



LICENCIATURA EN CIENCIAS DE DATOS
Propuesta de tema de tesis

Desarrollo e implementación de un modelo basado en self-supervised learning para la predicción de episodios de sueño a partir de datos de uso de smartphones con sistema operativo Android.

En el presente proyecto se propone diseñar, desarrollar y evaluar un modelo un modelo basado en self-supervised learning para la predicción de episodios de sueño a partir de datos de uso de smartphones con sistema operativo Android. La idea principal del proyecto es la de utilizar una filosofía de modelado similar a la utilizada por [Yuan et al., 2023] pero con la diferencia que en nuestro caso los datos para la etapa de self-supervised learning serían las interacciones con el smartphone y para la etapa de supervised learning nuestra verdad (con la cual taggearíamos los datos) será el status de despierto o dormido obtenido a partir de los datos de actigrafía (un relojito que mide la actividad similar a un Fitbit). El proyecto ya cuenta con datos taggeados de ~50 sujetos por ~20 noches utilizando actiimetrías y con registros de actividad de Android de ~100 sujetos de ~600 noches cada uno. Vale la pena mencionar que estamos en proceso de toma de datos, y en medio de un relanzamiento del reclutamiento de participantes. [Yuan et al., 2023] (<https://www.medrxiv.org/content/10.1101/2023.07.07.23292251v1.full.pdf>)

Palabras clave: Sueño - Modelado del comportamiento humano - Inteligencia Artificial - Smartphones

Conocimientos deseables

Se espera que el candidato tenga experiencia con modelado de datos y preferentemente redes neuronales (pytorch).

¿Qué podría aprender quien realice esta tesis?

El estudiante deberá enfrentarse a datos reales de los cuales deberá extraer la información relevante para el desarrollo del modelo. También deberá familiarizarse con la implementación de modelos de DNN, en particular en modelos que tienen una etapa de self-supervised learning. Finalmente, el estudiante deberá evaluar la validez del modelo propuesto para su uso en el campo de la epidemiología del sueño.

Dirección de la tesis

Spiousas, Ignacio
Escuela de Educación, Universidad de San Andrés

Contacto: ispiousas@udesa.edu.ar

Más información en el pdf a continuación.

Desarrollo e implementación de un modelo basado en self-supervised learning para la predicción de episodios de sueño a partir de datos de uso de smartphones con sistema operativo Android.

Ignacio Spiouzas, Laboratorio Interdisciplinario del Tiempo y la Experiencia, Universidad de San Andrés, CONICET.

Departamento de Computación, Instituto de Ciencias de la Computación, Universidad de Buenos Aires, CONICET.

Resumen

En el presente proyecto se propone diseñar, desarrollar y evaluar un modelo un modelo basado en self-supervised learning para la predicción de episodios de sueño a partir de datos de uso de smartphones con sistema operativo Android. La idea principal del proyecto es la de utilizar una filosofía de modelado similar a la utilizada por [\[Yuan et al., 2023\]](#) pero con la diferencia que en nuestro caso los datos para la etapa de self-supervised learning serían las interacciones con el smartphone y para la etapa de supervised learning nuestra verdad (con la cual taggearemos los datos) será el status de despierto o dormido obtenido a partir de los datos de actigrafía (un relojito que mide la actividad similar a un Fitbit). El proyecto ya cuenta con datos taggeados de ~50 sujetos por ~20 noches utilizando actiimetrías y con registros de actividad de Android de ~100 sujetos de ~600 noches cada uno. Vale la pena mencionar que estamos en proceso de toma de datos, y en medio de un relanzamiento del reclutamiento de participantes.

Palabras clave

Sueño - Modelado del comportamiento humano - Inteligencia Artificial - Smartphones

Conocimientos deseables para desarrollar el plan propuesto

Se espera que el candidato tenga experiencia con modelado de datos y preferentemente redes neuronales (pytorch).

El estudiante deberá enfrentarse a datos reales de los cuales deberá extraer la información relevante para el desarrollo del modelo. También deberá familiarizarse con la implementación de modelos de DNN, en particular en modelos que tienen una etapa de self-supervised learning. Finalmente, el estudiante deberá evaluar la validez del modelo propuesto para su uso en el campo de la epidemiología del sueño.

Descripción

El déficit de sueño, tanto en cuanto a cantidad como a calidad, tiene gran relevancia para la salud pública en términos de mortalidad, morbilidad, accidentología, percepción de la propia salud y calidad de vida, así como en la salud ocupacional y bienestar general. Se calcula que alrededor de un tercio de los estadounidenses no obtienen suficientes horas de sueño y se estima que en América Latina una gran parte de la población sufre de una deficiencia crónica de horas de sueño.

Las herramientas utilizadas en la clínica y en investigación para el monitoreo de los patrones de sueño incluyen la polisomnografía, el uso de monitores de sueño no invasivos portátiles o fijos, diarios de sueño, y distintos tipos de cuestionarios de autorreporte. Cada una de estas

herramientas posee sus ventajas y limitaciones en un amplio abanico, en cuanto a la precisión y alcance de cada una.

Uno de los desafíos actuales de la ciencia del sueño es el desarrollo de herramientas prácticas, escalables y de bajo costo para el monitoreo de los hábitos de sueño tanto en la escala individual como en la poblacional. Una alternativa reciente es el uso de datos de aplicaciones móviles y redes sociales para el estudio de los mismos. Estas fuentes de información permiten a los investigadores acceder con facilidad a una gran cantidad de datos obtenidos en condiciones naturales, y trabajos recientes indican que, debido a su masividad, los datos obtenidos a través de estas vías permiten el estudio de fenómenos del sueño a gran escala con aceptable confiabilidad.

Por un lado, los datos y modelos que pueden derivarse de los análisis de cuerpos de big data pueden facilitar la mitigación de la prevalencia y las consecuencias de los distintos problemas derivados de la falta de sueño. Por otro lado, y en una escala mayor, los repositorios de grandes conjuntos de datos, así como los resultados de censos epidemiológicos, pueden ser aprovechados tanto para informar la toma de decisiones en marco de políticas públicas así como para el diseño y evaluación de intervenciones para mejorar los hábitos de sueño en poblaciones específicas.

En el presente proyecto se propone diseñar, evaluar y validar un modelo basado en self-supervised learning para la predicción de episodios de sueño a partir de datos de uso de smartphones con sistema operativo Android.

El proyecto ya se encuentra en desarrollo con datos taggeados de ~50 sujetos por ~20 noches utilizando actiimetrías (un relojito que mide la actividad similar a un Fitbit) y con registros de actividad de Android de ~100 sujetos de ~600 noches cada uno. Vale la pena mencionar que estamos en proceso de toma de datos, y en medio de un relanzamiento del reclutamiento de participantes.

La idea principal del proyecto es la de utilizar una filosofía de modelado similar a la utilizada por [\[Yuan et al., 2023\]](#) pero con la diferencia que en nuestro caso los datos para la etapa de self-supervised learning serían las interacciones con el smartphone y para la etapa de supervised learning nuestra verdad (con la cual taggearemos los datos) será el status de despierto o dormido obtenido a partir de los datos de actigrafía.

La presente propuesta se enmarca en un proyecto titulado "Diseño, desarrollo y validación de métricas basadas en datos masivos para el estudio de los hábitos de sueño en Argentina" dirigido por Ignacio Spiouzas y Leandro Casiraghi en el Laboratorio del Tiempo y la Experiencia ([LITERA](#)) de la Universidad de San Andrés y cuenta con financiamiento de la Agencia de Promoción Científica. El laboratorio cuenta con todo lo necesario para que el estudiante lleve adelante su tarea de investigación.