



UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA

Scuola di Scienze

Dipartimento di Informatica, Sistemistica e Comunicazione

Corso di Laurea Magistrale in Informatica

AGENT-BASED MODELING AND LEARNING FOR EPIDEMIOLOGICAL STUDIES

Relatore: Prof. Antoniotti Marco

Correlatore: Prof.

Relazione della prova finale di:

Matteo Stievano

Matricola 829535

April 27, 2023

Anno Accademico 2022-2023

Abstract

L'abstract, che probabilmente dovrà essere presentato in doppia lingua (ITA, ENG) rappresenta un breve sunto di tutto l'elaborato di cui fa da cappello. Ovviamente non dovrà entrare nel dettaglio, e anzi dovrà essere scorrevole e chiaro, senza soffermarsi troppo su alcun dettaglio, in quanto ci saranno le sezioni successive per espandere tali conoscenze. Ad occhio e croce potrebbe prendere al massimo mezza pagina (così che se necessario ci stiano le due versioni su una pagina).

Contents

1 Introduzione

L'utilizzo di metodi e tecniche sempre più sofisticati da parte della comunità scientifica, in particolare modo da parte di epidemiologi e medici, è sempre stato argomento di grande dibattito e interesse. Negli ultimi anni il mondo si è espanso esponenzialmente divenendo sempre più connesso, incrementando drasticamente la probabilità che un virus affligga a livello mondiale la popolazione creando un disastro senza precedenti.

La storia dell'uomo è costellata di epidemie, e solamente alcune tra loro si sono guadagnate il privilegio di essere ricordate e tra queste solamente una piccola parte ha ottenuto il primato di essere ricordata come una catastrofe. Forse è proprio questo che le ha rese così salde nell'immaginario comune aumentando la loro già imponente aurea di terrore.

Per fare degli esempi possiamo citare: la peste nera [?] che a partire dalla metà del 14esimo secolo ritornò in Europa uccidendo venti milioni di persone in soli sei anni, l'epidemia di tifo [?] che non solo fu fatale durante il periodo delle crociate, ma anche durante la seconda guerra mondiale all'interno dei campi di concentramento nazisti, oppure le varie influenze, prima tra tutte quella spagnola [?] la quale nel periodo del primo dopo guerra, ovvero tra il 1918 e il 1920, uccise 50 milioni di persone in tutto il mondo.

Se ci si sofferma un attimo pensando proprio a questo tipo di pandemia, quella influenzale, ci viene da tirare un sospiro di sollievo, in quanto oramai come cittadini del primo mondo l'idea di influenza non ci fa più così paura. Eppure dovrebbe, e sfortunatamente lo abbiamo ricordato nel peggior modo possibile.

La pandemia di SARS-CoV-2 [?] che scoppiò negli ultimi mesi del 2019 condizionò l'intera umanità per circa 3 anni, e nel momento che sto scrivendo queste righe continua a condizionarla. Questa pandemia si è macchiata di aver mietuto, allo stato attuale delle cose, quasi 7 milioni di vite accertate. Questa tragedia rimarrà impressa nella memoria umana in quanto è stata capace di mettere in crisi l'intera specie umana, creando uno stato di allarme, panico e alle volte perfino isteria, che pochi altri avvenimenti sono stati in grado di fare.

Osservando le statistiche proprie di questa epidemia, ciò che balza subito all'occhio è sicuramente il numero associato alle morti e agli infetti: quasi 7 milioni di morti e più di 700 milioni di infetti [?]. Numeri che da soli basterebbero a mettere a disagio qualsiasi lettore, eppure altri dati, nascosti a prima vista, possono fornire ulteriori macabre informazioni. Un esempio tra tutti è l'effetto che un'epidemia del genere ha avuto sull'economia [?], portando disagi generalizzati ovunque.

Legato al disagio economico vi è un altro dato preoccupante definito come poverty trap [?]. Questo fenomeno nasce in quegli ambienti in cui le condizioni di povertà economica e la prevalenza di malattie possono imprigionare una società in uno stato persistente di bassa sanità e sempre maggiore povertà; questo fenomeno ciclico si auto sostiene e prende il nome di positive feedback [?]. Per ultimo, ma non per questo meno importante, l'effetto di una pandemia può portare ad instabilità governative le quali possono portare a un arretramento del sistema sanitario e di welfare [?], ricadendo come sopra descritto all'interno del fenomeno di positive feedback.

Questi sono solamente alcuni esempi dei problemi che possono sorgere, e che sono sorti con lo scoppio di una pandemia globale come è stata quella del COVID-19. Per questo la comunità scientifica, in particolare gli epidemiologi cercano soluzioni sempre più efficaci e accurate per prevenire, arginare e contrastare avvenimenti del genere.

L'epidemiologia è una disciplina nata di recente evolutasi insieme alle esigenze della società ogni qualvolta una nuova emergenza sanitaria faceva irruzione nella quotidianità. La prima definizione di epidemiologia è stata data da Lilienfeld [?] nel 1978, e cita:

“l'epidemiologia è un modo di ragionare riguardo le malattie, e si occupa di effettuare inferenza biologica derivata dall'osservazione di fenomeni patologici all'interno di una popolazione”.

Con il tempo questa definizione ha subito molti cambiamenti, derivati anche e soprattutto dall'espansione degli ambiti relazionati all'epidemiologia; con l'aggiunta ad esempio della farmacoepidemiologia, dell'epidemiologia molecolare e dell'epidemiologia genetica. Non solo, ambiti come etica, filosofia ed epistemologia sono estremamente importanti ed influenti nella crescita e sviluppo di questa materia [?].

Attualmente con il termine epidemiologia si intende la disciplina biomedica che studia la distribuzione e la frequenza delle malattie ed eventi di rilevanza sanitaria nella popolazione. L'epidemiologia si occupa di analizzare le cause, il decorso e le conseguenze delle malattie [?]. Secondo Last et al (1998) l'epidemiologia è definita come:

“lo studio della distribuzione e dei determinanti delle situazioni o degli eventi collegati alla salute in una specifica popolazione, e l'applicazione di questo studio al controllo dei problemi di salute.”

Uno degli strumenti più utilizzati in epidemiologia sono le simulazioni software [?]. Dato lo scopo dell'epidemiologia, questa necessita di avvalersi di modelli matematici [?] che aiutano i ricercatori a trarre conclusioni sul sistema che analizza. Sistemi simili vengono definiti come complessi, [?] [?] ovvero sistemi dinamici a multicomponenti che tipicamente interagiscono tra loro, e che sono descrivibili tramite modelli matematici.

I primi modelli chiamati compartimentali [?] basano il proprio funzionamento sullo studio di gruppi di individui disgiunti che interagiscono tra loro, analizzabili tramite un sistema di equazioni ordinarie differenziali (ODE) [?]. Questo approccio è stato teorizzato da Kermack e McKendrick nel 1927 applicando una modellazione matematica al comportamento delle malattie infettive su un gruppo di individui, tenendo in considerazione la variabile del tempo. Da qui è nato il famoso modello Susceptible-Infectious-Recovered (SIR), il quale tuttora viene utilizzato ampiamente insieme alle sue varianti. Successivamente con l'unione di svariate discipline come: la teoria dei giochi, i sistemi complessi, i comportamento emergente, la sociologia computazionale, i sistemi multiagente e la programmazione evolutiva sono nati i modelli ad agente, modelli computazionali autonomi per la simulazione di sistemi complessi.

Grazie alla loro teorizzazione negli anni 1940 ma soprattutto grazie al loro uso concreto dagli anni 1990 il mondo della simulazione ha avuto un grande balzo in avanti. Ciò che rende questi sistemi molto flessibili e potenti è la capacità di far emergere spontaneamente dei comportamenti complessi da un insieme di regole semplici. Ovviamente però la richiesta di risorse e capacità computazionale è estremamente elevata in rispetto alla controparte di modelli puramente matematici SIR.

Ogni approccio ha dei pro e dei contro e la vera potenza di questi modelli sta principalmente nella acutezza di chi deve poi farne uso. In base ai compromessi e le assunzioni fatte durante la fase di modellazione ogni approccio può rivelarsi vincente. Uno dei compromessi maggiori che viene generalmente applicato a questi modelli è quello della discretizzazione dell'ambiente. La realtà è

a tutti gli effetti un insieme continuo di eventi, ma essendo tutti i dispositivi di calcolo discreti è impossibile simulare avvenimenti continui (siano essi nel tempo o nello spazio) in maniera diretta e perciò bisogna fare dei compromessi, trasformando il proprio spazio di lavoro in uno più adatto alle macchine che lo devono simulare.

Un'altra assunzione generalista e approssimativa per definizione, ma necessaria per la costruzione di un modello che analizzi il decorso di un'epidemia all'interno di una società, è quella legata alla tipologia di comportamento che verrà mostrato dagli esseri umani in condizioni di pericolo. Queste sono solo alcune delle assunzioni e compromessi che bisogna fare quando ci si appropria al mondo della simulazione. Tuttavia applicare delle assunzioni, alle volte anche forti e controintuitive, non sempre è sinonimo di errore.

Alle volte tramite lo studio della causalità degli eventi è possibile astrarre un set minimale di assunzioni che se applicate danno la capacità al modello di rappresentare molto bene il comportamento desiderato. Certamente ci saranno alcune discrepanze soprattutto in casi estremi, ma è un'eventualità che viene tenuta in considerazione ogni qualvolta si parla di simulazione, e che non è possibile eliminare del tutto.

La causalità o causa effetto, è quell'influenza per cui un evento, un processo, uno stato o un oggetto contribuiscono nella produzione di un nuovo evento, processo, stato o effetto, dove la causa è parzialmente responsabile dell'effetto e l'effetto è parzialmente dipendente dalla causa. A prima vista non sembra una tematica molto complessa o di difficile approccio, complice il fatto che in quanto esseri umani siamo una specie che si è evoluta per trovare una correlazione tra gli eventi, ma correlazione non significa, e soprattutto non implica, causalità. Questo tema si riferisce all'incapacità legittima di dedurre la relazione di causa - effetto tra due eventi o variabili solamente sulla base di una osservazione della loro associatività o correlazione. Diventa chiaro come una delle parti più complesse dell'epidemiologia sia proprio quella di stabilire le cause di un dato fenomeno e comprendere come un determinato intervento su di esse influenzi quest'ultimo.

Questa rapida introduzione all'epidemiologia e alla simulazione di sistemi complessi tramite l'utilizzo in particolare di modelli ad agente sarà l'argomento cardine dell'intero elaborato. Nelle sezioni successive verrà proposta un'analisi dello stato dell'arte dell'attuale panorama epidemiologico e simulativo, con alcune rapide digressioni sul problema della discretizzazione, problema assai sentito nell'ambito della simulazione.

Successivamente verranno portati alla luce alcuni esempi e modelli di sistemi ad agente che si occupano di modellare differenti tipi di epidemie, tutte caratterizzate però dal fatto di essere epidemie infettive a diffusione principalmente aerea, come ad esempio l'ebola, l'avaria e l'influenza.

Successivamente vi sarà un'analisi più dettagliata riguardo la pandemia da COVID-19 che recentemente ci ha colpito. Infine verranno analizzati alcuni metodi di monitoraggio di queste simulazioni con un focus su alcuni metodi di intervento.

2 Stato dell'Arte

2.1 Modelli Compartmentali

2.1.1 Modelli Deterministici

2.1.2 Modelli Stocastici

2.2 Modelli ad Agente

2.2.1 Discretizzazione

Tempo

Spazio

2.3 Epidemiologia

2.4 Julia

Julia è un linguaggio di programmazione ad alto livello, multi-paradigma e open-source ideato per compiere analisi numerica ed effettuare operazioni di computer science in maniera rapida e stabile. Julia è nato ufficialmente come linguaggio di programmazione nell'anno 2012 con lo scopo di fornire uno strumento potente, robusto e veloce tanto se non più dei linguaggi considerati in questo ambito lo stato dell'arte, ovvero C e Fortran; ma anche facile da approcciare, al contrario dei linguaggi sopra citati. Julia è un linguaggio di programmazione scritto in C++ e Scheme, ma gran parte della sua composizione è scritta in Julia stesso.

le caratteristiche principali di questo linguaggio sono principalmente:

- Alte performance: lo scopo per cui Julia è nato è stato quello di offrire un linguaggio estremamente performante con la capacità di poter compilare i programmi in codice nativo per molteplici piattaforme grazie all'utilizzo di LLVM
- Dinamico: la scelta di rendere Julia un linguaggio dinamicamente tipizzato lo rende di facile utilizzo in quanto rende molto più semplice il suo approccio anche a chi non ha una base solida di programmazione, in quanto ritorna la stessa sensazione di immediatezza di un linguaggio di scripting. Inoltre questo permette un alto supporto per l'uso interattivo
- Ambiente riproducibile: lo scopo del linguaggio è quello di poter permettere all'utente di ricreare le stesse condizioni ogni volta su ogni macchina su cui un programma viene eseguito. Questo può essere ottenuto tramite l'utilizzo di file binari pre compilati
- Componibile: Julia utilizza l'approccio multiple dispatch come paradigma, permettendo una grande flessibilità nell'esprimere una elevata quantità di pattern di programmazione, dall'object-oriented al funzionale
- General Purpose: lo scopo del linguaggio è quello di creare un ecosistema in grado di poter soddisfare qualsiasi esigenza di un utente, permettendo la creazione di applicativi e microservizi senza dover ricorrere ad integrazioni con codice non nativo Julia
- Open source: Julia abbraccia la filosofia open source, e il codice sorgente dell'intero linguaggio, così come di tutte le librerie è disponibile sulla piattaforma GitHub sotto la licenza MIT. Questo permette una crescita eterogenea grazie al contributo di più di 1000 utenti che si impegnano a migliorare il linguaggio

2.4.1 Agents.jl

Seguendo la filosofia propria del linguaggio di programmazione in cui è sviluppata, la libreria Agents.jl è stata sviluppata con l'idea di essere facile da imparare e usare ed estendibile, con forte attenzione sulla creazione ed evoluzione di modelli veloci e soprattutto scalabili. Molteplici esempi comparativi sono stati effettuati mostrando come il framework sviluppato permetta di avere un notevole guadagno prestazionale rispetto ai maggiori competitor attualmente presenti sul mercato (Mesa, Netlogo, MASON).

La facilità di interazione con questa libreria non è da confondersi con una mancanza di opzioni durante lo sviluppo, in quanto nativamente Agents.jl permette l'integrazione con altre librerie che

in maniera altrettanto semplice e veloce offrono all'utente la possibilità di addentrarsi nel mondo del machine learning, in particolar modo il mondo del Scientific Machine Learning, branca che soprattutto grazie alla pandemia da Covid-19 ha visto un enorme interesse per l'analisi di dati per lo sviluppo di policy di prevenzione e contenimento della pandemia.

2.4.2 SciML.ai

SciML è una collezione di librerie dedite al calcolo scientifico e non solo al machine learning. Questo framework permette di avere tutti gli strumenti per poter utilizzare facilmente, velocemente e in maniera robusta tecniche di analisi numerica molto avanzata, così da poter sviluppare applicazioni semplicemente senza essere banali.

Durante la pandemia da Covid-19 questo framework è stato utilizzato per sviluppare applicazioni le quali grazie all'utilizzo di tecniche di scientific machine learning riuscivano sia a prevedere in maniera molto accurata l'andamento dell'epidemia seppur in presenza di una scarsa quantità di dati (uno dei grandi problemi dei modelli di machine learning e di artificial intelligence sono gli enormi dataset necessari per addestrare le reti in maniera robusta) e le stesse presentavano misure di contenimento e prevenzione che si sono dimostrate essere efficaci.

Questa suite è stata largamente integrata e supportata anche da Microsoft.

SafeBlues

Un esempio di un modello di scientific machine learning può essere il modello denominato SafeBlues il quale simulando una rete bluetooth in cui gli individui potevano venire infettati da un virus e poi infettare a loro volta tutti gli individui nella rete con una certa probabilità, aveva riprodotto fedelmente l'andamento della pandemia da Covid-19. In aggiunta questa soluzione, aveva mostrato come l'applicazione di policy per il contenimento del virus bluetooth erano perfettamente applicabili anche al caso reale della pandemia.

3 Metodi e Modelli

3.1 COVID-19

3.1.1 Modelli SEIR

3.1.2 Modelli ad Agente

3.2 Altre epidemie

3.2.1 Ebola

3.2.2 Aviaria

3.2.3 Influenza

3.3 Monitoraggio e Intervento

4 Risultati Ottenuti

5 Conclusioni