

КОМП'ЮТЕРНІ ЗАСОБИ, МЕРЕЖІ ТА СИСТЕМИ

V. Pisarenko, J. Pisarenko,
O. Gavrylyuk, A. Boiko

ABOUT DYNAMICS OF INFORMATION FLOWS

The current issues of identification, the reasons for the emergence and study of the spreading dynamics of the information operations are discussed.

Key words: mathematical model, information flows, distribution, dynamics.

Рассмотрены актуальные вопросы идентификации, выявления причин возникновения и изучения динамики распространения информационных операций.

Ключевые слова: математическая модель, информационные потоки, распространение, динамика.

Розглянуто актуальні питання ідентифікації, виявлення причин виникнення і вивчення динаміки поширення інформаційних операцій.

Ключові слова: математична модель, інформаційні потоки, поширення, динаміка.

© В.Г. Писаренко, Ю.В. Писаренко,
О.М. Гаврилюк, А.Г. Бойко,
2018

УДК 004.942

В.Г. ПИСАРЕНКО, Ю.В. ПИСАРЕНКО,
О.М. ГАВРИЛЮК, А.Г. БОЙКО

ПРО ДИНАМІКУ ІНФОРМАЦІЙНИХ ПОТОКІВ

Інформаційний простір формується сьогодні не тільки вербально. Велику роль в цьому відіграють інформаційні технології. Вони стали силою, але водночас і точкою вразливості сучасних держав, які істотно залежать у своєму функціонуванні від інформаційних технологій, оскільки за допомогою них ведеться не тільки накопичення і обробка інформації, а й управління і виробництвом, і військовими діями. Сучасний стан систем масових комунікацій надає можливість доведення інформаційних повідомлень у різних форматах практично до будь-якої людини. Це створює умови для ефективного проведення інформаційних операцій через різні канали поширення інформації. Інформаційна операція – це взаємопов'язана послідовність інформаційних впливів для досягнення поставленої мети. Своєї ефективності інформаційна операція (ІО) досягає у разі максимального охоплення цільової аудиторії [1]. Основним інструментом для проведення ІО є соціальні засоби масової інформації (ЗМІ). Отже, питання ідентифікації ІО, виявлення причини їх виникнення та вивчення динаміки поширення є актуальними. Вивчення наукової літератури та відкритих джерел дозволило виділити два основних підходи до виявлення ІО: класичний та на основі динаміки інформаційних потоків.

Класичний підхід. Ідея підходу полягає у тому, що проводиться оцінка кількості повідомлень негативної тональності в рівні проміжки часу. Якщо їх кількість вища за порогове значення, робиться висновок про проведення інформаційної операції [2, 3]. Переваги та недоліки:

- точність автоматичного визначення тональності повідомлень українською мовою складає близько 70 %;
- новинні повідомлення нейтральної та позитивної тональності також можуть бути частиною інформаційної операції (наприклад, недостовірні повідомлення, негативний контекст події, освітленої у ЗМІ);
- відсутність обґрунтування порогового значення;
- виявлення певної кількості ІО проходить у час їх проведення.

Підхід на основі динаміки інформаційних потоків. Ідея підходу полягає у побудові графіка зміни інтенсивності (кількість в одиницю часу) і порівнюється з заданим шаблоном (рис. 1). У разі їх збігу робиться висновок про здійснення ІО.

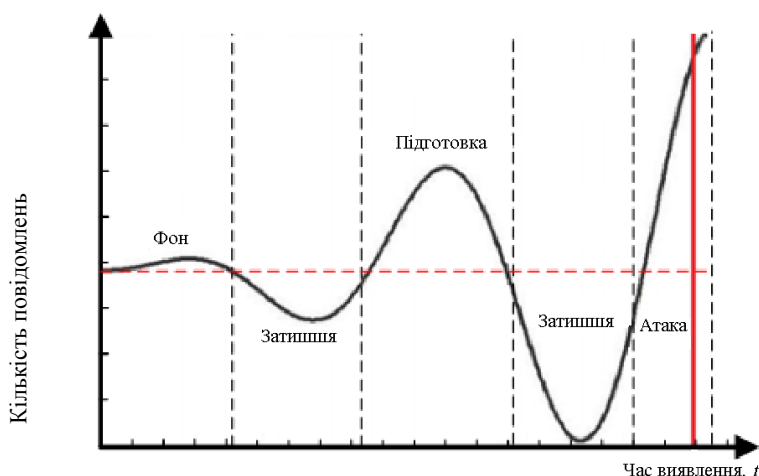


РИС. 1. Динаміка кількості тематичних повідомлень у час ІО

Переваги та недоліки:

- наявність сплесків графіка може говорити не тільки про штучну появу публікацій (наявності ІО), а й про подальший розвиток якоїсь події і появу нової інформації і нових повідомлень;
- простота реалізації (групування за сюжетами, побудова графіка інтенсивності повідомлень і порівняння його з шаблонами).

Під своєчасним виявленням ІО розуміється здатність виявляти її до моменту початку пасивної фази ІО. При своєчасному виявленні ІО залишається можливість ефективної протидії.

На рис. 2 надано загальні закономірності проведення ІО.

В час активної фази публікуються схожі повідомлення в різних, частіш за все маловідомих, ЗМІ. Оскільки повідомлення з'являються впродовж невеликого проміжку часу, системи моніторингу ЗМІ сприймають їх як повідомлення про нову, «популярну» подію.

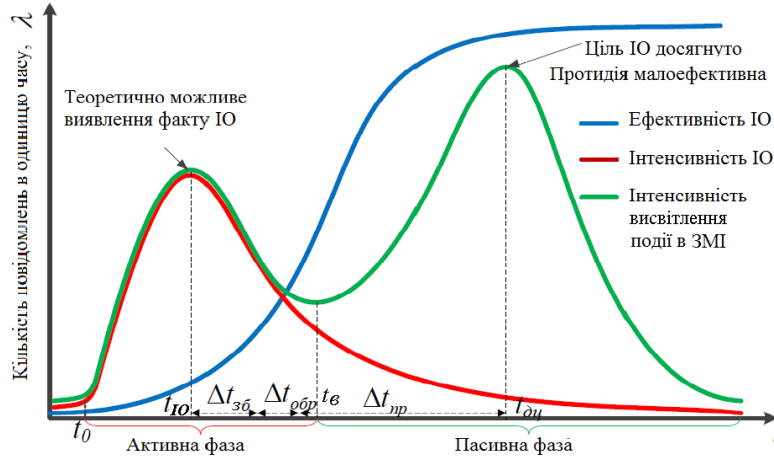


РИС. 2. Узагальнені закономірності здійснення ІО: t_0 – час початку ІО; t_{IO} – час максимальної інтенсивності у активній фазі ІО; $\Delta t_{зб}$ – інтервал часу збору повідомлень ЗМІ; $\Delta t_{обр}$ – інтервал часу обробки повідомлень ЗМІ; $\Delta t_{пр}$ – інтервал часу протидії ІО; $t_{доц}$ – час досягнення цілі ІО; t_ϵ – час виявлення ІО (неправдивої новини); λ – інтенсивність появи повідомлень

В час пасивної фази частина інших ЗМІ реагує на повідомлення активної фази ІО. Тим самим аудиторія збільшується, ефективність ІО підвищується.

Теоретично можливе виявлення ІО у момент максимальної інтенсивності активної фази ІО t_{IO} , однак необхідно враховувати час збору $\Delta t_{зб}$ та $\Delta t_{обр}$ повідомлень ЗМІ:

$$t_\epsilon = t_{IO} + \Delta t_{зб} + \Delta t_{обр}, \quad (1)$$

де t_ϵ – час виконання ІО.

Система нерівностей визначає необхідні умови забезпечення своєчасного виявлення ІО:

$$\begin{cases} t_\epsilon < t \\ \Delta t_{обр} < \frac{1}{\lambda_{max}} \end{cases}, \quad (2)$$

де $\Delta t_{обр}$ – середній час обробки одного повідомлення, що містить інформацію; λ_{max} – максимальна інтенсивність виявлення повідомлень. Згідно з (2), виявлення ІО має бути до моменту досягнення її цілі, коли інтенсивність появи нових повідомлень під час пасивної фази ІО максимальна. Після цього моменту виявляти ІО для протидії їй не є доцільним, оскільки не є ефективним. Окрім того, необхідно забезпечити швидкість обробки повідомлень не менше за максималь-

ну інтенсивність їх надходження. Система (2) не є достатньою умовою ефективного використання ІО, оскільки до моменту початку пасивної фази ІО необхідно прийняти рішення про можливу протидію, виробити заходи та привести їх у дію. Однак (2) – необхідна умова виявлення ІО для ефективної протидії.

Для створення умов оптимальної протидії ІО необхідно забезпечити:

$$\max \Delta t_{ep} \begin{cases} \Delta t_{np} = t_{\partial u} - t_e \\ t_e < t_{\partial u} \\ \Delta t_{обp} < \frac{1}{\lambda_{max}} \end{cases} . \quad (3)$$

Для постановки задачі треба обрати інформативні ознаки та алгоритм класифікації для виявлення ІО до початку пасивної фази з необхідною точністю. Оскільки активну фазу ІО складають повідомлення соціальних ЗМІ, поширення яких носить координований цілеспрямований характер, необхідно визначити інформативні ознаки для сукупності повідомлень і зв'язків між ними, як одного образу. У загальному випадку, такі повідомлення мають такі особливості:

- високий ступінь подібності текстів, що не є повними дублікатами. Такі тексти називаються нечіткими дублікатами [4, 5];
- невеликий інтервал часу, в який відбувається опублікування новин. Це дозволяє імітувати появу нової значущої події;
- джерела, що опублікували їх, є маловідомими, регіональними або спеціалізованими ЗМІ. Поява повідомлень у великих інформаційних агентствах можливо тільки у час пасивної фази.

Приклад структури поширення повідомлень за однією темою. Множина образів ІО, що розпізнаються:

- Ω_1 – клас образів сукупності повідомлень ЗМІ та зв'язків між ними, що складають активну фазу ІО;
- Ω_2 – клас образів сукупності повідомлень ЗМІ та зв'язків між ними, що не є частиною ІО;
- образ ω_{e_k} надається у вигляді зваженого орієнтованого графу розповсюдження повідомлень за темою e_k :

$$\omega_{e_k} = (\{news\}_{e_k}, H_{e_k}), \quad (4)$$

де $\{news\}_{e_k}$ – повідомлення ЗМІ мережі Інтернет за повідомленням e_k ;

H_{e_k} – матриця зв'язності, що визначає напрямки ребер між вузлами графу.

Новинне повідомлення включає у себе повний текст, час опублікування, джерело ЗМІ та значення тематичного класифікатора:

$$news_k = \langle c_k, t_k, src_j, \langle n_{k_1}, n_{k_2}, \dots, n_{k_0} \rangle \rangle, \quad (5)$$

$$n_{k_a} = \begin{cases} 1, & \text{якщо } c_k \in e_a \\ 0, & \text{якщо } c_k \notin e_a \end{cases}, \quad (6)$$

де c_k – текст новини $news_k$; t_k – дата і час появи повідомлення $news_k$; src_j – джерело ЗМІ, в якому було опубліковано повідомлення; n_{k_a} – індикатор відповідності тексту повідомлення $news_k$ темі e_a .

Текст новинного повідомлення надано у вигляді послідовності слів (що відповідають векторному представленню тексту):

$$c_k = \langle s_{k,1}, s_{k,2}, \dots, s_{k,n} \rangle, \quad (7)$$

де $s_{k,i}$ – i -е слово в тексті c_k .

Для розрахунку ваг ребер $h(news_k, news_i)$ між повідомленнями $news_k$ та $news_i$ необхідно враховувати ступінь подібності текстів $d(c_k, c_i)$ та час опублікування повідомлень:

$$h(news_k, news_i) = f(d(c_k, c_i), t_i, t_k). \quad (8)$$

Ваги вузлів графа поширення повідомлень відповідають ваговим коефіцієнтам джерел ЗМІ, що опублікували їх. Ваги джерел мають враховувати тип ЗМІ. Для інформаційних агентств великих державних ЗМІ, які забезпечують достовірність і неупередженість інформації, ваговий коефіцієнт має бути максимальним. Граф поширення інформації з певної теми може складатися з декількох підграфів, які не мають шляхів один між одним. Такі підграфи в орієнтованому графі називаються компонентами з слабкою зв'язністю (КСС). Наявність декількох КСС в структурі поширення повідомлень говорить про різні описи або інтерпретації однієї і тієї ж події. Це можливо і в разі проведення ІО. Отже, образам для розпізнавання будуть самі КСС графа поширення повідомлень. Виділення КСС в графі реалізується за допомогою алгоритму пошуку в глибину [6] та має поліноміальну складність. Априорний словник ознак $x(\omega_r)$ образів поширення повідомлень з певної теми включає у себе параметри графів [7], частину з яких згенеровано евристично.

Формування робочого словника ознак $x^q(\omega_r)$ проводиться на основі априорного словника ознак $x(\omega_r)$ і вибірки структур поширення повідомлень із заданою належністю до класів Ω_1 і Ω_2 . Робочий словник ознак $x^q(\omega_r)$ включає інформативні ознаки для розпізнавання ІО у повідомленнях ЗМІ. При цьому вид функції інформативності та значення елементів платіжної матриці C залежить від критерію ефективності розв'язуваної задачі. Якщо метою є досягнення максимальної повноти виявлення ІО (мінімальної помилки пропуску цілі), то функ-

цією інформативності може бути кількість правильно розпізнаних ІО в навчальній вибірці, а платіжна матриця наступного виду:

$$C = \begin{vmatrix} 0 & 0,9 \\ 0,1 & 0 \end{vmatrix}. \quad (9)$$

Тобто штраф за помилкове розпізнавання образу, що є інформаційною операцією, як елемента класу Ω_2 , тобто пропуск цілі набагато вище за віднесення образу, що не є ІО, до класу Ω_1 . Для вибору оптимального набору інформативних ознак з апріорного словника використовувати повний перебір не є можливим, якщо кількість ознак апріорного словника більше 20. Отже, необхідно використовувати методи перебору наборів ознак меншої складності, що дозволяють знаходити оптимальне або близьке до оптимального рішення.

Висновок. Розглянуто два основних підходи до виявлення ІО: класичний та на основі динаміки інформаційних потоків. При розгляданні підходу на основі динаміки інформаційних потоків було доведено, що неправдива новина характеризується тими ж самими часовими характеристиками, що й інформаційні операції: етапи фону, затишшя, підготовки, затишшя і атаки. Вказані етапи можуть бути виокремлені із графічного представлення кількості результатів пошукової видачі за запитом по неправдивій новині.

СПИСОК ЛІТЕРАТУРИ

1. Расторгуев С.П. Математические модели в информационном противоборстве (экзистенциальная математика). М.: АНО ЦСОиП. 2014. 276 с.
2. Ландэ Д.В. Динамика информационных потоков и информационные операции. *Информационные технологии для менеджмента*. 2010. № 10. С. 22–27.
3. Додонов А.Г., Ландэ Д.В. Методика аналитического исследования динамики событий на основе мониторинга веб-ресурсов сети Интернет. Информационные технологии и безопасность: основы обеспечения информационной безопасности: Мат. международной научной конференции ИТБ-2014. Киев: ИПРИ НАН Украины. 2014. С. 3–17.
4. Зеленков Ю.Г., Сегалович И.В. Сравнительный анализ методов определения нечетких дубликатов для Web-документов. Тр. 9-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Переславль-Залесский: Изд-во ИПС РАН. 2007. С. 166–174.
5. Загоруйко Ю.А., Саломатина Н.В., Серый А.С., Сидорова Е.А., Шестков В.К. Выявление нечетких дубликатов при автоматическом формировании тематических коллекций документов на основе Web-публикаций. Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2013. Т. 11. № 4. С. 59–70.
6. Tarjan R. Depth first search and linear graph algorithms SIAM journal on computing. 1972. Vol. 1, N 2. P. 146–160.
7. Татт У. Теория графов: Пер. с англ. М.: Мир. 1988. 424 с.

Одержано 06.11.2018