

ДОСЛІДЖЕННЯ ІНФОРМАЦІЙНИХ ПОТОКІВ, ЯК ДИНАМІЧНИХ МУЛЬТИАГЕНТНИХ СИСТЕМ

А. М. Грайворонська^{1, а}, Д. В. Ланде²

¹Національний технічний університет України «Київський політехнічний інститут»

²Інститут проблем реєстрації інформації НАН України

Анотація

В сучасному інформаційному просторі публікується безліч інформаційних документів, і інтерес представляють закономірності формування та розвитку сукупності документів, що відповідають певній тематиці. В статті запропонована модель тематичного інформаційного потоку на основі агентів, які відповідають опублікованим документам. В результаті моделювання отримані часові ряди, які відображають зміни кількості агентів в системі і дослідженні їх властивості.

Ключові слова: інформаційний потік, мультиагентна модель, часовий ряд

Вступ

Сучасний інформацій простір є динамічною системою, яка складається з інформаційних елементів (документів) пов'язаних між собою за тематикою [1]. Багато інформаційних ресурсів, які публікують інформацію он-лайн, мають потокову структуру [2]. Прикладами таких ресурсів є соціальні мережі, стрічки новин, блоги. Публікацію наукових статей також можна розглядати як інформаційний потік.

Основним об'єктом моделювання є тематичний переріз інформаційного потоку – тобто послідовність документів, які відповідають певній тематиці. Такому тематичному інформаційному потоку ставлять у відповідність часовий ряд, який складається з кількості повідомлень або документів опублікованих за певний проміжок часу. Аналіз часових рядів є фундаментальною задачею математичної статистики. В сучасних дослідженнях також використовують методи дисперсійного, фрактального та вейвлет-аналізу. Інтерес представляють шаблони поведінки, виявлення аномалій та трендів, також актуальною є задача прогнозування [3].

На динаміку реального інформаційного потоку впливає безліч факторів. На невеликих часових проміжках зміну кількості документів за тематикою можна описувати простими лінійною або експоненціальною моделями. Проте більш актуальним є дослідження динаміки інформаційного потоку від появи нової тематики до його "старіння". Прикладом моделі зміни актуальності тематики є модель дифузії інформації, яка заснована на концепції клітинних автоматів [1]. Ще один сучасний підхід, який можна застосовувати для досліджень інформаційного простору є мультиагентні моделі, які будуть розглянуті далі в статті.

1. Мультиагентні моделі

В мультиагентних моделях агент – це абстрактна сутність, для якої характерна автономна поведінка на основі правил. Принциповою особливістю моделей на основі агентів є децентралізованість. В таких моделях визначаються правила поведінки окремих агентів, їх локальна взаємодія і взаємодія із середовищем. Такий підхід є зручним в імітаційних моделях, коли глобальна динаміка системи дуже складна або недостатньо вивчена для безпосереднього моделювання.

В роботі [4] описані такі особливості мультиагентних моделей:

- Різноманітність (heterogeneity). Індивідуальність агентів може відрізнятися певним чином. Наприклад правилами прийняття рішень, місцем положення, пам'яттю, накопиченими ресурсами, генетично і т.п.
- Автономність (autonomy). В мультиагентних моделях немає централізованого контролю індивідуальної поведінки.
- Визначений простір (explicit space). Еволюція моделі відбувається у певному просторі, де визначене поняття локального околу.
- Локальна взаємодія (local interactions). За звичай взаємодія відбувається між сусідніми (в сенсі простору) агентами.
- Обмежена раціональність (bounded rationality). Агенти використовують лише локальну інформацію і діють на основі простих правил поведінки.
- Нерівноважна динаміка (non-equilibrium dynamics).

2. Модель інформаційного потоку на основі агентів

В роботі представлена мультиагентна модель, у якій агенти відповідають інформаційним повідом-

^аnastya_graiv@ukr.net

лення (документам) з певною тематикою. В початковий момент часу ($t = 0$) в системі існує лише один агент. Далі в процесі еволюції системи агенти можуть народжуватися і помирати, що відповідає появі повідомлення в мережі та його видаленню або втраті актуальності. Події, які відбуваються з агентом протягом його існування, відповідають основним операціям із записами в соціальних мережах: like, dislike, repost. Далі розглянемо докладно правила, за якими відбувається еволюція системи.

Будемо розглядати модель з дискретним часом. Нові агенти можуть з'являтися в системі двома способами. По-перше в кожен момент часу існує ймовірність p_s самонародження агента. Це відповідає появі повідомлення із зовнішнього світу, наприклад публікації людиною нової статті. У випадку самонародження агент може дати посилання на існуючого агента. Другий спосіб появи агента – це копіювання існуючого, тобто операція repost.

Кожен агент народжується із початковою енергією E_0 . З кожним часовим відліком енергія агента зменшується на 1. Якщо значення енергії стало менше за 0, то агент помирає. Події, які відбуваються з агентом, змінюють значення енергії за наступними правилами:

- якщо поставили like, то енергія збільшується на 1;
- якщо поставили dislike, то енергія зменшується на 1;
- якщо зробили repost, то енергія збільшується на 2;
- якщо інший агент зробив посилання на даного агента, то енергія збільшується на 1.

Ймовірність того, що відбудеться певна подія, залежить від енергії агента. Нехай в n -ий момент часу енергія рівна E_n , тоді:

$$\begin{aligned} P\{\text{при } t = n \text{ поставили like}\} &\equiv P_{like}^n = P_{like_0} \varphi(E_n), \\ P\{\text{при } t = n \text{ - dislike}\} &\equiv P_{dislike}^n = P_{dislike_0} \varphi(E_n), \\ P\{\text{при } t = n \text{ зробили repost}\} &\equiv P_{repost}^n = P_{repost_0} \varphi(E_n) \end{aligned}$$

$$\text{де } \varphi(E) = \begin{cases} \frac{E}{2(E_0 - E)}, & E \in [0; \frac{4E_0}{3}]; \\ 1, & E > \frac{4E_0}{3} \end{cases};$$

Величини P_{like_0} , $P_{dislike_0}$, P_{repost_0} є параметрами моделі. Функція φ – монотонна зростаюча, тому ймовірність будь-якої події є більшою для агента з більшим значенням енергії.

3. Результати моделювання

Для візуалізації роботи моделі створюється спеціальний граф. Рис. 2 є ілюстрацією основних позначень.

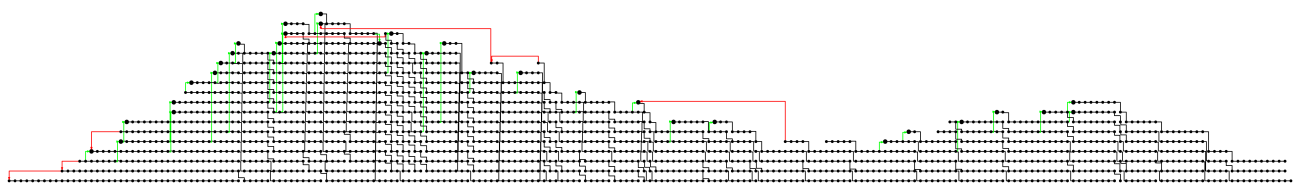


Рис. 1. Візуалізація еволюції системи агентів у вигляді графу

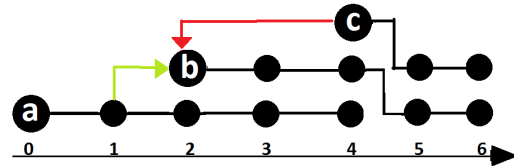


Рис. 2. Приклад візуалізації системи у вигляді графу з трьома агентами (a, b, c)

Внизу показана часова вісь, за якою видно, що в при $t = 0$ є один агент **a**. Агент **a** існує у проміжок часу від 0 до 4, що відображено його лінією життя. Також на Рис. 2 зображені агенти **b** і **c**, які народились при t рівне 2 і 4 відповідно, і померли в момент часу 6. Зеленою стрілкою показано, що в момент часу 1 зробили repost агента **a**, отже агент **b** є його копією. Агент **c** самонародився, і при цьому зробив посилання на агента **b**, що зображено червоною стрілкою.

Приклад роботи моделі з параметрами $P_{like_0} = 0.4$, $P_{dislike_0} = 0.05$, $P_{repost_0} = 0.1$ і $E_0 = 35$ зображений на Рис. 1.

У вступі було зауважено, що основним об'єктом дослідження є часовий ряд, що відповідає тематичному інформаційному потоку.

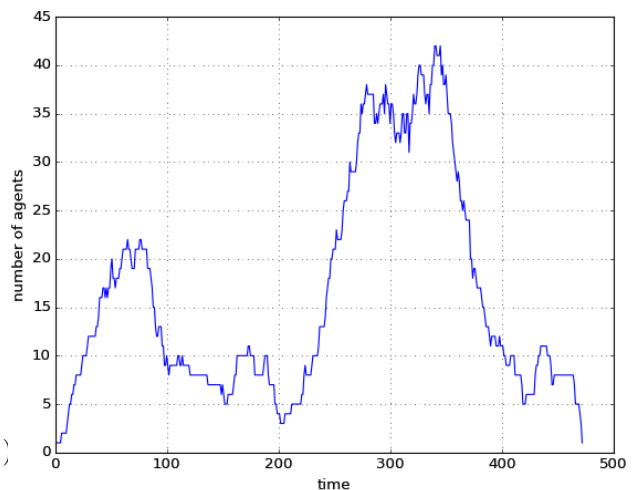


Рис. 3. Результат моделювання – часовий ряд

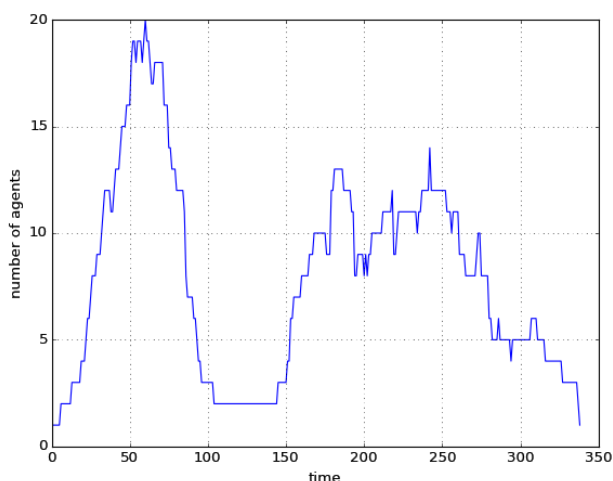


Рис. 4. Результат моделювання – графік залежності кількості агентів в системі від часу

У роботі [3] обґрунтовується, що для таких часових рядів характерні два основних локальних екстремуми, тобто два піки популярності тематики. За допомогою даної мультиагентної моделі вдається отримати часові ряди, які відповідають описаним характеристикам (Рис. 4, Рис. 3).

Реальним часовим рядам, які відповідають тематичним інформаційним потокам, притаманна хаотична поведінка. Поширеним методом аналізу самоподібності часових рядів є R/S -аналіз [1]. За допомогою цього методу обчислюється показник Херста (H), який характеризує різні типи кореляційної динаміки. Значення $H < 0.5$ відповідає некорельованій поведінці значень ряду, а значення $0.5 < H < 1$ визначають рівень автокореляції ряду. Показник R/S для заданого часового ряду $\{X(t), t = 1, 2, \dots, N\}$ визначають через стандартне відхилення (S):

$$S = \sqrt{\frac{1}{N} \sum_{t=1}^N (X(t) - \langle X \rangle_N)^2},$$

де $\langle X \rangle_N = \frac{1}{N} \sum_{t=1}^N X(t)$,

і розмах (R):

$$R = \max_{1 \leq t \leq N} X(t, N) - \min_{1 \leq t \leq N} X(t, N),$$

де $X(t, N) = \sum_{i=1}^t (X(i) - \langle X \rangle_N)$.

Для багатьох часових рядів справедливе наступне співвідношення:

$$R/S = \left(\frac{N}{2}\right)^H$$

Для часових рядів, які були отримані в результаті моделювання, показник Херста має значення в межах $[0.82 - 0.96]$. Такі значення відповідають високому ступеню автокореляцій в ряді.

Висновки

Для тематичних інформаційних потоків характерна складна динаміка, яка залежить від багатьох факторів. Для моделювання еволюції таких потоків застосований мультиагентний підхід. При цьому кожному агенту відповідає документ і зміна стану системи відбувається відповідно до локальних правил поведінки агентів. Розглянута модель також є ймовірнісною, тому дозволяє отримати різні типи еволюції інформаційного потоку.

В результаті моделювання отримані часові ряди, які відповідають кількості агентів в різні моменти часу. В багатьох випадках для отриманих рядів характерна наявність двох чітко виділених локальних максимумів. Також виявлений високий ступінь автокорельованості ряду. Вказані особливості також характерні для багатьох реальних інформаційних потоків.

Перелік використаних джерел

1. Горбулін В.П., Додонов О.Г., Ланде Д.В. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання: монографія. — К.: Інтертехнологія, 2009. — 164 с.
2. Kleinberg J. Temporal dynamics of on-line information streams // Data Stream Management: Processing High-Speed Data Streams. — Springer, 2006. — 18 p.
3. Додонов О.Г., Ланде Д.В. Мультиагентная модель поведения тематических информационных потоков — Материалы 6-ой Всероссийской мультиконференции — 2013. — Т. 4, № 6. — с. 102–107.
4. J. M. Epstein Remarks on the foundations of agent-based generative social science — Handbook on Computational Economics — North Holland Press, 2005. — V. II.