

Predicción del comportamiento de una enfermedad simulada en autómatas celulares con un algoritmo propuesto en redes neuronales

May 25, 2021

1 Formulación del problema

Se han desarrollado gran cantidad de estudios que marcan un avance importante en el campo de la epidemiología, entre los que cabe destacar trabajos como el de sir Ronald Ross en 1911, con la incidencia y el control de la malaria, a partir de un modelo determinista, y el modelo de Kermack-McKendrick en 1927, también conocido como modelo SIR, por la relación que describe entre tres estados: susceptible-infectado-recuperado. [10]

Se han desarrollado múltiples variaciones del modelo SIR con el objetivo de analizar diferentes tipos de enfermedades, de las cuales cabe mencionar el modelo SIS, que describe la relación entre dos tipos de poblaciones: susceptibles e infectados, donde al igual que en el modelo SIR, se establecen las interacciones entre estados, a partir de un sistema de ecuaciones diferenciales acopladas.[9]

Los avances tecnológicos de las últimas décadas han sido de gran ayuda para la elaboración de modelos y simulaciones que permiten analizar características que no eran posibles con los modelos anteriormente mencionados. Además, con el reciente brote de coronavirus se identificó la necesidad de un modelo que tuviera en cuenta las relaciones sociales, lo que ocasionó que se destacarán aplicaciones de la epidemiología en redes neuronales y complejas, ambas con una fuerte influencia de la teoría de grafos. Se han realizado trabajos que analizan el impacto de la movilidad y las interacciones entre individuos usando redes neuronales, por ejemplo el descrito en [8], que toma como nodos a regiones de un país y las aristas las identifica como la movilidad humana de una región a otra, para posteriormente analizar sus implicaciones, y el desarrollado en [11] para describir la propagación del VIH en una población determinada, que tiene en cuenta la cantidad de interacciones de cada individuo en una población determinada.

La inexistencia de un algoritmo que tenga en cuenta las relaciones cercanas de cada individuo es un problema que se ha evidenciado en los avances epidemiológicos recientes, principalmente debido a la naturaleza de las bases de datos disponibles.

2 Justificación

A pesar del fuerte acercamiento de los modelos epidemiológicos a las dinámicas sociales con la llegada del coronavirus, se consideran interacciones a escalas que podrían ser mayores a las que se contagia la enfermedad, si bien existen modelos que analizan dinámicas de persona a persona, sus capacidades predictivas resultan limitadas por la naturaleza con la que se generan las bases de datos, como ocurre en [2, 8].

Teniendo en cuenta la aplicabilidad de los autómatas celulares y las cualidades predictivas de los modelos en redes neuronales, se pretende simular el comportamiento de una enfermedad y desarrollar un algoritmo que permita realizar pronósticos sobre su comportamiento, para responder a la pregunta: ¿Qué impacto tienen las relaciones sociales cercanas, en la propagación de una enfermedad?

Este trabajo permitirá comprender de manera más certera, como las interacciones de las personas afectan la propagación de alguna enfermedad y podrá establecer características en las bases de datos, que pueden dar lugar a análisis predictivos más realistas. Así mismo, se abre la puerta a que se propongan diferentes algoritmos con el fin de obtener diferentes resultados óptimos que alimenten el estudio epidemiológico.

3 Objetivos

Objetivo general

Diseñar un algoritmo en redes neuronales que permita realizar pronósticos en el comportamiento de una enfermedad simulada con autómatas celulares, teniendo en cuenta aspectos topológicos que describan las interacciones entre individuos.

Objetivos específicos

- Establecer la topología que describa las relaciones entre autómatas.
- Diseñar el conjunto de reglas probabilísticas que modelen las dinámicas de la enfermedad.
- Implementar las dinámicas de comportamiento de los agentes en Python, con el propósito de simular la propagación de la enfermedad.
- Identificar el conjunto de datos de entrenamiento adecuado para el algoritmo en redes neuronales.
- Definir las métricas que se van a monitorear con el algoritmo en redes neuronales.
- Diseñar e implementar un algoritmo en redes neuronales que permita realizar pronósticos para el comportamiento de la enfermedad simulada.

- Validar el comportamiento del algoritmo en ambientes topológicos distintos.

4 Antecedentes

Los modelos epidemiológicos compartimentales clásicos, descritos por Kermack y McKendrick, han sido fuente de numerosos estudios y simulaciones relacionadas con el estudio epidemiológico. Como caso particular, encontramos el modelo propuesto por Fuentes y Kuperman en [3], el cual permite estudiar la evolución de un núcleo de infectividad en varias condiciones y para dos tipos de enfermedades epidemiológicamente distintas, en autómatas celulares. A diferencia de los modelos clásicos en ecuaciones diferenciales [9], proponen un conjunto de reglas de evolución para determinar el estado de un individuo en función de su propio estado y el de sus vecinos en pasos de tiempo anteriores, considerando un radio de "interacción".

En trabajos como el de Mikler-Venkatachalam-Abbas en [7], los autómatas son interpretados como regiones, de modo que se puedan tener en cuenta factores geográficos, demográficos, de medio ambiente y patrones de migración, usando un autómata celular estocástico global (GSCA). Algo comparable se ve en [13], en donde H. White, Martín del Rey y G. Rodríguez implementan la misma noción de autómata espacial, para analizar entre otras cosas el proceso de vacunación para una enfermedad modelada con un enfoque determinista.

Un enfoque diferente se ve en [6], en donde Marr y Hutt se enfatizan en la relación entre topología y teoría de grafos, para analizar dinámicas de redes biológicas. Mostrando que la distribución de grados puede estar relacionada con las propiedades dinámicas de la red, además de mencionar posibles mejoras basadas en ideas de topología de redes para analizar diferentes grados de entropía.

Encontramos también aplicaciones entre la teoría de redes y la epidemiología. Moo Kim realiza un análisis interesante para la propagación de enfermedades en redes complejas, puntualmente en la red de Barabasi-Albert, presente en diversas redes reales [5]. Realiza además, una implementación de los modelos compartimentales clásicos en redes regulares y complejas.

Desplazándonos un poco a la teoría de redes neuronales y los procesos de machine-learning, encontramos trabajos como el de Schmidt en [12], en donde implementa una red neuronal artificial, basada en modelos compartimentales para predecir la propagación del virus de la influenza, para posteriormente realizar una comparación con un modelo de regresión lineal múltiple.

Otra aplicación interesante es el modelo desarrollado en [1], el cual se centra en la predicción de los ILI a largo plazo, usando datos con un rango de tiempo limitado a 20 semanas. Teniendo presente la pérdida de información temporal, exploran un modelo de propagación en grafos con representaciones espaciales. Usan además, las redes neuronales recurrentes para capturar las dependencias secuenciales en los datos de las series temporales locales y diseñan un módulo de convolución temporal para extraer automáticamente las depen-

dencias temporales cortas y largas de los datos de series temporales de múltiples localizaciones.

La reciente aparición del coronavirus a finales del 2019, ha incentivado estudios que proponen diversos enfoques basados en redes neuronales convolucionales de grafos (GCNs), memorias de corto plazo (LSTMs) y modelos de aprendizaje de transferencia basados en el meta-aprendizaje (MAML) [2, 4]. Mientras que en [4] proponen un modelo basado en machine-learning para estimar los parámetros de un modelo epidemiológico que se ajusta a tendencias observadas sobre el movimiento de individuos, en [2] proponen un enfoque de fusión entre redes neuronales basadas en grafos y modelos epidemiológicos para combinar directamente los datos de movilidad diádica y conectividad derivadas de Facebook con información estructural y espacial de distritos alemanes.

Panagopoulos, Nikolentzos y Michalis Vazirgiannis realizan un trabajo comparable en [8], en donde se centran en predecir el comportamiento de la enfermedad basándose en los datos de movilidad de masa, y consideran la movilidad dentro de una región como un indicador de interacción. Exponen además, la inexistencia de una base de datos que considere las posibles interacciones entre individuos y sus implicaciones.

Después de revisar diferentes desarrollos en el campo de la epidemiología, se ratifica la importancia de un análisis predictivo que tenga en cuenta las relaciones o interacciones más cercanas, de modo que se establezcan estrategias de prevención y contención mucho más certeras.

5 Metodología

Este proyecto estará dividido en cuatro fases principales, que se articulan entre sí y con los objetivos específicos del proyecto. Las fases se ejecutarán en forma secuencial y cada una de ellas generará una serie de módulos concretos que servirán de insumo a las fases subsecuentes. A continuación, se describe brevemente cada una de las fases metodológicas junto con la articulación con los objetivos específicos del proyecto:

Objetivos Específicos	Fases Metodológicas
1. Establecer la topología que describa las relaciones entre autómatas.	Fase 1: Definición de reglas de interacción entre individuos.
2. Diseñar el conjunto de reglas probabilísticas que modelen las dinámicas de la enfermedad.	
3. Implementar las dinámicas de comportamiento de los agentes en Python, con el propósito de simular la propagación de la enfermedad.	Fase 2: Simulación de la enfermedad.
4. Identificar el conjunto de datos de entrenamiento adecuado para el algoritmo en redes neuronales.	
5. Definir las métricas que se van a monitorear con el algoritmo en redes neuronales.	Fase 3: Diseño e implementación del algoritmo en redes neuronales.
6. Diseñar e implementar un algoritmo en redes neuronales que permita realizar pronósticos para el comportamiento de la enfermedad simulada.	
7. Validar el comportamiento del algoritmo en ambientes topológicos distintos.	Fase 4: Entrenamiento del algoritmo y desarrollo de pruebas.

Fase 1: Definición de reglas de interacción entre individuos.

La primera fase del proyecto consistirá en definir las interacciones para los diferentes participantes dentro de nuestra investigación. Partiremos de la definición de una topología definida por un sistema de vecindades locales, con el fin de simular las relaciones sociales entre individuos.

Posteriormente se hará uso de reglas totalísticas para definir las dinámicas de la enfermedad, teniendo en cuenta los comportamientos locales para cada individuo.

Fase 2: Simulación de la enfermedad.

Una vez se establezcan las dinámicas de la enfermedad y de los individuos, se implementarán en Python usando las librerías: numpy, random, matplotlib, entre otras, para cálculos con arreglos numéricos, valores aleatorios y gráficos respectivamente.

Fase 3: Diseño e implementación del algoritmo en redes neuronales.

Durante esta fase se definirán las métricas de interés para el estudio epidemiológico y se establecerán las neuronas que harán parte del algoritmo.

Al igual que la simulación de la enfermedad, el algoritmo definido en esta etapa, se desarrollará sobre Python usando las diferentes librerías disponibles para redes neuronales.

Fase 4: Entrenamiento del algoritmo y desarrollo de pruebas.

Para entrenar el algoritmo propuesto, se tomarán las bases de datos simulados en la fase 2. Posteriormente se realizarán modificaciones en los parámetros de la enfermedad, el sistema de vecindades locales y la condición de frontera del conjunto de individuos, para poner a prueba las cualidades predictivas del algoritmo propuesto.

6 Cronograma

Actividades a realizar	Semanas de ejecución de cada actividad															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Análisis de modelos SIS y SIR	■															
Definición e implementación del sistema de vecindades		■														
Diseño e implementación de reglas de comportamiento (enfermedad e individuos)			■	■												
Definición de métricas					■	■										
Diseño del algoritmo de clasificación en redes neuronales							■	■	■							
Entrenamiento del algoritmo										■	■					
Diseño e implementación de los escenarios de prueba												■	■	■	■	
Aplicación de pruebas sobre el algoritmo en redes neuronales y conclusiones																■
Redacción del documento final del trabajo de grado		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

References

- [1] Songgaojun Deng et al. “Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction”. In: *arXiv* (2019).
- [2] Cornelius Fritza, Emilio Dorigattia, and David Rugamera. “Combining Graph Neural Networks and Spatio-temporal Disease Models to Predict COVID-19 Cases in Germany”. In: *arXiv* (2021).
- [3] M.A. Fuentes1 and M.N. Kuperman. “Cellular automata and epidemiological models with spatial dependence”. In: *ELSEVIER* (1999).
- [4] Valerio La Gatta et al. “An Epidemiological Neural Network exploiting Dynamic Graph Structured Data applied to the COVID-19 outbreak”. In: *ELSEVIER* (2021).
- [5] Kyung Moo Kim. “Algunos modelos epidemiológicos en grafos completos y redes complejas”. MA thesis. Universidad de Sao Paulo, 2019.
- [6] Carsten Marr and Marc-Thorsten Hutt. “Cellular Automata on Graphs: Topological Properties of ER Graphs Evolved towards Low-Entropy Dynamics”. In: *entropy* (2012).
- [7] Armin R. Mikler, Sangeeta Venkatachalam, and Kaja Abbas. “Modeling infectious diseases using global stochastic cellular automata”. In: *World Scientific Publishing Company* (2005).

- [8] George Panagopoulos, Giannis Nikolentzos, and Michalis Vazirgiannis. “Transfer Graph Neural Networks for Pandemic Forecasting”. In: *arXiv* (2021).
- [9] Diego de Pereda Sebastián. “Modelización matemática de la difusión de una epidemia de peste porcina entre granjas”. MA thesis. Universidad Complutense de Madrid, 2010.
- [10] Mario Castro Ponce, Manuel de León, and Antonio Gómez Corral. “Las matemáticas contra la malaria y el modelo SIR”. In: *Matemáticas y sus fronteras* (2020).
- [11] Kao RR. “Networks and Models with Heterogeneous Population Structure in Epidemiology”. In: *Network Science*, 2010. Chap. 4.
- [12] Carina Schmidt. *Prediction of the Influenza Virus Propagation by using different Epidemiological and Machine Learning Models*. 2019.
- [13] S. Hoya White, A. Martín del Rey, and G. Rodríguez Sánchez. “Modeling epidemics using cellular automata”. In: *ELSEVIER* (2006).