УДК 519.876.5

МАТЕМАТИЧНІ МОДЕЛІ ДИНАМІКИ ПОШИРЕННЯ ІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕДІА

Ю. В. Наконечна[1, a], А. Б. Качинський[1]

[1]*Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Фізико-технічний інститут*

Анотація

У віртуальному просторі користувачі залишають величезну кількість інформаційних слідів – повідомлення, коментарі, фото, відео, особисті дані. Велика кількість джерел інформації та її споживачів, особливості її впливу та взаємодії із користувачами мережі роблять тему моделювання поширення інформації актуальною. Вивчення цього процесу дає можливість оцінити кількість кінцевих споживачів, прогнозувати швидкість та отримувати дані про стан поширення інформації. В роботі було вивчено математичні моделі клітинного автомата, Дейлі-Кендалла та розширену модель типу хижак-жертва та застосовано їх для вивчення часових та просторових характеристик інформації в процесі її поширення. За результатами роботи отримано адаптовані математичні моделі, що дають змогу не тільки отримувати інформацію про реальний стан процесу поширення деякої новини, а й прогнозувати масштаби поширення інформаційних повідомлень.

Ключові слова: математична модель, модель Дейлі-Кендалла, клітинний автомат, модель хижак-жертва, поширення інформації, соціальні медіа

Вступ

Якщо інформація потрапила в Інтернет і здалася цікавою комусь із користувачів, вона може бути моментально поширена будь-якою зацікавленою особою на теоретично необмежену кількість інтернет-майданчиків різного роду (таких як сайти з можливістю постингу та коментування, соціальні мережі, блоги, форуми і ряд інших).

Оцінка ефекту від поширення інформації через ЗМІ є складним завданням. Однак, з появою веб-блогів та інших мережевих медіа (новинні портали, форуми та ін.), аналізувати поширення інформації стало простіше. Вся інформація зберігається в мережі у відкритому доступі, користувачі відкрито діляться своїми думками з приводу інформації як в текстовому вигляді, так і через рейтингові системи (наприклад, позначки «мені подобається/лайк», «поділиться/репост»). Це все дозволяє більш точно вивчати процеси дифузії інформації, оцінювати ефект від поширення. Однак, моделювання дифузії в соціальних мережах залишається складним завданням. Досить складно отримати велику кількість різноманітної інформації з різних джерел, обробити і відстежити елементи соціальних мереж такі як: рекомендації, посилання, теги, повідомлення, фрази і «меми».

Однак, величезна кількість інформації у відкритому доступі має властивості, що дають змогу застосовувати до процесу поширення інформації математичні моделі з метою вивчення його особливостей, виконувати статистичну обробку даних про поширення деякого інформаційного повідомлення тощо. В свою чергу, інформаційна війна, в стані якої знаходиться Україна на даний момент, є джерелом повідомлень, аналіз поширення яких є завданням не лише математики, але й кібернетичної безпеки [1].

1. Моделі поширення інформації

Розширена модель SIR

Модель SIR – це класична віральна модель поширення інфекції в межах популяції в залежності від часу. Процес поширення інформації та хвороби має подібність.

Модель SIR характеризується наявністю об'єктів трьох типів:

S – не заражені,

I – заражені,

R – вилікувані об'єкти, що мають імунітет.

Для процесу поширення інформації нехай:

S – агенти мережі (користувачі), які не отримали інформаційне повідомлення,

I – отримали повідомлення та вважають надану в ньому інформацію актуальною,

R – забули новину/втратили до неї інтерес.

Загальна структура системи, що описується моделлю SIR, може бути представлена у вигляді:

$$S(t) + I(t) + R(t) = N$$

де $S(t)$, $I(t)$, $R(t)$ – кількість об'єктів в кожному класі [2, 3].

[a]nakonechna.yu@gmail.com

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)}{N}S(t),$$
$$\frac{dI(t)}{dt} = \frac{\beta I(t)}{N}S(t) - \gamma I(t),$$
$$\frac{dR(t)}{dt} = \gamma I,$$
$$\frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = 0.$$

де β – частота зараження (отримання новини користувачем),

γ – частота лікування або швидкість імунізації (швидкість забування повідомлення або його переходу до неактуального стану).

Зауважимо, що поширення новини можливо лише за умови $\beta < \gamma$.

Соціальній мережі властива мінливість в часі – це означає, що агенти можуть приєднатись до мережі або покинули її. Розширимо модель. Позначимо параметром μ середню частоту приєднання до мережі в одиницю часу. Параметром δ будемо вважати середню частоту виходу агента із мережі в одиницю часу, тобто частоту «забування» новини в одиницю часу. Система рівнянь буде мати вигляд [1, 3]:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)}{N}\mu(N - S(t)) + \alpha R(t),$$
$$\frac{dI(t)}{dt} = \frac{\beta I(t)}{N}S(t) - \gamma I(t) - \delta I(t),$$
$$\frac{dR(t)}{dt} = \gamma I - \delta R(t) - \alpha R(t),$$
$$\frac{dS(t)}{dt} + \frac{d(t)I}{dt} + \frac{dR(r)}{dt} = 0$$

Модель Дейлі-Кендалла (DK-модель)

Метод Далея-Кендала – математична модель імітації процесу поширення інформації (чуток, повідомлень), яка також називається DK-модель. Дана модель ділить всіх учасників поширення деякого інформаційного повідомлення на три групи:

1) група, що починає поширення повідомлення (U);
2) група, що після поширення повідомлення продовжує його поширювати (V);
3) група, що після отримання повідомлення приймає рішення не поширювати його (W).

Модель представлена на рис. 1 :

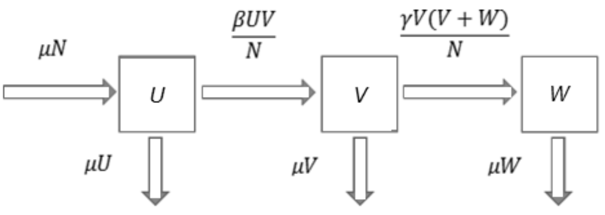

Рис. 1. DK-модель поширення повідомлень соціальною мережею.

У моделі N – число учасників процесу поширення. Повідомлення поширюється з ймовірністю $\frac{\beta}{N}$. Ступінь прийняття повідомлення визначена параметром μ. Коли поширювач повідомлення зіштовхується з аудиторією W, то поширення припиняється і ймовірність того, що це відбудеться, дорівнює $\frac{\gamma V(V-W)}{N}$. Новина, подана в повідомленні, втрачає свою цінність із плином часу. Така ймовірність визначається фактором γ. Це можна пояснити тим, що новина перестає бути новинкою або не залишається частин, які можна передати. Модель можна представити у вигляді рівнянь [1, 4, 5]:

$$\frac{dU(t)}{dt} = \mu N - \frac{\beta U(t)V(t)}{N} - \mu U(t),$$
$$\frac{dV(t)}{dt} = \frac{\beta U(t)V(t)}{N} - \frac{\gamma V(t)(V(t) - W(t))}{N} - \mu N,$$
$$\frac{dW(t)}{dt} = \frac{\gamma V(t)(V(t) - W(t))}{N} - \mu W(t).$$

Модель клітинного автомата

Клітинний автомат – це дискретна динамічна система, що включає однорідні клітини, з'єднані один з одним. Всі клітини утворюють клітинний автомат. Стан кожної клітини визначається клітинами, що знаходяться в околиці даної клітини. Набір «найближчих сусідів» називається околом кінцевого автомата з номером j. Стан клітинного автомата j в момент часу $t + 1$ визначається наступним чином:

$$y_j(t + 1) = F(y_i(t), O(j), t)$$

де F – деяке правило(виражене, наприклад, мовою булевої алгебри), $O(j)$ – сусіди, t – крок.

Клітинний автомат визначається правилами:

1) зміна значень кожної клітини відбувається одночасно (кроком є зміна одиниці часу);
2) мережа клітинного автомата є однорідною, тобто правила зміни однакові для всіх клітин;
3) число станів клітини є скінченним.

Теорія клітинних автоматів використовується для аналізу дифузії інновацій, цей процес дуже схожий на поширення новин в Інтернеті. Найпростіша функція перетворення моделі відповідає наступним правилам: індивідуум відповідає одній клітці, яка може приймати два стани: 1 – новина прийнята, 0 – новина не прийнята. Передбачається, що один раз прийнявши інформацію, стан залишається незмінним. Автомат приймає рішення про прийняття новини орієнтуючись на думку найближчих сусідів, якщо серед сусідів m підтримали інновацію і p – ймовірність прийняття новини (генерується в ході роботи моделі), тоді якщо $pm > R$, де R – фіксоване порогове значення, клітина приймає новину. Крім того, можуть бути накладені додаткові умова на тип новини: клітина має свіжу новину (чорний колір), у клітини знаходиться застаріла інформація (сірий колір), клітина не має інформації або забула про неї (білий колір) [6, 7].

2. Результати дослідження

Обрані для роботи моделі було застосовано на даних, отриманих з новинного порталу УНІАН. Аудиторія каналу становить більше 200 тис. підписників у Facebook, та активну щоденну аудиторію ресурсу близько 9 тис. чоловік, з яких 1428 переглянули новину про відсутність депутатів на своїх робочих місцях 06.04.18. Зважаючи на популярність даної новини станом на день публікації, було обрано відповідні параметри рівнянь моделей. Після моделювання отримали результати, показані на рис. 2 і 3 (пунктиром позначено реальний стан поширення новини):

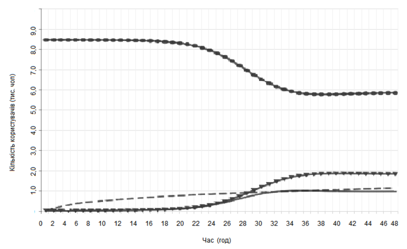

Рис. 2. Розширена SIR-модель поширення новини.

На рис. 2 видно, що за моделлю можливо оцінити кількість людей, що переглянули новину, але вона не відповідає динаміці збільшення переглядів у часі.

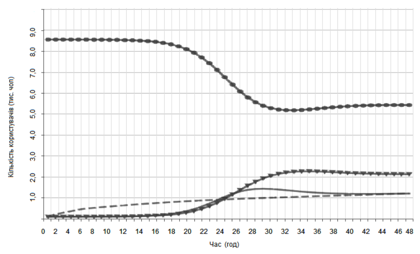

Рис. 3. DK-модель поширення поширення новини.

Модель Дейлі-Кендалла на рис. 3 показує хороші результати щодо визначення кількості користувачів, які переглянуть новину, але показує надто швидке зростання кількості переглядів в середині розглянутого періоду, що не відповідає дійсності.

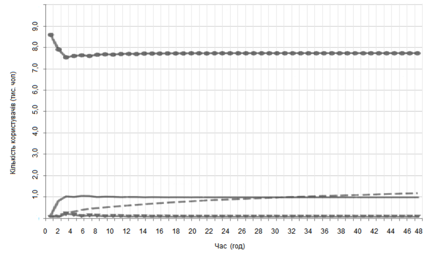

Рис. 4. DK-модель поширення повідомлень соціальною мережею.

Модель клітинного автомату рис. 4 досить точно оцінює кінцеву кількість користувачів, що прочитають новину, але не відповідають реальному стану речей на початку часу поширення новини.

Висновки

Вивчення процесу поширення інформації є актуальним в багатьох галузях, від маркетингу до безпеки суспільства і держави. Правильна оцінка характеристик, притаманних поширенню інформації, допомагає передбачувати кількість кінцевих користувачів деякого інформаційного повідомлення. З цією метою в роботі було розглянуто математичні моделі поширення інформації, що є простими для використання та реалізації, але достатньо ефективними для отримання шуканого результату.

Таким чином, розширена модель SIR дає змогу оцінити охват аудиторії, але не відповідає реальності стосовно відображення динаміки поширення новини у часі, а також наявність параметра одужання робить використання моделі складнішим, адже оцінка швидкості переходу новини у неактуальний стан є комплексним показником та потребує експертної оцінки.

DK-модель має такий же недолік, і, хоча кількість залученої аудиторії за використання моделі оцінюється достатньо точно, через деякий час після публікації новини модель показує значний стрибок у кількості переглядів новини, що не відповідає дійсності.

Модель клітинного автомата за умови її ускладнення дає змогу оцінювати не тільки кількість споживачів новини, але й канали її поширення, але значним її недоліком є невідповідність у відображенні динаміки поширення новини у часі, а також значний час та складність обчислювальних кроків.

Для покращення результатів роботи моделей необхідна їх перевірка на більшій кількості користувачів та різних типах новин для їх подальшого доопрацювання та уточнення параметрів.

Перелік використаних джерел

1. Д.К. Горковенко. Обзор моделей распространения информации в социальных сетях // Молодой ученый. — 2017. — № 8(142). — С. 23–28.
2. W.O. Kermack, A.G. McKendrick. A contribution to the mathematical theory of epidemics // Proc. of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 1927. — no. 115.
3. H.W. Hethcote. The mathematics of infectious diseases // SIAM Review. — 2000. — no. 42(4).
4. D.J. Daley, D.G. Kendall. Stochastic rumors // J. Inst. Math. Appl. — 1965. — Vol. 142. — P. 42–55.
5. A. Montanari, A. Saberi. The spread of innovations in social networks. // Proc Nat Acad Sci USA. — 2010. — no. 107(12).
6. В. Носова М., И. Сенникова Л. Моделирование распространения информации в децентрализованных сетевых системах с нерегулярной структурой // Новые информационные технологии в автоматизированных системах. — 2014. — № 17. — С. 8–15.
7. C. Castellano, S. Fortunato, V. Loreto. Statistical physics of social dynamics // Rev Mod Phys. — 2009. — Vol. 81. — P. 591–646.