УДК 519.67

МОДИФІКОВАНА SEIRA МОДЕЛЬ ПОШИРЕННЯ ДЕЗІНФОРМАЦІЇ В ІНТЕРНЕТІ

Віолета Третиник, Микола Давиденко

MODIFIED SEIRA MODELS OF THE SPREAD OF DISINFORMATION ON THE INTERNET

Violeta Tretynyk, Mykola Davydenko

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Берестейський , 37, Київ, 03056, Україна

Abstract*. In the context of a full-scale war, a country's information security becomes critically important. The information space is highly saturated with various sources, and fake news often serves as a tool for destabilizing public sentiment. The spread of disinformation can be interpreted similarly to the spread of a virus, where users transition between different states. This paper explores approaches to detecting disinformation and modeling the dynamics of its diffusion.*

Гібридна модель поширення дезінформації враховує два основні процеси:

1) розпізнавання дезінформації;
2) моделювання динаміки поширення фейка в Інтернеті.

Для розпізнавання дезінформації запропоновано використати графові нейронні мережі (GNN). GNN - це тип глибокої нейронної мережі, яка може використовуватись для навчання на графових даних[1]. GNN застосовують механізм передачі повідомлень для агрегування інформації від сусідніх вузлів, що дозволяє їм захоплювати складні взаємозв'язки в графах[1]. GNN ефективні для виконання різних завдань, таких як класифікація вузлів, передбачення зв'язків та кластеризація[1]. Вибір такого підходу зумовлено особливістю даних, що ми отримуємо з публікацій (контекст новин, активність користувачів в соціальних мережах, реакції, коментарі тощо). Таке рішення дає можливість будувати нетривіальні залежності в неевклідовому просторі між об'єктами, та враховувати більше, ніж характеристики ізольованого об'єкта. Тобто його зв'язки та дії відносно інших користувачів, також аналізуються, що особливо корисно в контексті вирішення даної задачі.

Новина у графі подається у вигляді вузла, а зв'язки між вузлами — це взаємодії. Дані з соціальних мереж природно підходять для GNN. Повідомлення несуть не лише зміст, але і контекст. Таким чином ми можемо враховувати латентні залежності та приховані патерни фейкової новини, які важче зафіксувати класичними методами.

Щодо ознак характерних фейкам, можна виділити такі:

1) централізоване поширення з фальшивих акаунтів;
2) аномальна динаміка активності;
3) чутливі теми та маніпуляції.

Урахування характеристик не лише ізольованого об'єкта, дозволяє краще виявляти фейкові новини. Ця важлива перевага GNN дає можливість більш чутливо ідентифікувати дезінформацію.

Для автоматизованих систем моніторингу цей підхід стане ефективною складовою, хоча варто зазначити, що таке рішення вимагає значних обчислювальних ресурсів.

В подальшому досліджені можлива модифікація GNN у графову згорткову мережу(GCN) [2]. Вона є доречною тому, що GCN використовує спектральну згортку для агрегації ознак вузлів на основі їх сусідів у графі, що дозволяє ще більш ефективно моделювати структуровані дані. Модель GCN використовує ефективне правило пошарового поширення, яке базується на наближенні першого порядку [2]. Модель GCN здатна кодувати як структуру графа, так і характеристики вузла способом корисним для напівконтрольованої класифікації зображення спектральних згорток на графіках [2].

Безпосереднє моделювання розповсюдження фейкової новини відбувається з використанням модифікованої SEIRA-моделі. Такий вибір зумовлено можливістю детальніше описувати поведінку користувачів в соціальних мережах. Система диференційних рівнянь включає в себе 5 окремих станів: S – сприятливі; E – байдужі; I – інфіковані; R – вилікувані; A – антифіковані.

$$\begin{cases} \frac{dS}{dt} = -\beta\frac{SI}{N} + \gamma R + \alpha A \\ \frac{dE}{dt} = \beta\frac{SI}{N} - \xi E \\ \frac{dI}{dt} = \xi E - \delta I - \theta I + \alpha A, \\ \frac{dR}{dt} = \delta I - \gamma R \\ \frac{dA}{dt} = \frac{\theta EI}{N} - \alpha A \end{cases}$$

де, N – загальна кількість користувачів, β – ймовірність зараження дезінформацією, ε – швидкість переходу зі стану E до I, γ – ймовірність, що користувач знову буде поширювати інформацію, δ – показник, того як швидко, користувач перестає поширювати інформацію, θ – швидкість з якою користувачі переходять до антифікованих, α – швидкість з якою, антифіковані стають сприятливими.

В такому варіанті диференційних рівнянь з'являється стан Е, що описує ігнорування людиною новини. Цей стан справді є одним з найважливіших показників залученості аудиторії в реальних випадках. Додатково з'являються антифіковані користувачі, які розуміють та аргументують фальшивість інформації іншим станам. Разом з тим з'являються повернення в попередні стани, передумовою для чого може бути: втома від висвітлення дезінформації чи повернення до інфікованих, через слабке розуміння проблеми предметної області.

Модифікація рівняння $\frac{dS}{dt}$ є суттєво гнучкішою в порівнянні з класичним варіантом SIR-моделі. Таким чином ми отримуємо можливість враховувати повторне зараження, втрату мотивації та спостерігати за циклічністю станів. Рівняння $\frac{dI}{dt}$ також відображає згасання інтересу у аудиторії до свідомої боротьби за правду окрім природньої втрати інтересу користувачів. Ще одним рівняння є $\frac{dR}{dt}$, де описано групу користувачів, що вилікувались (втратили інтерес та перестали поширювати фейк), проте все одно залишається частка користувачів, які могли взяти паузу, щоб повернутись з новими силами. Тобто ми можемо відобразити циклічну особливість та схильність користувачів вірити фейковим новинам.

Тож, ми маємо можливість моделювати схильність користувачів вестись на маніпуляції та розглядаємо фактор вагань (користувачі змінюють стани). Рішення про опис динамік поширення дезінформації такими рівняннями обумовлено більш гнучким та реалістичнішим рішенням, оскільки такий підхід дозволяє описати взаємодію між різними станами. Це може дати точніші інтерпретації процесів поширення дезінформації в Інтернеті.

Таким чином ми можемо врахувати нестабільну поведінку людей в Інтернеті на що мають вплив такі фактори: емоції, критичне мислення, вплив оточення тощо. На виході отримуємо адаптивну модель з урахуванням змін поведінки користувачів. Тож, поєднання графової нейронної мережі та моделювання динаміки розповсюдження фейка дає сильний допоміжний інструмент для прийняття рішень в боротьбі з дезінформацією.

Бібліографія

1. Zhang J, et al. Graph neural networks: A review of methods and applications. Journal of Big Data. 2023. https://doi.org/10.1186/s40537-023-00876-4
2. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. 2016. arXiv preprint arXiv:1609.02907.