

Daily Movements Identifier in Real Time

Abstract

Este proyecto desarrolló un sistema de identificación de movimientos humanos utilizando un modelo de clasificación XGBoost. Mediante refinamientos iterativos, se ajustaron características y se hizo recolección de datos para mejorar la precisión al diferenciar movimientos. El modelo logró identificar acciones como sentarse, pararse, caminar hacia atrás y adelante para poder ser utilizado en otros proyectos futuros.

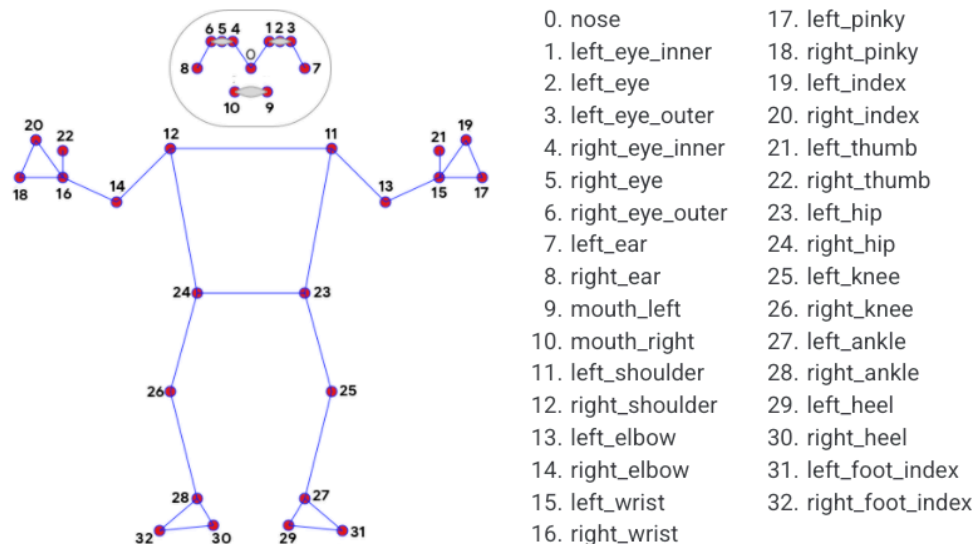
Introducción

Este proyecto presenta el desarrollo de un sistema de identificación de movimientos humanos en tiempo real utilizando técnicas de aprendizaje automático y machine learning, con un enfoque en actividades cotidianas como caminar, girar, sentarse y pararse. A partir de un conjunto de datos de videos etiquetados, el proyecto emplea XGBoost, un algoritmo de boosting de gradiente, como modelo principal para la clasificación de actividades. El sistema demuestra alta precisión en la identificación de los diversos movimientos, destacando su potencial en aplicaciones como monitoreo de salud, seguimiento de actividad física y asistencia a personas.

Theory

Machine learning: El machine learning se refiere a la disciplina encargada de desarrollar algoritmos y modelos estadísticos que permiten a los sistemas computacionales realizar tareas sin recibir instrucciones explícitas. En su lugar, estos algoritmos identifican patrones en grandes volúmenes de datos históricos y generan resultados basados en inferencias. Este enfoque facilita la mejora de la precisión en tareas específicas a partir de datos de entrada. Por ejemplo, en el ámbito médico, se pueden entrenar aplicaciones para diagnosticar enfermedades como el cáncer utilizando imágenes de rayos X previamente etiquetadas con diagnósticos.

Media Pipe: Para la identificación del cuerpo de las personas hacemos uso de Media Pipe Pose que nos permite realizar la silueta de una persona mediante un video o incluso con una cámara en tiempo real mientras se procesa el video.



Mr. Pose: An Easy Guide for Pose Estimation with Google's MediaPipe

Hacemos uso de los puntos 0, 11-16 y 23-28 que fueron aquellos más importantes para la detección de los movimientos.

XG Boost: XGBoost (Extreme Gradient Boosting) es un algoritmo de aprendizaje automático basado en árboles de decisión que optimiza el proceso de boosting de gradiente, haciendo énfasis en la eficiencia, flexibilidad y escalabilidad. Se utiliza ampliamente en tareas de clasificación y regresión y se ha convertido en una herramienta ampliamente utilizada debido a su capacidad para manejar grandes conjuntos de datos y complejas tareas de predicción con alta precisión.

Methodology:

El desarrollo de este proyecto se llevó a cabo basándonos en la metodología crisp-dm se realizaron 3 iteraciones antes de obtener el producto final.

Para la primera entrega del proyecto se evaluaron diversas herramientas y tecnologías disponibles para implementar el sistema al igual que el propósito del proyecto.

Finalmente, decidimos hacer uso Media Pipe Pose para la extracción de landmarks de movimientos humanos y el modelo XGBoost como clasificador principal debido a que mediante las diferentes pruebas que realizamos con el modelo durante el desarrollo se obtuvieron mejores métricas y mejores resultados con este modelo a diferencia de Random Forest y SVC.

Para la recolección de datos se grabaron videos de todos los movimientos cotidianos con la colaboración de los integrantes del grupo de trabajo y voluntarios que quisieron participar en esta recolección de datos. Al principio grabamos unos datos estáticos que a pesar que funcionaban para el modelo a la hora de probarlo se le hacía complicado al cambiar las

distancias de las personas, por lo cual hicimos una segunda toma de datos en la cual realizamos los movimientos pero manejando diferentes distancias de la cámara que nos logró dar mejores resultados.

Los videos etiquetados con los movimientos grabados se procesaron utilizando los datos de etiquetado para extraer los landmarks proporcionados por Media Pipe Pose, al inicio hacíamos uso de los 33 landmarks pero al final nos dieron mejores resultados usando ciertos puntos específicos por lo cual en la exploración de los datos decidimos quitarlos y solo quedarnos con los landmarks 0, 11-16 y 23-28. De igual manera al en la segunda iteración decidimos realizar el cálculo de deltas del landmark de los hombros de la persona de tal manera que agrupamos 5 frames de video y hacíamos el cálculo para identificar el movimiento lo cual no nos dio un buen resultado, por lo cual para esta última entrega decidimos solo hacer el agrupamiento por 10 frames y hacer el cálculo de deltas de la mayoría de landmarks que nos terminó dando mejores resultados.

Como último paso de las iteraciones realizamos un Grid Search para identificar los mejores hiper parámetros para el modelo XGBoost, maximizando su rendimiento y se evaluó con métricas como accuracy, recall, precisión y f1-score.

Finalmente, se desarrolló y desplegó una aplicación que integra el modelo entrenado que se exporta a un archivo .pkl, permitiendo la identificación en tiempo real de movimientos cotidianos.

Results:

- **Random Forest Model:**

```
Exactitud del modelo: 0.9789770264412657
```

Matriz de Confusión:

[653	12	22	0	2]
[2	1747	0	5	4]
[23	1	643	3	1]
[2	6	2	764	3]
[0	3	2	4	710]]

Reporte de Clasificación:

	precision	recall	f1-score	support
0	0.96	0.95	0.95	689
1	0.99	0.99	0.99	1758
2	0.96	0.96	0.96	671
3	0.98	0.98	0.98	777
4	0.99	0.99	0.99	719
accuracy			0.98	4614
macro avg	0.98	0.97	0.98	4614
weighted avg	0.98	0.98	0.98	4614

- **SVM Model:**

```
Exactitud del modelo: 0.9605548331166016

Matriz de Confusión:
[[ 636  20  30  1  2]
 [  4 1737  4  6  7]
 [ 32  18 619  2  0]
 [  1  17  6 751  2]
 [  3  23  2  2 689]]

Reporte de Clasificación:

```

	precision	recall	f1-score	support
0	0.94	0.92	0.93	689
1	0.96	0.99	0.97	1758
2	0.94	0.92	0.93	671
3	0.99	0.97	0.98	777
4	0.98	0.96	0.97	719
accuracy			0.96	4614
macro avg	0.96	0.95	0.96	4614
weighted avg	0.96	0.96	0.96	4614

- **XG Boost:**

```
Exactitud del modelo: 0.9755093194625054

Matriz de Confusión:
[[ 650  18  19  1  1]
 [  4 1744  0  4  6]
 [ 21  5 642  3  0]
 [  5 13  2 755  2]
 [  0  4  2  3 710]]

Reporte de Clasificación:

```

	precision	recall	f1-score	support
0	0.96	0.94	0.95	689
1	0.98	0.99	0.98	1758
2	0.97	0.96	0.96	671
3	0.99	0.97	0.98	777
4	0.99	0.99	0.99	719
accuracy			0.98	4614
macro avg	0.97	0.97	0.97	4614
weighted avg	0.98	0.98	0.98	4614

Análisis de Resultados

La primera iteración del modelo no lograba generalizar, dado que asumía que todo movimiento era un giro, sin importar la distancia ni la velocidad, por lo tanto nuestro primer acercamiento fue retirar puntos innecesarios, es decir, los puntos de la cara fueron retirados para evitar que el modelo se confunda con todos ellos.

Nuestro siguiente acercamiento fue añadir la visibilidad a los puntos del modelo, con el objetivo de que pueda identificar el giro como un cambio de visibilidad, y después de prueba

y error decidimos solo dejar la visibilidad de la nariz, hombros, y cadera, este cambio finalmente llevó a que el modelo detectara el acercarse a la cámara.

Luego, viendo que remover puntos no utilizados fue útil, y que aparte significaba que el modelo era más ligero de correr, decidimos retirar los puntos correspondientes a las manos y los pies, dejando solamente las muñecas y tobillos, esto no generó mayor cambio en la precisión del modelo, pero en este punto nos dimos cuenta que el modelo tenía dificultad diferenciando entre caminar hacia el frente y hacia atrás, y entre pararse y sentarse.

Finalmente, con el propósito de solucionar la dificultad de diferenciación entre diferentes opuestos, decidimos añadir a los conjuntos de frames el promedio de la diferencia en diferentes ejes, es decir, el promedio del cambio en el eje z para el punto derecho del hombro, esto nos permitió diferenciar los opuestos y finalmente detectar pararse, aunque sentarse seguía siendo muy inconsistente.

Por último, decidimos doblar la cantidad de frames que componen cada conjunto, con el propósito de diferenciar entre movimientos con mayor precisión, esto finalmente permitió que se pudiera detectar la acción de sentarse, y permitió una separación más consistente entre las otras acciones.

Un estudio realizado por el Dr. Celal Alagöz que se presentó en el 1st Bilsel International Sur Scientific Researches Congress nos muestra una investigación de identificación de actividades humanas haciendo la utilización del modelo XG Boost con datos de sensores de movimientos y sus resultados son muy cercanos a los nuestros.

Model	Accuracy	F1	AUC
XGBoost	0.990	0.990	0.999
Minirocket	0.988	0.989	0.993
SGD	0.446	0.427	0.664
Naive Bayes	0.736	0.747	0.734
DT	0.748	0.746	0.850
kNN	0.707	0.706	0.895
RF	0.818	0.818	0.966
NN	0.856	0.857	0.974
SVM	0.878	0.872	0.988
LSTM	0.900	-	-
CNN	0.975	-	-

Tomada de: *Comparative analysis of XGBoost and MiniRocket algorithms for human activity recognition using smartphone sensor data.*

Como observamos para XG Boost como para SVM ambos resultados son muy similares en la identificación de movimientos donde incluso presentamos un mejor accuracy para el modelo de SVM, lo que nos permite decir que nuestros resultados no estan muy alejados de otros modelos en la actualidad.

Conclusión:

Para concluir, implementamos un sistema de identificación de movimientos humanos que inicialmente contaba con dificultades para generalizar, clasificando incorrectamente todos los movimientos como giros. Este proceso nos enseñó que la simplificación y selección de características específicas son esenciales para la robustez y precisión del modelo. Aun así, se podría mejorar explorando técnicas avanzadas y alimentando el modelo con una mayor cantidad de datos diversos con respecto a los movimientos y tamaño de las personas generando datasets más diversos. Finalmente, es posible que pueda haber un aumento de precisión simplemente siendo más cuidadosos con el proceso de etiquetado, todo esto ayudaría a mejorar la generalización y aplicabilidad del sistema.

Bibliografía

Siva, L. (2022, 31 octubre). Mr. Pose: An Easy Guide for Pose Estimation with Google's

MediaPipe. Medium.

<https://logessiva.medium.com/an-easy-guide-for-pose-estimation-with-googles-media-pipe-a7962de0e944>

¿Qué es el machine learning? - Explicación sobre el machine learning empresarial - AWS. (s.

f.). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/machine-learning/>