

УДК 51-7

БАРАНОВСЬКИЙ Олексій Миколайович

КАЧИНСЬКИЙ Анатолій Броніславович

ДОСЛІДЖЕННЯ ЧАСОВОЇ ЗАЛЕЖНОСТІ ІНФОРМАЦІЙНИХ ПОТОКІВ МЕТОДОМ R/S АНАЛІЗУ

Постановка проблеми. Функціонування та розвиток суспільства відбуваються у різних формах соціальних процесів. Своєчасна й об'єктивна інформація про різноманітні форми соціальної активності дає змогу прогнозувати можливі наслідки прийняття важливих рішень, особливо в умовах швидких динамічних політичних змін у країні.

Однією з підстав успішного моделювання соціальних процесів і прогнозування результатів певних соціальних процедур є урахування взаємозв'язку подій з інформаційним середовищем, зокрема його найбільш динамічною та сучасною частиною – множиною інформаційних ресурсів мережі Інтернет. Завдання вивчення властивостей інформаційного простору, зокрема інформаційних потоків, – багатопланове, припускає активне використання методів, що дозволяють глибше зрозуміти природу соціальних явищ, специфіку предметної галузі [1].

Аналіз останніх досліджень та публікацій. Питанням дослідження й моделювання інформаційного простору мережі Інтернет приділяється увага у працях Д.В.Ланде, О.Г.Додонова, В.М.Фурашева. Зокрема приклади використання методів фрактального аналізу для дослідження характеристик новинних інформаційних потоків наведені в роботах [2; 3].

При цьому в абсолютній більшості робіт лише підтверджується факт наявності самоподібності та довгострокової залежності в інформаційних потоках без дослідження характеристик і причин виникнення цієї залежності. Водночас кількість аналітичних моделей інформаційних потоків на основі фрактальних випадкових процесів досі є невеликою, тому це питання потребує додаткових досліджень.

Метою статті є виявлення характеристик, які можуть бути використані для розроблення аналітичної моделі інформаційних потоків на основі фрактальних випадкових процесів.

Виклад основного матеріалу. Для досягнення поставленої мети вирішуються такі завдання:

- дослідження інформаційного потоку на наявність неперіодичних циклів, довгострокової та короткострокової залежності;
- дослідження динаміки реалізації інформаційного потоку;
- видалення, в разі необхідності, короткострокової залежності та аналіз динаміки залишку.

Як об'єкт дослідження в роботі використовується реалізація інформаційного потоку, який характеризує прояв громадської активності “Демонстрації”. Під інформаційним потоком розуміється щоденна кількість

повідомлень електронних засобів масової інформації, які були зареєстровані системою контент-моніторингу в період 01.01.2010 – 31.12.2011 та відповідають визначеній вербальній моделі [4].

Щодо вибору та обґрунтування методу, то стохастичний процес є статистично самоподібним із параметром, якщо для довільного дійсного $\alpha > 0$ процес $\alpha^{-H} X(t)$ має ті самі статистичні характеристики, що й сам процес X [5]. Параметр $0 < H < 1$ називається показником Херста та характеризує ступінь самоподібності.

Довгострокова залежність означає повільне (гіперболічне) спадання в часі автокореляційної функції $r(k)$ самоподібного випадкового процесу: $r(k) \sim |k|^{-\beta}$ при $|k| \rightarrow \infty$, де $0 < \beta < 1$. Параметр β пов'язаний із показником Херста співвідношенням $H = 1 - \beta$.

Традиційно самоподібність та довгострокова залежність реалізації розпізнається шляхом визначення показника Херста. Звичайно той факт, що $0,5 < H < 1$ вважається достатньою основою для визнання процесу самоподібним. Одним із найбільш використовуваних методів оцінки показника Херста є R/S-аналіз (метод нормованого розмаху). При дослідженні R/S-статистики часового ряду X довжиною τ визначається відношення $R(\tau)/S(\tau)$, де R_1 – розмах кумулятивного ряду $x^{\text{cum}}(t)$, S_1 – середнє квадратичне відхилення ряду:

$$R/S = \frac{R(\tau)}{S(\tau)} = \frac{\max(x^{\text{cum}}(t, \tau)) - \min(x^{\text{cum}}(t, \tau))}{\sqrt{\frac{1}{\tau-1} \sum_{t=1}^{\tau} (x(t) - \bar{x})^2}} \quad (1)$$

$$t = \overline{1, \tau}$$

$$x^{\text{cum}}(t, \tau) = \sum_{i=1}^t x(i) - \bar{x}(\tau)$$

$$\bar{x}(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} x(t)$$

Для самоподібного процесу відношення $R(\tau)/S(\tau)$ при великих значеннях τ має таку характеристику:

$$M\left[\frac{R}{S}\right] = (c\tau)^H, \quad (2)$$

де c – деяка постійна величина.

Якщо прологорифмувати обидві частини цього виразу, отримаємо:

$$\log\left(M\left[\frac{R}{S}\right]\right) = H \log c + H \log \tau \quad (3)$$

Показник H обчислюється як тангенс куту нахилу лінійної залежності $\log\frac{R(\tau)}{S(\tau)}$ від $\log\tau$.

У випадку $0,5 < H < 1$ роблять висновок про персистентну поведінку процесу або про те, що процес має довгу пам'ять. Тобто, якщо упродовж деякого часу в минулому спостерігався додатний приріст процесу, тобто виконувалось збільшення, то і далі в середньому буде проходити збільшення. У випадку $0 < H < 0,5$ кажуть про антиперсистентний процес. Тут великі значення процесу йдуть за низькими та навпаки. При $H = 0,5$ відхилення процесу від середнього є справді випадковими та не залежать від попередніх значень.

Звичайно по одному часовому ряду отримують *точкову оцінку* показника Херста – число, яке характеризує ступінь довгострокової залежності: чим ближче параметр H до 1, тим повільніше при збільшенні часової затримки k спадають кореляції між пакетами $r(k) = \frac{1}{k^{2(1-H)}}$.

У роботі для аналізу типу часової залежності застосовується *модифікована методика*, розглянута в [6]. У цьому випадку показник Херста розглядається як функція кількості відліків $H(\tau) = f\left[\log\frac{R}{S}(\tau)\right]$. Поведінка функції $H(\tau)$ дає змогу визначити такі характерні особливості часового ряду:

- незалежність випадкових даних;
- середню довжину неперіодичного циклу;
- довгострокову та короткострокову залежність.

За допомогою R/S-аналізу можливо виявити циклічність процесу та визначити середню довжину неперіодичних циклів, характерних для хаотичних систем. Оскільки в цьому випадку динаміка системи обмежена аттрактором (періодичним або хаотичним),

починаючи з визначеного періоду значення показника H (нахил кривої $\log \frac{R(\tau)}{S(\tau)}$) перестають змінюватись. Цей період характеризує середню довжину хаотичного циклу.

R/S-аналіз дозволяє визначити та ліквідувати короткострокову залежність, яка характерна для авторегресійних процесів. Лінійна залежність зміщує криву показника Херста та демонструє довгострокову пам'ять. Для видалення короткострокової залежності необхідно ряд приростів реалізації процесу $S(\tau)$ регресувати як залежну змінну проти $S(\tau - 1)$ та знайти лінійну залежність між ними. Після цього проводиться R/S-аналіз залишку $X(t) = S(t) - (a + b \cdot S(t - s))$. Якщо вихідний ряд мав довгострокову залежність, то вона зберігається, натомість короткострокова залежність видаляється.

Реалізація інформаційного потоку в мережі Інтернет, що відповідає типу громадської активності "Демонстрації", має особливу структуру, що зберігається на багатьох масштабах – в реалізації завжди присутня деяка кількість дуже великих викидів при відносно невеликому середньому рівні показника (рис. 1, рис. 2).

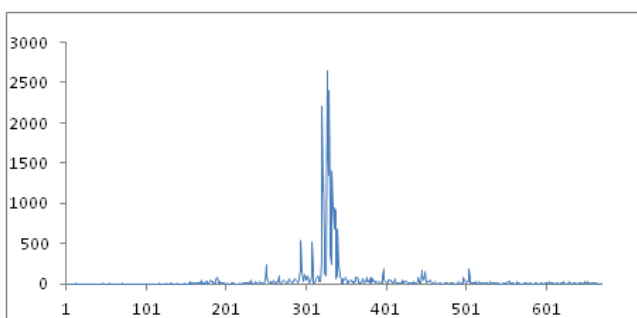


Рис. 1. Реалізація інформаційного потоку, що відповідає виду активності "Демонстрації" за весь період дослідження

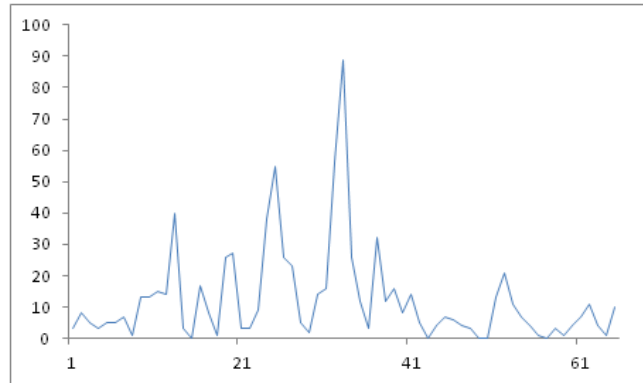


Рис. 2. Фрагмент реалізації інформаційного потоку, що відповідає виду активності "Демонстрації"

На рис. 3 показана автокореляційна функція реалізації, представлена на рис. 1.

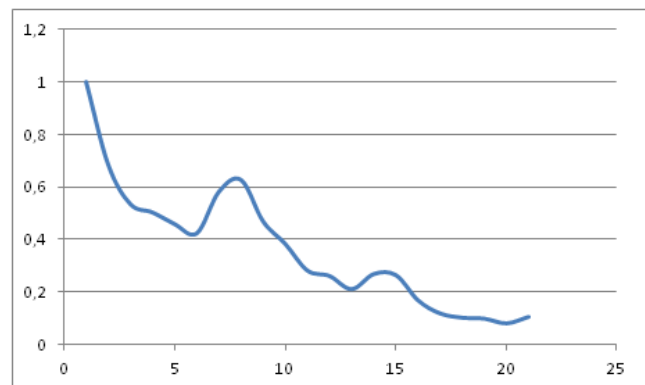


Рис. 3. Кореляційна функція інформаційного потоку, що відповідає виду активності "Демонстрації"

Характерні сплески на графіку кореляційної функції інформаційного потоку дають змогу дійти висновку про наявність короткострокової залежності.

На рис. 4 показано визначення показника Херста для досліджуваного інформаційного потоку. Обчислене значення складає 0,96, що підтверджує наявність довгострокової залежності.

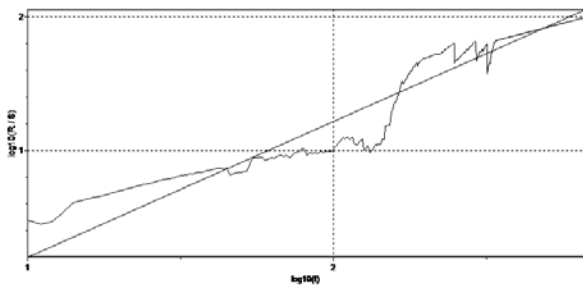


Рис. 4. Залежність $\log \frac{R(t)}{S(t)}$ від $\log t$ для інформаційного потоку, що відповідає виду активності “Демонстрації”

Візуальний аналіз рис. 4 дозволяє простежити наявність нерегулярного циклу.

У реалізації була виявлена сильна авторегресійна залежність. Після видалення короткострокової залежності згідно з обгрунтованою методикою показник Херста зменшився (0,8), залишаючись при цьому більшим 0,5. На рис. 5 представлено графічне відображення визначення показника Херста з видаленою короткостроковою залежністю.

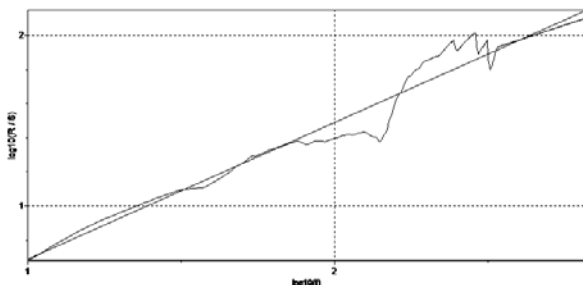


Рис. 5. Залежність $\log \frac{R(t)}{S(t)}$ від $\log t$ для інформаційного потоку, що відповідає виду активності “Демонстрації” з видаленою короткостроковою залежністю

Отже, реалізація інформаційного потоку має істинну довгострокову залежність.

Висновки. Наукова новизна роботи полягає у тому, що як аналітична модель інформаційних потоків, які характеризують соціальні процеси, а саме прояви громадської активності, пропонується використовувати стохастичні процеси, що є сумою процесів із довгостроковою та короткостроковою залежністю, тобто самоподібного процесу й процесу авторегресії. Параметрами цих моделей є коефіцієнт авторегресії та показник Херста істинної довгострокової залежності.

Вибір моделі може здійснюватись залежно від мети дослідження. Наприклад, для короткострокового прогнозування в окремих випадках доцільно використовувати моделі авторегресії. При довгостроковому прогнозуванні основними математичними моделями соціальних процесів стають випадкові процеси з довгостроковою залежністю.

Питання причини виникнення авторегресійних характеристик та факторів, що визначають фрактальну поведінку інформаційних потоків, що відображають певні соціальні процеси, є завданням подальших досліджень.

Список використаних джерел

1. Додонов О.Г. Державна інформаційна політика і становлення інформаційного суспільства в Україні / О.Г.Додонов, О.С.Горбачик, М.Г.Кузнєцова // Стратегічна панорама. – 2002. – № 1. – С. 166 – 170.

2. Горбулін В.П. Інформаційні операції та безпека суспільства: загрози, протидія, моделювання : моногр. / В.П.Горбулін, О.Г.Додонов, Д.В.Ланде. – К. : Інтертехнологія, 2009 – 164 с.

3. Моделирование динамики новостных текстовых потоков / Д.В.Ландэ, А.А.Снарский,

С.М.Брайчевский, А.Т.Дармохвал // Интернет – математика 2007: сборник работ участников конкурса. – 2007. – С. 98–107.

4. Прикладні аспекти застосування моделей взаємозв'язку “ЗМІ-соціум – державна політика” на прикладі реформування податкової системи України / О.М.Барановський, Є.Л.Добровольський, А.Б.Качинський, Д.В.Ланде // Інформаційна безпека людини, суспільства, держави. – 2011. – № 2 (2) – С. 140–146.

5. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / Э.Петерс. – М. : Мир, 2000. – 277 с.

6. Кириченко Л.О. Исследование долгосрочной зависимости сетевого трафика методом R/S-анализа / Л.О.Кириченко, Т.А.Радивилова // Автоматизированные системы управления и приборы автоматики. – 2006. – Вып. 135. – С. 51–55.

Аннотація: В роботі представлені результати численного анализа інформаційних потоків мережі Інтернет, виконаного з метою визначення наявності і типів часової залежності в них. Дослідження, проведені з допомогою R/S аналізу, показали, що інформаційні потоки мають властивості як довгострокової, так і короткострокової пам'яті, що дозволяє представити потоки в вигляді сумми реалізацій самоподобного процесу і процесу авторегресії.

Ключевые слова: інформаційні потоки, фрактальний аналіз, громадянська активність, показател ь Херста.

Abstract: This article is about the quantitative analysis of Internet information flows performed to determine the types of temporal dependence. With results of R/S analysis we can prove that information flows have long-term and short-term memory, that allows to image information flows as the sum of implementation of self-similar process and autoregressive process.

Key words: information flows, fractal analysis, civic engagement, Hurst exponent.