

Predicción de la progresión de enfermedades neurodegenerativas

Daniel Polanco Pérez (@Polandani)
María Fernanda Suárez González (@MaferGlez)
Adrián Navarro Foya (@adrianfoya)

Universidad de La Habana, La Habana, Cuba

October 17, 2024

<https://github.com/MaferGlez03/IA-Project.git>

Resumen

Este trabajo presenta una simulación computacional de un hospital de enfermedades neurodegenerativas, basada en agentes inteligentes, con el objetivo de evaluar la eficacia de diferentes tipos de tratamientos y medicamentos para la positiva evolución de los pacientes. La simulación contempla dos tipos de agentes fundamentales: doctores y pacientes. Los doctores tienen una serie de equipos a su disposición para la recuperación de los pacientes, así como medicamentos y tratamientos varios. Para escoger cuales son las mejores opciones para el paciente se aplica un algoritmo de búsqueda.

Los pacientes acudirán a las consultas buscando mejorar su estado. Serán diagnosticados por un sistema que combina el uso de un LLM y un algoritmo que aplica técnicas de conocimiento. El primero se encarga de realizar la entrevista con el paciente y sonsacar los síntomas perceptibles. El segundo, a partir de los síntomas detectados y un previo entrenamiento con una base de datos de pacientes, predice cuales son las posibles enfermedades que posea el paciente.

Introducción

El futuro se acerca con avances tecnológicos cada vez más veloces. Las ciencias médicas no se quedan atrás y en el último siglo las enfermedades más mortíferas han sido erradicadas. En cambio han surgido nuevas y más complejas enfermedades a las que debemos enfrentar.

No es sorpresa para nadie saber que las enfermedades del futuro van a ser las enfermedades neurodegenerativas. Con tanto progreso en la nanobiotecnología, el resto de órganos van a ser infalibles en cuestión de algunas décadas pero el cerebro seguirá siendo un problema.

Este trabajo se centra en la detección y la búsqueda de mecanismos para detener el progreso de las enfermedades neurodegenerativas, siendo el eje central la interacción entre dos tipos de agentes: doctores y pacientes; y el medio: el hospital. Los doctores con el deber de detectar las enfermedades y conducir al paciente por una serie de tratamientos para invertir el avance de la enfermedad. Los pacientes con su principal objetivo de mejorar su estado de salud, también juegan con la incertidumbre que genera ciertos tratamientos a largo plazo, con efectos secundarios agresivos y la posibilidad de abandonar sus tratamientos.

El hospital, como medio en el que se desarrolla la simulación, posee los recursos necesarios para que los agentes puedan cumplir sus objetivos, proporcionar información acerca del estado de cada uno de ellos e interactuar entre ellos. Los agentes serán influenciados por el medio impidiendo su normal funcionamiento y afectando los deseos y creencias de los mismos.

El principal objetivo del presente trabajo es detallar el avance de distintas enfermedades neurodegenerativas, la influencia que puede ejercer sobre ellas distintos tratamientos o medicamentos y la importancia de las creencias y decisiones de los portadores para el progreso de la misma en el paciente. También se desea analizar la precisión del sistema de predicción de enfermedades a partir de ciertos síntomas e incluso el mejoramiento del mismo.

Desarrollo

Método

Para los estudios realizados en este trabajo, se tiene como principal herramienta y método de generación y obtención de datos, la simulación computacional conformada para este trabajo en particular en el lenguaje de programación Python. Por esta razón, es necesario conocer los detalles de su implementación.

Simulación

Esta, es una simulación basada en agentes inteligentes que interactúan entre sí con el medio, cuyas implementaciones serán abordadas en las secciones posteriores. Como el objetivo es analizar la influencia de distintos tratamientos y medicamentos en las diferentes enfermedades, es necesario manejar la variable temporal para poder comprobar el progreso, por tanto, la simulación recreada es dinámica.

Medio

Nuestro medio principal es el hospital, para ello se hace una representación del mismo mediante una clase. En ella tenemos propiedades y métodos que recrean el funcionamiento de un hospital convencional. Es loable mencionar las clases ***Procedure y Utility*** que permiten obtener detalles de los servicios que presta el centro, así como de los recursos que dispone. De ahí

es posible extraer información como los nombres, funcionalidades o disponibilidad de estos servicios.

Agentes

En la simulación intervienen dos clases de agentes distintas, como ya hemos mencionado, los doctores y los pacientes. Todos los agentes poseen una arquitectura **Beliefs Desires Intentions (BDI)** que les permite tomar decisiones a partir de sus propios intereses y las circunstancias externas.

Los doctores buscan cumplir con su deber de mejorar el estado de salud de los pacientes, para ello se enfocan en disminuir tanto la gravedad de la enfermedad como la presencia de síntomas, para ello se apoyan en el uso de tratamientos y medicamentos que ofrece el hospital como servicios. A su vez, los galenos tienen un grupo de restricciones como lo son la vida útil de los recursos, la disponibilidad de los servicios o la seguridad del mismo.

Los pacientes en cambio se enfocan en soportar todos los procesos negativos asociados al tratamiento de enfermedades neurodegenerativas cumpliendo con la máxima de mejorar su estado de salud.

Inteligencia Artificial (IA)

El uso de algoritmos de inteligencia artificial en este trabajo resultó crucial para el cumplimiento de los objetivos e estudio. A continuación se expondrán los algoritmos utilizados y con qué fin.

Algoritmo de conocimiento

El código presentado implementa un modelo de predicción de enfermedades a partir de síntomas utilizando un algoritmo **RandomForestClassifier**. En primer lugar, se realiza la lectura de los datos desde un archivo CSV, y luego se emplea la ontología cargada para crear un mapa de síntomas con el fin de identificar de manera precisa cada uno de ellos en el conjunto de datos. Además, el código utiliza la librería MultiLabelBinarizer para transformar las listas de síntomas en vectores binarios, de tal manera que se puedan procesar correctamente en los modelos de predicción.

Posteriormente, el código divide los datos en conjuntos de entrenamiento y prueba para entrenar un modelo de RandomForest por cada enfermedad, prediciendo de manera individual la probabilidad de que el paciente sufra una de las enfermedades neurodegenerativas en el conjunto considerado. Al final, se aplican las predicciones para calcular la precisión de los modelos en el conjunto de prueba y, posteriormente, se construye un método de predicción que permite introducir nuevos síntomas y obtener un diagnóstico basado en los modelos entrenados, devolviendo las probabilidades de sufrir cada una de las enfermedades en un formato ordenado por probabilidad.

Algoritmo de búsqueda

El algoritmo A^* es una técnica de búsqueda informada que combina la exploración de los posibles caminos con un enfoque guiado por una función de costo acumulado y una heurística. En cada iteración, A^* selecciona el estado que minimiza la suma del costo acumulado desde el estado inicial y una estimación (**heurística**) del costo restante hasta el objetivo. La implementación del proyecto utiliza el algoritmo de búsqueda A^* para optimizar el proceso de

diagnóstico y tratamiento médico. El sistema tiene como objetivo ayudar a los médicos a determinar el mejor camino a seguir, aplicando pruebas y medicamentos que maximicen la mejora en el estado del paciente. Este se modela a través de la clase *State*, que incluye atributos como los medicamentos administrados, los procedimientos realizados, los síntomas presentes, el progreso en el tratamiento y el estado general del paciente. Clases adicionales como *Symptom*, *Medication*, *Procedure*, *Progress* y *StateGeneral* permiten representar estos componentes de manera detallada y modular.

En este proyecto, la función de costo incluye los precios de los procedimientos y medicamentos, así como penalizaciones por la presencia de síntomas severos o un progreso insuficiente. La heurística utilizada es admisible, lo que significa que nunca sobreestima el costo real desde cualquier estado hasta el objetivo. Esto garantiza que el algoritmo A^* encuentre el camino óptimo, es decir, el que minimiza el costo total.

El sistema genera sucesores de cada estado a través de la función *generate_successors*, que simula la aplicación de pruebas y tratamientos, actualizando los síntomas, el progreso y el estado general del paciente. Estos sucesores se evalúan en función del estado objetivo, implementado en la clase *GoalCheck*, que busca alcanzar al menos un 80% de progreso y eliminar todos los síntomas severos. Para garantizar que el progreso sea adecuado, se ajustan los efectos de algunos medicamentos clave que mejoren significativamente el progreso del paciente.

Large Language Model (LLM)

Cuando un paciente ingresa al centro neurológico lo primero que se le realiza es la consulta verbal, la cual es guiada por un LLM al cual se le especifica ahondar en la sintomatología del paciente para poder hacer un diagnóstico más profundo.

Posteriormente para un análisis médico más exhaustivo, se requiere de unas métricas que definan el estado general del paciente, las cuales son el nivel de energía y el nivel de dolor. Para ello es utilizado nuevamente el LLM, al cual se le muestra toda la conversación del paciente con el doctor y se le pide que escoja un valor entre 0 y 10 que defina los niveles para ambos parámetros siendo 0 el mínimo y 10 el máximo.

Finalmente un LLM es el encargado de realizar un resumen clínico del paso del paciente por el hospital, reseñando todos procesos por los que recorrió y cuál fue su estado inicial y su final, así como reportar numéricamente el progreso del paciente entre 0 y 100.

Estudios

Estableciendo criterio de parada por paciente

Una de las cuestiones más importantes para el correcto funcionamiento de la simulación es decidir hasta que punto un paciente sigue siendo atendido en el hospital. La elección es un balance entre el tiempo de estancia del mismo y el progreso de la enfermedad. Para ello se definieron distintas simulaciones con diferentes niveles de progreso de enfermedad mínimos para poder liberar una cama del centro y permitir la entrada de un nuevo paciente. Para nivel mínimo se realizaron 100 simulaciones y luego se hizo una prueba de **Mann-Whitney U** para analizar la cantidad de pacientes atendidos contra la probabilidad de empeorar por el nivel de progreso de enfermedad. Los resultados y discusión serán expuestos en la próxima sección.

Servicio más utilizado

Es importante conocer cuales son los servicio más usados por los doctores para poder conocer en el futuro que equipos y tratamientos son los que deben reponer mas pronto o aumentar su disponibilidad. Se realizaron 100 simulaciones aplicando cada criterio de parada seleccionado en al punto anterior. Se analizó el servicio mas usado de cada uno y una comprobación de la relación entre el criterio de parada y el servicio más utilizado.

Resumen del LLM vs progreso real

El uso del LLM en el proyecto se divide en dos secciones, primero para la entrevista con el paciente y segundo para entregar un resumen de todo el proceso medico. Al final de este proceso se le pide al LLM que analice el progreso del paciente según todo lo vivido en el hospital. Este progreso será contrastado con el cambio del progreso de la enfermedad desde la primera consulta hasta la última y se compararán resultados. Para esta comparación se realizarán 100 simulaciones y se tomará la varianza entre los progresos del LLM y los progresos reales. Se realizó, además un test de **correlación de Spearman** para analizar la relación existente entre la clasificación del LLM y la satisfacción real.

Conclusiones

La simulación realizada utilizando agentes inteligentes ha resultado de gran utilidad para evaluar la eficacia de las distintas estrategias de medicación y tratamientos. Los resultados nos muestran que un correcto funcionamiento del hospital combinado con un colectivo médico de alto nivel capaz de decidir que tratamientos y medicinas asignar en cada momento son claves para el positivo progreso de las enfermedades neurodegenerativas en los pacientes. Además se ha reflejado cual es la mejor estrategia para la gestión de un numero limitado de camas en hospitales de tan importante labor como este y como manejarlos para evitar un colapso y poder atender a la mayor cantidad de pacientes posibles sin descuidar la calidad de la atención.

Recomendaciones

Es sugerente mencionar que la introducción de distintos agentes de menor importancia podrían alterar sustancialmente el funcionamiento de la simulación dotando a la misma de un funcionamiento más cercano al de un hospital real. Entre estos posibles agentes podrían ser los acompañantes de los pacientes, los enfermeros, entre otros. Además sería interesante modelar cuanto puede cambiar la simulación al adicionar la variable económica.