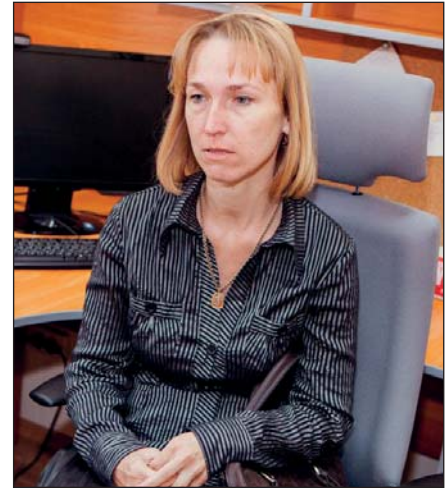


ПРОФЕСОР НАТАЛІЯ ШАХОВСЬКА:

«50,5 МЛН ЛЮДЕЙ У СВІТІ ЗАХВОРИЛИ НА COVID-19...»



Національний фонд досліджень України у жовтні 2020 року визначив перелік проєктів конкурсу «Наука для безпеки людини та суспільства», які рекомендуються до реалізації за рахунок грантової підтримки. Серед проєктів – «Система підтримки прийняття рішень моделювання поширення вірусних інфекцій», автором якого є завідувачка кафедри систем штучного інтелекту Інституту комп'ютерних наук та інформаційних технологій Національного університету «Львівська політехніка», доктор технічних наук, професор Наталія Шаховська, інтерв'ю з якою пропонується Вашій увазі.

— Пані Наталю! Чим викликана поява Вашого проєкту?

— На перебіг хвороб, викликаних інфекціями та вірусами (навіть з відомими схемами запобігання на лікування) впливають різні фактори, а саме: мінливість штамів, характер взаємодії, особливості території поширення. Саме тому розроблення імітаційних моделей поширення та протікання захворюваності на різного роду інфекції та віруси є складним науковим завданням. Це завдання характеризується: багатокритеріальністю: тип поширення (епідемічне поширення, контрольоване поширення у легкій формі захворювання), початкові дані поширення, територія поширення; залежністю від часу; інтервалом моделювання, різнотипністю вхідних даних.

Коронавірусне захворювання (COVID-19) – це інфекційна хвороба, яку спричиняє новий штам коронавірусу.

На даний час 50,5 млн людей у всьому світі захворіли на COVID-19. З них –33,1 млн були виліковані. Від хвороби померло понад 1,26 млн людей.

— У багатьох країнах?

— У 203 країнах.

— В чому полягає новизна проєкту, запропонованого Вами?

— Пропонується створити інформаційну систему моделювання та прогнозування поширення інфекцій різного роду, що має здатність масштабування (країна, регіон, місто). Дана система буде базуватись на двох підходах до моделювання поширення інфекцій та визначення причин відхилень даних. Перший метод зосереджений на моделюванні перебігу хвороби та визначенні основних факторів впливу на конкретного індивіда, другий – імітація в режимі реального часу просторового поширення інфекції з можливістю динамічної зміни параметрів.

Перший метод передбачає пошук прихованих залежностей даних для визначення характеру перебігу хвороби індивіда. Аналіз великих обсягів даних вимагає визначення груп атрибутів, які утворюють функціональні залежності. Однак у реальних наборах даних, отриманих з різних джерел, важливі залежності визначені лише для підмножини значень групи атрибутів (існують залежності, наприклад, між перенесеними раніше захворюваннями та характером перебігу хвороби зараз – така залежність встановлена між підмножинами значень різних кортежів і не може бути знайдена існуючими методами пошуку прихованих даних). Ми будемо називати такі залежності частковими функціональними залежностями. Відповідно, рівень підтримки таких залежностей є низьким, що не дає змоги використовувати їх для подальшого аналізу даних. Водночас, часткові функціональні залежності є модифікованими асоціативними правилами, але такими, що виконуються лише для частини даних, і залежить від фактору часу. Тому пропонується розробити апарат ймовірнісних секвенційних залежностей. Метод пошуку таких залежностей базуватиметься на основі відкладених обчислень (модифікація FP-дерева). Це дозволяє зменшити часову складність та використовувати паралельний і розподілений режим для розрахунку. Отже, алгоритм пошуку залежностей може бути реалізований на MapReduce.

— А імітація?

— Імітація – в режимі реального часу поширення інфекції. Як результат отримується динаміка (просторова та статистична) таких показників, як кількість інфікованих, здорових, тих, що перехворіли, летальних випадків та інші показники. Також буде можливим імітувати різні сценарії запобігання поширенню інфекції на кшталт карантину чи інших заходів. Оцінювати їх ефективність та науково-обґрунтовано приймати стратегічні рішення.

— Ви вже виробили і методологію дослідження?

— Звичайно. Задача пошуку залежностей в даних потребує аналізу залежностей між десятками параметрів досліджуваного процесу та сотнями можливих джерел впливу на цей процес. Залежності носять недетермінований характер, і тому моделювання потребує застосування статистичних методів аналізу випадкових процесів. Значна частина інформації часто прихована від спостереження або ж спостереження за нею не ведеться. Це вносить багато труднощів у процес аналізу зібраної інформації.

— Наприклад?

— На сьогодні розроблені методи статистичного аналізу дають змогу працювати з частково невизначеними розмитими процесами. Але наявні методи мають істотні обмеження в галузі застосування і типах даних, що можуть аналізуватись цими методами.

Іншою особливістю медичних даних є їх ієрархічність та мережевість. До мережевих даних належить інформація про супутні патології, алергічні реакції тощо, що також є прямим або непрямим чинником, який визначає характер захворюваності індивіда. Отже, необхідним є відшукування не тільки лінійних залежностей у даних.

— Словом, роботи – не початий край...

— Безперечно. Адже усі вищезазначені фактори можуть негативно впливати на проведення, інтерпретацію і узагальнення результатів досліджень, а також на розуміння й тлумачення досліджуваного феномену. У роботі пропонується розроблення підходу моделювання характеру захворюваності індивіда на основі підходу великих даних

— Вироблено вже й метод?

— Так. Метод складається з двох частин: 1) пошук ймовірнісних продукційних залежностей на основі моделі великих даних; 2) використання ймовірнісних продукційних залежностей для моделювання характеру захворюваності. В основі цього методу буде запропоновано розробку вже спеціальних методів формування навчальної множини даних і попередньої обробки атрибутів з урахуванням специфіки контенту медичних даних та даних навколишнього оточення. А ще розроблення ансамблів моделей імпутації даних на основі базових моделей різнобічної природи у складі спеціалізованої інформаційної технології відновлення пропущених даних для автоматизованого опрацювання інформації.

Наступним етапом буде прогнозування динаміки поширення інфекції та моделювання різних сценаріїв впливу з боку держави. Ця модель базується на теорії агентних систем та може бути реалізована методами імітаційного моделювання. Для моделювання необхідні дані, отримані з першої частини розробки та відомих статистичних даних.

— Про що йдеться?

— Це – кількість населення центру та всіх районів області, густина населення обласного центру та районів, соціальна дистанція між людьми, тривалість хвороби, імовірність захворювання при контактах людей, рівень смертності, наявність місць скупчення людей (супермаркети, церкви, аптеки, ринки, об'єкти будівництва, спортзали), відсоток людей, що переносять хворобу безсимптомно, наявність процедури ізоляції хворих людей, можливість переміщення людини з району до обласного центру та назад, дотримання людьми необхідної дистанції, інкубаційний період та інші.

Для стійкості результату необхідно використовувати ансамблі моделей, які легко розпаралелити. Тому реалізація розрахункового ядра має виконуватись на комп'ютерних кластерах.

— Хто ще, окрім Вас, дотичний до поданого проєкту?

— Професор **Виклюк Я.І.**, доценти **Мельникова Н.І.**, **Кривенчук Ю.П.**, **Ізонін І.В.**, які готові взяти участь у даному проєкті, враховуючи характер їхньої наукової діяльності.

— І, головне: а які очікувані результати від виконання проєкту «Система підтримки прийняття рішень моделювання поширення вірусних інфекцій», науковим керівником якого Ви є?

— Очікуємо отримати таку наукову і науково-технічну продукцію:

1) метод та алгоритм формування та аналізу великих даних для побудови інформаційного портрету досліджуваного об'єкта за необхідності їх консолідації та паралельної обробки;

2) метод та алгоритм передбачення складності перебігу хвороби на основі секвенційних асоціативних залежностей;

3) метод пошуку залежностей у багатовимірних даних на основі ансамблю моделей (стеккінг) та з використанням регуляризації інформаційних ознак;

4) імітаційна модель поширення захворюваності населення, спричиненої вірусом чи інфекцією, на основі мультиагентних систем;

5) інформаційна технологія для моделювання поширення захворюваності із візуалізацією на карті, а також рекомендаційна лікарська система моделювання перебігу захворюваності індивідом.

— В такому випадку, пані Наталю, і світ буде здоровим, і Україна. Чи не так?

— А чому би й ні? Людей потрібно лікувати – світ потрібно рятувати. Правда ж?

— Правда, пані Наталю! Дякую Вам за розмову, а передусім – за Вашу з колегами роботу, яку Вам вдасться (я в це вірю) провести з думкою про людину і світ. ■

Спілкувався
Богдан Залізняк,
керівник пресцентру наукової журналістики
ЗНЦ НАН України і МОН України,
член НСПУ і НСЖУ
м. Львів