifica las actividades realizadas en seis tipos según su intensidad: tumbado, sentado, andando, corriendo a 3 MET¹, corriendo a 5 MET, y corriendo a 7 MET. Los datos fueron recogidos al minuto incluyendo la frecuencia cardíaca, pasos dados, distancia recorrida y calorías quemadas.

El conjunto de datos en total incluye 6264 minutos de datos recopilados por los *smartwatches*, divididos en 3656 y 2608 minutos entre el Apple Watch y el Fitbit respectivamente. El objetivo del modelo a desarrollar será el de predecir la actividad realizada, entre sentado, tumbado y las diferentes intensidades de correr y andar, y determinar la calidad del modelo. Para esta tarea se hará uso de la plataforma de *machine learning* DeepInt², que permitirá desarrollar distintos modelos de aprendizaje automático y compararlos entre sí.

¹El equivalente metabólico de una tarea (MET), o sencillamente el equivalente metabólico, es una medida fisiológica que expresa la intensidad de las actividades físicas. Un MET es el equivalente de la energía gastada por una persona mientras está sentada en reposo.

²<https://deepint.net/>

2. Preprocesamiento de datos

El conjunto de datos a utilizar contiene cuatro características relativas a los participantes, cuatro medidas durante la actividad realizada, ocho derivadas de las mediciones, una para indicar el dispositivo utilizado en las mediciones, y por último la etiqueta de la categoría de la actividad. En la Tabla 1 se puede ver un resumen de las características del conjunto de datos.

El primer paso del preprocesamiento de datos es cargarlos en DeepInt, para ello se sube el archivo CSV tal cual a la plataforma. El conjunto de datos con el que se trabaja ha sido limpiado con anterioridad, por lo que no hay valores vacíos o no válidos, no obstante hay que eliminar las dos primeras columnas correspondientes al identificador del registro, ya que no aportan ninguna información relevante para el modelo. Para ello se accede al menú de edición de la fuente de datos y se eliminan las dos primeras columnas. Además, hay características que se han identificado con un tipo incorrecto, como la del sexo que se ha identificado como numérica en lugar de categórica, por lo que se cambia el tipo de dato.

Una vez se ha preprocesado correctamente el conjunto de datos se puede ver la distribución de las clases en el conjunto de datos. Para ello, se crea una nueva visualización en DeepInt, donde se crea un diagrama de sectores donde los gajos representarían las distintas actividades y su valor corresponderá al número total de registros de cada clase. En la Figura 1 se puede ver el diagrama de sectores resultante. Como se puede observar, la distribución está razonablemente equilibrada, aunque hay una ligera tendencia al alza en la actividad “Lying”, la distribución es lo suficientemente equilibrada como para no tener que realizar un sobremuestreo o submuestreo de las clases.

Para evitar problemas debidos a las diferencias en la medición de las características entre los dos dispositivos, se ha decidido separar el conjunto de datos en dos, uno para cada dispositivo. Para ello se crean dos nuevas fuentes de datos en DeepInt, ambas derivadas del conjunto de datos anterior. En el primero se indica que la variable “device” debe ser igual a “apple watch”, y en el segundo que debe ser igual a “fitbit”. De esta forma se obtienen dos conjuntos de datos, uno para cada dispositivo, con las mismas características a las anteriores, exceptuando la de dispositivo, que se descarta.

Si se clona la visualización anterior y se cambia el origen de los datos a los dos nuevos conjuntos de datos, se puede ver la distribución de las clases en cada uno de los dispositivos. Como se puede ver en la Figura 2 y en la Figura 3, la distribución de las clases sigue manteniéndose razonablemente equilibrada en ambos conjuntos.

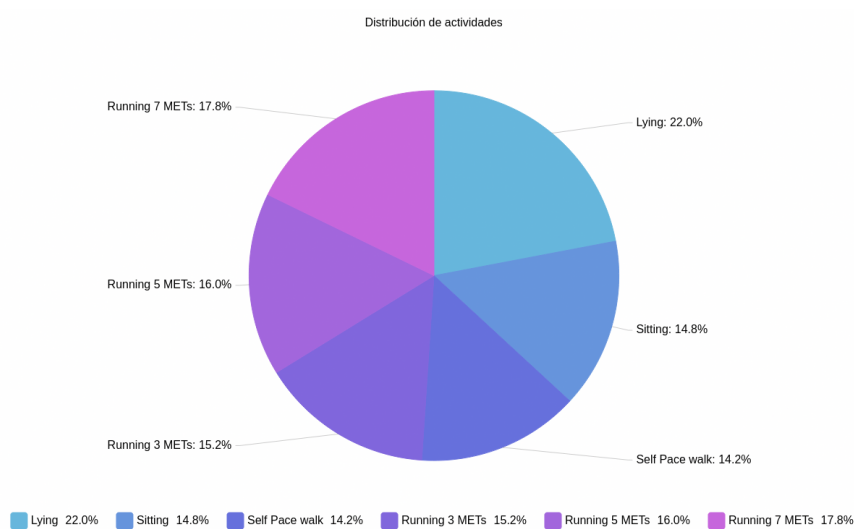


Figura 1: Distribución de las actividades.

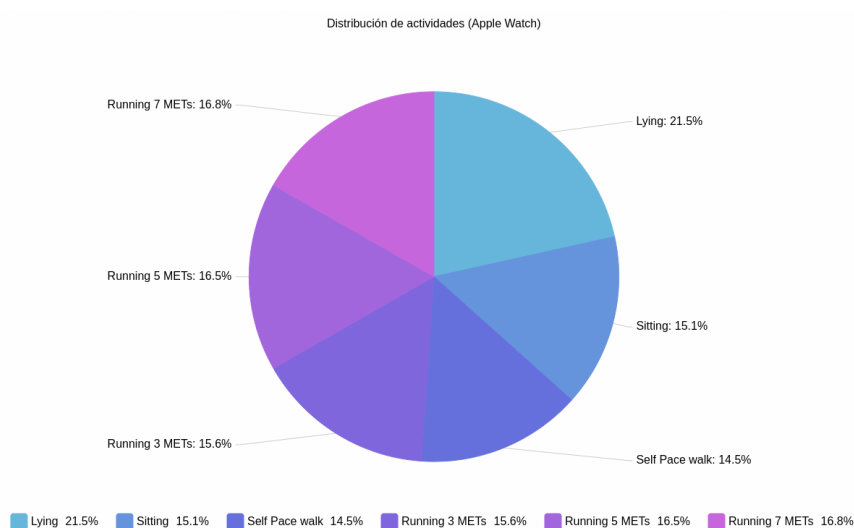


Figura 2: Distribución de las actividades (Apple Watch).

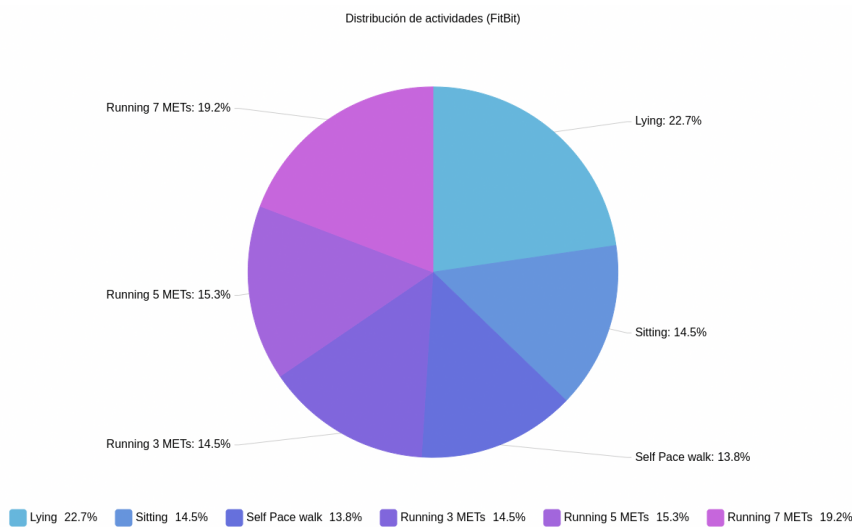


Figura 3: Distribución de las actividades (FitBit).

Variable	Description
X1	Identificador del registro de datos
Edad	Edad del individuo en años
Sexo	Sexo del individuo
Altura (cm)	Altura del individuo
Peso (kg)	Peso del individuo
Pasos	Número de pasos por minuto
Frecuencia Cardíaca (bpm)	Frecuencia cardíaca promedio en 1m
Calorías	Cantidad de energía gastada
Distancia (m)	Distancia recorrida
Entropía de la Frecuencia	Medida de la variabilidad en la frecuencia Cardíaca
Entropía de los Pasos	Medida de la variabilidad de los pasos
Frecuencia Cardíaca en Reposo (bpm)	Percentil 10 de las frecuencias medidas
Correlación Frecuencia Cardíaca y Pasos	Coefficiente de correlación entre frecuencia cardíaca y pasos
Frecuencia Cardíaca Normalizada	Frecuencia cardíaca normalizada
Intensidad (Karvonen)	Intensidad de la actividad calculada usando la fórmula de Karvonen
SD Frecuencia Cardíaca Normalizada	Desviación estándar de la frecuencia cardíaca normalizada
Pasos por Distancia	Producto del número total de pasos y la distancia recorrida
Dispositivo	Tipo de dispositivo utilizado para recopilar datos
Actividad	Tipo de actividad realizada

Tabla 1: Descripción de las características.

3. Importancia de características

La importancia de las características es un concepto que se utiliza para determinar cuáles son las características más relevantes para un modelo de aprendizaje automático. Las características con mayor importancia son las que más influyen en el modelo, por lo que se pueden utilizar para reducir el número de características y simplificar el modelo. Mientras que las características con menor importancia son las que menos influyen en el modelo, por lo que se podrían eliminar sin que el modelo pierda precisión.

Para determinar la importancia de las características de cada conjunto de datos se ha utilizado el algoritmo de *Random Forest* con el objetivo de obtener un ranking de las características más importantes. El algoritmo de *Random Forest* consiste en un conjunto de árboles de decisión, donde cada árbol se entrena con un subconjunto de datos y características, y la predicción final se obtiene mediante la media de las predicciones de cada árbol.

En DeepInt se ha seleccionado los conjunto de datos procesados anteriormente, y se ha seleccionado como variable objetivo la actividad realizada, además se ha establecido un valor de semilla igual a 42 para asegurar la reproducibilidad de los resultados. El tamaño del conjunto de test se ha establecido en un 20 % del total de los datos, empleando también un valor de semilla igual a 42.

Después de entrenar los modelos de *Random Forest* se puede ver el ranking de las características más importantes para cada conjunto de datos. En la Tabla 2 se representa este ranking de las 10 características más importantes para cada dispositivo. Como se puede observar las características más importantes en ambos casos son las mismas, pero en distinto orden. En el caso del Apple Watch, la característica más importante es la frecuencia cardíaca, seguida de la frecuencia cardíaca normalizada y la intensidad de la actividad. En el caso del FitBit, la característica más importante es el número de calorías quemadas, seguida del número de pasos y la frecuencia cardíaca normalizada.

Para comprobar si la mejora de los modelos mejora al utilizar solo las 10 clases más significantes, se crearán dos nuevas fuentes de datos en DeepInt, cada una derivada de su conjunto correspondiente, manteniendo las 10 características más importantes.

#	Apple Watch	FitBit
1	Frecuencia Cardíaca	Calorías
2	Frecuencia Cardíaca Normalizada	Pasos
3	Intensidad (Karvonen)	Frecuencia Cardíaca Normalizada
4	SD Frecuencia Cardíaca Normalizada	Pasos por Distancia
5	Calorías	Distancia
6	Distancia	Frecuencia Cardíaca
7	Pasos	Intensidad (Karvonen)
8	Correlación Frecuencia Cardíaca y Pasos	SD Frecuencia Cardíaca Normalizada
9	Pasos por Distancia	Entropía de los Pasos
10	Entropía de los Pasos	Correlación Frecuencia Cardíaca y Pasos

Tabla 2: Top 10 características para el Apple Watch y FitBit.

4. Entrenamiento de modelos

Obtenidas las características del conjunto de datos y sabiendo cuáles son las más importantes, se procede a entrenar distintos modelos de aprendizaje automático para determinar cuál es el algoritmo más adecuado para este caso de estudio. Al estar trabajando con un problema de aprendizaje supervisado, concretamente de clasificación, ya que trabajamos con conjuntos de datos con una característica objetivo, nos centraremos exclusivamente en los algoritmos aplicables a este tipo de problema.

Para comparar los distintos algoritmos se utilizará la métrica de exactitud (*accuracy*), que se define como el número de predicciones correctas sobre el número total de predicciones. Esta métrica es adecuada para este caso de estudio, ya que las clases están razonablemente equilibradas, por lo que no hay una clase que predomine excesivamente sobre las demás.

Aprovechando que se entrenó un modelo de *Random Forest* para determinar la importancia de las características, se extraerán los valores de exactitud de este modelo para compararlos con los de los nuevos modelos. La exactitud en este caso es del 82.24 % sobre el conjunto de datos del Apple Watch y del 91.57 % sobre el conjunto de datos del FitBit. A continuación se entrenará de nuevo modelos *Random Forest*, pero esta vez sobre los sets filtrados con solo las 10 características más relevantes, manteniendo los mismos hiperparámetros y semillas. La exactitud obtenida en ambos casos es inferior, por lo que se desechan los resultados.

Posteriormente se sigue la misma metodología para el algoritmo de **K-vecinos más cercanos (k-NN)**. Este algoritmo consiste en asignar una clase a un nuevo dato basándose en la clase de los k datos más cercanos. Se ha obtenido una exactitud del 68.99 % para el Apple Watch, y del 69.73 % para el FitBit. Esta vez se ha obtenido una mejora substancial al sustituir el conjunto de datos por el filtrado con las 10 características más relevantes, pero solo en el caso del FitBit, donde la exactitud sube al 82.18 %.

El siguiente algoritmo a probar es el de **máquinas de soporte vectorial (SVM)**. Este algoritmo consiste en asignar una clase a un nuevo dato basándose en la clase de los vectores de soporte, que son los datos más cercanos al hiperplano que separa las clases. Se ha obtenido una exactitud del 50.27 % para el Apple Watch, y del 50.00 % para el FitBit. Se obtiene una mejora substancial al sustituir el conjunto de datos por el filtrado con las 10 características más relevantes en el caso del FitBit, donde la exactitud sube al 55.94 %, en el caso del Apple Watch solo hay una mejora marginal, subiendo al 50.82 %.

El siguiente algoritmo es el de **regresión logística**. Este algoritmo consiste en asignar una clase a un nuevo dato basándose en la probabilidad de que pertenezca a cada clase. Se ha obtenido una exactitud del 33.47 % para el Apple Watch, y del 45.98 % para el FitBit. Se vuelve a obtener una mejora substancial al sustituir el conjunto de datos por el filtrado con las 10 características más relevantes en el caso del FitBit, donde la exactitud sube al 50.00 %, en el caso del Apple Watch solo hay una mejora marginal, subiendo al 34.84 %.

Otro algoritmo a probar es el de *Gradient Boosting*. Este algoritmo consiste en asignar una clase a un nuevo dato basándose en la suma de los resultados de varios modelos de predicción, como árboles de decisión. Se ha obtenido una exactitud del 69.40 % para el Apple Watch, y del 89.46 % para el FitBit. Se obtiene una mejora al sustituir el conjunto de datos por el filtrado con las 10 características más relevantes, en el caso del FitBit la exactitud sube al 92.15 %, superando al mejor caso anterior, en el caso del Apple Watch hay una mejora menor, subiendo al 70.49 %.

En la Tabla 3 se puede ver un resumen de los resultados obtenidos para cada algoritmo y conjunto de datos. Como se puede observar, el algoritmo que mejor resultados ha obtenido ha sido el de *Gradient Boosting* con un 92.15 % de exactitud sobre el conjunto de datos del FitBit, seguido del *Random Forest* con un 91.57 % de exactitud sobre el mismo conjunto de datos. En el caso del Apple Watch, el algoritmo que mejor resultados ha obtenido ha sido el de *Random Forest* con un 82.24 % de exactitud, seguido del *Gradient Boosting* con un 70.49 % de exactitud.

Algoritmo	Apple Watch	FitBit
Random Forest	82.24 %	<i>91.57 %</i>
Random Forest (@10)	81.97 %	90.04 %
k-NN	68.99 %	69.73 %
k-NN (@10)	68.31 %	82.18 %
SVM	50.27 %	50.00 %
SVM (@10)	50.82 %	55.94 %
Regresión Logística	33.47 %	45.98 %
Regresión Logística (@10)	34.84 %	50.00 %
Gradient Boosting	69.40 %	89.46 %
Gradient Boosting (@10)	<i>70.49 %</i>	92.15 %

Tabla 3: Exactitud de los distintos algoritmos.

5. Evaluación de modelos

Una vez se han entrenado los modelos y se han obtenido los resultados, se procede a evaluarlos para determinar el rendimiento de ellos para nuestro caso de estudio. Para ello se utilizará la matriz de confusión, que es una matriz que muestra el número de predicciones correctas e incorrectas para cada clase. En la Tabla 4 se puede ver la matriz de confusión para el modelo de *Gradient Boosting* entrenado con el conjunto de datos del FitBit filtrado. Como se puede observar, la mayoría de las predicciones son correctas, aunque en algún caso se confunde entre actividades de diversa intensidad.

Observado/ Predicho	Lying	Sitting	Self Pace walk	Running 3 METs	Running 5 METs	Running 7 METs	Σ
Lying	100	1	1	3	1	0	111
Sitting	2	65	1	0	3	3	74
Self Pace walk	3	0	60	3	1	0	67
Running 3 METs	4	1	0	79	0	1	85
Running 5 METs	1	3	0	0	73	0	77
Running 7 METs	0	0	2	1	1	104	108
Σ	110	70	64	86	79	108	522

Tabla 4: Matriz de confusión para el modelo de Gradient Boosting (FitBit filtrado).

En la Tabla 5 se puede ver la matriz de confusión para el modelo de *Random Forest* entrenado con el conjunto de datos del Apple Watch. Como se puede observar, la mayoría de las predicciones también son correctas, aunque se ve que hay un mayor número de errores que en el caso anterior.

Observado/ Predicho	Lying	Sitting	Self Pace walk	Running 3 METs	Running 5 METs	Running 7 METs	Σ
Lying	127	8	8	13	1	2	159
Sitting	9	70	3	0	8	5	95
Self Pace walk	3	10	93	5	3	0	114
Running 3 METs	15	1	2	98	4	1	121
Running 5 METs	3	7	0	1	112	5	128
Running 7 METs	4	5	0	1	3	102	115
Σ	161	101	106	118	131	115	732

Tabla 5: Matriz de confusión para el modelo de Random Forest (Apple Watch).

Las métricas de precisión, exhaustividad y F1 se pueden ver en la Tabla 6 y la Tabla 7. Como se puede observar, las métricas son similares en ambos casos, aunque en el caso del FitBit se obtienen mejores resultados en general. Ambos modelos

serían efectivos para predecir la actividad realizada, aunque el modelo del FitBit sería más efectivo.

Métrica	Valor
Exactitud	0.92
Cohen kappa	0.91
Matthews Phi	0.91
Precisión (ponderada)	0.92
Exhaustividad (ponderada)	0.92
F1 (ponderada)	0.92

Tabla 6: Métricas de Gradient Boosting (FitBit)

Métrica	Valor
Exactitud	0.82
Cohen kappa	0.79
Matthews Phi	0.79
Precisión (ponderada)	0.82
Exhaustividad (ponderada)	0.82
F1 (ponderada)	0.82

Tabla 7: Métricas de Random Forest (Apple Watch)

6. Conclusión

Se puede considerar exitoso el intento de predecir la actividad realizada por una persona a partir de los datos de un *smartwatch*, ya que se han obtenido modelos con una exactitud superior al 90 % en el caso del FitBit, y superior al 80 % en el caso del Apple Watch. Estos modelos se pueden utilizar para predecir la actividad realizada por una persona, y se pueden utilizar para detectar anomalías en la intensidad de éstas, como por ejemplo para detectar si una persona está realizando una actividad de mayor intensidad de la que debería.

Referencias

- [1] Daniel Fuller. Replication Data for: Using machine learning methods to predict physical activity types with Apple Watch and Fitbit data using indirect calorimetry as the criterion. Versión V1, 2020. DOI: 10.7910/DVN/ZS2Z2J.