

УДК 004.91

В. О. Додонов, Д. В. Ланде, В. Г. Путятін
Інститут проблем реєстрації інформації НАН України
вул. М. Шпака, 2, 03113 Київ, Україна

Мультиагентний підхід до моделювання інформаційно-аналітичної системи

Розглянуто питання реалізації мультиагентного підходу до моделювання інформаційно-аналітичної системи, який реалізується у вигляді відповідної інформаційної технології, для подальшого застосування сценарного підходу. Представлено мультиагентну модель поширення інформаційних повідомлень, які містять посилання на інформаційні ресурси в мережі Інтернет. Результати моделювання підтверджено шляхом дослідження реальної мережі мікроблогів Twitter. Описано етапи створення корпоративної системи моніторингу мережесвих інформаційних ресурсів, склад яких визначається посиланнями в мікроблогах. Наведено переваги такого підходу.

Ключові слова: мультиагентна модель, активний агент, моделювання, система, сценарій.

Вступ

Метою цієї роботи є обґрунтування можливості застосування мультиагентного підходу при побудові інформаційного забезпечення інформаційно-аналітичних систем, що застосовуються для формування сценаріїв інформаційного впливу.

Науковим фундаментом з питань прикладного характеру сценарного підходу стала низка робіт [1–4], де можна виділити основну особливість даного методу, що складається у наступному: будуються сценарії можливих варіантів розвитку подій у майбутньому, на підставі чого формулюються, а потім відбираються стратегічні альтернативи, які працюють у кожному сценарії і служать підставою для прийняття рішень про вибір інтегрованої стратегії.

Сценарії являють собою комплекс моделей, за допомогою яких при певних умовах функціонування системи визначаються послідовності експертно-значущих подій (фазових станів системи), з яких складається загальний процес управління.

Сценарне моделювання передбачає проведення альтернативних розрахунків з даними, що відповідають різним варіантам розвитку ситуації, а також розробку

логічно та методологічно обґрунтованої схеми розвитку або прогнозу ситуації з подальшою оцінкою реалізації того чи іншого варіанту розвитку подій. При цьому може здійснюватись як оцінка різних параметрів, що пов'язані з реалізованим процесом, так і оцінка впливу окремих його складових і зовнішніх умов на якість рішення системою поставлених функціональних завдань.

Важливе місце тут займає мультиагентний підхід, що дозволяє проводити багатоваріантний ситуаційний аналіз системи, що моделюється. Сутність мультиагентного підходу при побудові сценаріїв полягає в побудові середовища активних агентів, визначення алгоритмів їхнього функціонування та взаємодії, виявлення нових закономірностей, зв'язків, когнітивних карт, а також комплексу математичних моделей формування сценаріїв на комп'ютерному моделюючому комплексі.

Методи побудови сценаріїв отримали широкий розвиток після створення спеціальних засобів і мов програмування сценаріїв. Прикладом може служити нотація MSDL (Military Scenario Definition Language) [1], розроблена Simulation Interoperability Standards Organization, яка орієнтована на створення військових сценаріїв процесу моделювання.

На практиці моделювання на основі мультиагентного сценарного підходу реалізується у вигляді відповідної інформаційної технології.

Отримані в результаті сценарії можуть бути організовані в єдину ієрархічну структуру. Верхній рівень створюють складні сценарії реалізації комплексних завдань управління, що обумовлені цільовою функцією системи; в такі сценарії як дочірні елементи можуть входити самостійні сценарії, що реалізують менш складні завдання.

Ієрархічна структура сценаріїв забезпечує механізм побудови складних сценаріїв з використанням набору більш простих, при цьому складність генерованого сценарію визначається поставленим завданням і глибиною деталізації технологічного процесу.

На етапі створення комп'ютерного моделюючого комплексу здійснюється всебічний аналіз об'єктів управління, процесів інформаційної взаємодії, функціональних завдань і вимог до інформаційних ресурсів для їхнього вирішення, визначаються принципові підходи до організації програмно-апаратного комплексу та стратегії подальшого розвитку системи.

Далі розглянемо застосування мультиагентного моделювання для дослідження можливості та параметрів бази даних актуальних інформаційних ресурсів, на базі моніторингу соціальних мереж в Інтернеті.

Задача формування бази даних інформаційно-аналітичної системи

Обсяги інформаційних ресурсів у веб-просторі сьогодні ускладнюють оперативне отримання необхідної користувачам інформації навіть за допомогою найпотужніших мережових пошукових систем (Google, Baidu, Yandex і ін.) [5]. Одним із шляхів вирішення цієї проблеми є підключення оцінок великої кількості людей, експертів у предметній області. Така можливість, яку можна назвати краудсорсинговою, відкривається за допомогою змістовного аналізу соціальних мереж, де користувачі «голосують» за ті чи інші інформаційні матеріали, встановлюючи на

них гіперпосилання. Особливо це актуально у сегментних предметних областях, які відповідають потребам корпоративних користувачів. Незважаючи на те, що аналіз соціальних мереж сам по собі є складною науково-технічною задачею, існуючі пошукові можливості деяких з них дають надію на рішення названої проблеми.

У рамках даної роботи пропонується підхід до створення корпоративної системи моніторингу мережевих інформаційних ресурсів, склад яких визначається посиланнями у соціальних мережах, зокрема, мережі мікроблогів Twitter [6].

Разом з тим, процеси поширення інформації у мережах, що містить посилання на інформаційні ресурси, вимагають детального аналізу. Моделювання поширення інформації дозволяє досліджувати відповідні інформаційні процеси, виявляти закономірності, які можуть використовуватись як при вивченні механізмів передачі інформації у таких мережах, так і рівня її впливу на людей [7].

Мультиагентна модель розповсюдження повідомлень

Для створення мультиагентної моделі поширення інформації, перш за все, необхідно сформувати близький до реальності віртуальний інформаційний простір, населений віртуальними агентами, з якими асоціюються окремі повідомлення у соціальній мережі, і які інкапсулюють в себе гіперпосилання на інформаційні ресурси мережі Інтернет [8, 9]. Передбачається, що окремі агенти можуть [10]:

- 1) самозароджуватися;
- 2) породжувати нових агентів шляхом репостінгу (repost);
- 3) «вмирати» — зникати з простору агентів;
- 4) отримувати лайки (like) від інших агентів.

Агент має «потенціал», що залежать від його віку, авторитетності (посилань на нього) і плодючості (кількості породжених їм агентів, репостів). Варіювання цими параметрами управління дозволили змоделювати профілі поведінки інформаційних сюжетів. У результаті проведених досліджень було реалізовано програму еволюції простору агентів, досліджено еволюцію мультиагентної системи, знайдено аналогії з реальними тематичними інформаційними потоками. Було виявлено статистичні закономірності, що відносяться до життєвого циклу окремих повідомлень, розподіл яких відповідає розподілу Вейбулла. Дані моделювання було перевірено шляхом дослідження реальної мережі мікроблогів Twitter. Збіг результатів моделювання і параметрів розподілу реальної мережі дозволяють говорити щодо закономірностей, властивих реальним мережам, а також щодо адекватності моделі.

Моделювання динаміки всього інформаційного потоку починається з одного агента. Поява нового агента можлива двома способами. Перший полягає в копіюванні існуючого агента за допомогою операції репост. Також можливе самозародження агента, що відповідає публікації нового повідомлення. Таким чином, у кожен момент часу з певними можливостями, з кожним із агентів, може статися будь-яка з подій. Також у будь-який момент часу з імовірністю p_s може з'явитися новий агент у результаті самозародження.

Розглянемо життєвий шлях одного агента. Агент з'являється з початковим значенням енергії E_0 , і далі його енергія змінюється залежно від подій, які з ним

відбуваються. Будемо вважати, що можливі дві події: лайк і репост. За одиницю часу може статися одна з цих подій, обидві одночасно або не відбутися жодної.

Позначимо через ε_t значення енергії агента на момент часу t . Тоді значення енергії у наступний момент часу можна записати в такий спосіб:

$$\varepsilon_{t+1} = \varepsilon_t + \delta_t,$$

де δ_t є випадковою величиною зі значенням у $\{-1, 0, 1, 2\}$. Згідно з правилами зміни енергії, введеними вище, збільшення енергії на 2 відповідає тому, що відбулись одночасно лайк і репост; збільшення на 1 — стався тільки репост; енергія не змінюється, якщо був лайк, і зменшується на 1, якщо не відбулося ні однієї з подій. Отже, можна вказати умовний розподіл δ_t при відомій енергії ε_t :

$$\begin{aligned} P(\delta_t = 2 | \varepsilon_t = E) &= p_{like}^{(E)} p_{repost}^{(E)}; \\ P(\delta_t = 1 | \varepsilon_t = E) &= (1 - p_{like}^{(E)}) p_{repost}^{(E)}; \\ P(\delta_t = 0 | \varepsilon_t = E) &= p_{like}^{(E)} (1 - p_{repost}^{(E)}); \\ P(\delta_t = -1 | \varepsilon_t = E) &= (1 - p_{like}^{(E)}) (1 - p_{repost}^{(E)}). \end{aligned}$$

Ці формули справедливі при $E > 0$. Далі будемо використовувати позначення $P_{\Delta}^{(E)} = P(\delta = \Delta | \varepsilon = E)$. Процес зміни енергії агента можна розглядати як цілочисельне випадкове блукання з перехідними ймовірностями:

$$p_{ij} = \begin{cases} P_{j-i}^{(i)}, & (j-i) \in \{-1, 0, 1, 2\}, \quad i > 0; \\ 1, & i = j = 0; \\ 0, & \text{інакше.} \end{cases}$$

Так як значення енергії у наступний момент часу залежить тільки від значення енергії у попередній момент часу, то стохастична послідовність $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots)$ є марківським ланцюгом з перехідними ймовірностями p_{ij} .

Позначимо через τ_{E_0} час життя агента з початковим значенням енергії E_0 або, що те ж саме, час, за який з E_0 потрапили до 0. У реалістичній моделі необхідно мати оцінку $P(\tau_{E_0} > T_{\max}) < \varepsilon$ для малого ε і не великого значення T_{\max} , для того, щоби можна було замість нескінченних послідовностей $(\Delta_1, \Delta_2, \dots, \Delta_t, \dots)$ розглядати скінченні.

Розглянемо функцію $\rho_T(E) = P(\tau_E > T)$. Справедливо рекурентне співвідношення:

$$\rho_T(E) = P_2^{(E)} \rho_{T-1}(E+2) + P_1^{(E)} \rho_{T-1}(E+1) + P_0^{(E)} \rho_{T-1}(E) + P_{-1}^{(E)} \rho_{T-1}(E-1).$$

Систему таких рекурентних співвідношень можна вирішити, використовуючи початкові умови:

$$\rho_0(E) = \begin{cases} 0, & E = 0; \\ 1, & E \neq 0. \end{cases}$$

При початкових параметрах $p_{i_0} = 0,4$; $p_{r_0} = 0,1$ з рішення рекурентного рівняння можна отримати оцінку $P(\tau_{E_0} > 1,5E_0) < 10^{-3}$. Тобто час життя агента обмежено $1,5E_0$ з великою ймовірністю і, отже, для отримання досить точних оцінок розподілу для $(\delta_0, \delta_1, \dots, \delta_i, \dots)$ можна розглядати вектори скінченної довжини $T_{\max} = 1,5E_0$.

У результаті моделювання було визначено, що розподіл середнього час життя, кількості лайків і репостів у даній моделі відповідає щільності розподілу Вейбулла [8]:

$$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Параметри розподілу Вейбулла k і λ були отримані методом максимальної правдоподібності. При зазначених початкових параметрах отримані значення: $k = 1,9$; $\lambda = 3,8$.

Отримані результати моделювання порівнювалися з проведеними результатами дослідження життєвого циклу новинних повідомлень у мережі мікроблогів Twitter, де, зокрема, аналізувалися показники зростання кількості спеціальних репостів (ретвіттів) обраних повідомлень. Розподіл лайків і ретвіттів у цьому випадку, як і в моделі, відповідав стандартному розподілу Вейбулла, причому параметр з високою точністю збігся з модельним.

Формування бази даних актуальних інформаційних матеріалів з мережі Інтернет

Авторами протягом травня 2016 року здійснювався експеримент зі збору повідомлень з мережі мікроблогів Twitter, для чого в пошуковому інтерфейсі цієї мережі на періодичній основі оброблявся пакет із 100 запитів з банківської і бізнес-тематики. В результаті було отримано наступні кількісні дані, що пов'язані з кількістю інформаційних ресурсів мережі Інтернет, на які були вказані посилання з Twitter-повідомлень.

1. Відскановано близько 100 тисяч повідомлень по 100 елементарним запитам до Twitter за травень 2016 р.
2. 58 % повідомлень містять гіперпосилання на веб-ресурси мережі Інтернет.
3. Кількість унікальних гіперпосилань — 48 %.

4. Функція розподілу кількості гіперпосилань на одні й ті ж джерела — степенева.

5. Є проблеми ідентифікації зовнішніх посилань, що пов'язані, перш за все, з використанням «коротких посилань» — переадресації з такими базовими адресами:

- <http://migre.me/>
- <http://bit.ly/>
- <http://ow.ly/>
- <http://tinyurl.com/>
- <https://lnkd.in/>
- <https://goo.gl/>
- <http://wp.me/>
- <http://j.mp/>
- <http://dlvr.it/>

6. Виявлено найбільш часто цитовані ресурси мережі Інтернет:

- <https://youtu.be> (<https://www.youtube.com>)
- <http://fb.me/> (<https://www.facebook.com/>) — публічні сторінки
- <https://vk.com/>
- <https://twitter.com/> — посилання на ту ж саму соціальну мережу
- <https://plus.google.com/>
- <http://livejournal.com/>

Таким чином, практика показала, що розподіл інкапсульованих у повідомлення соціальних мереж посилань на інформаційні ресурси мережі Інтернет відповідає степеневому розподілу (рис. 1), а не розподілу Вейбулла. Відповідно з цим фактом було наведено доповнення до зазначеної вище моделі, а саме, зміна ймовірності репостінгу повідомлень, які містять посилання на зовнішні інформаційні ресурси.

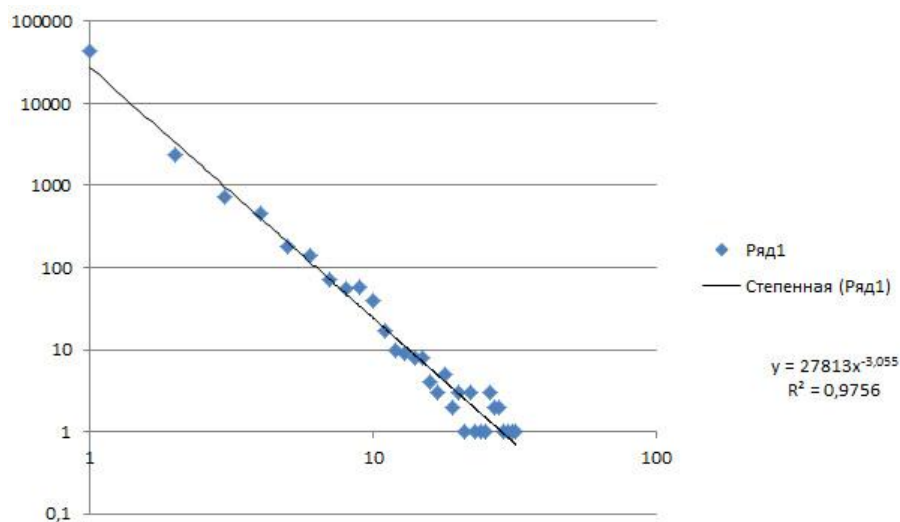


Рис. 1. Розподіл кількості ретвітів повідомлень, що містять посилання на зовнішні інформаційні ресурси

На основі отриманої інформації про розподіл повідомлень, на які реалізовані посилання з мережі мікроблогів, пропонується наступна «краудсорсингова» схема формування бази даних корпоративної системи моніторингу мережевих інформаційних ресурсів (рис. 2).

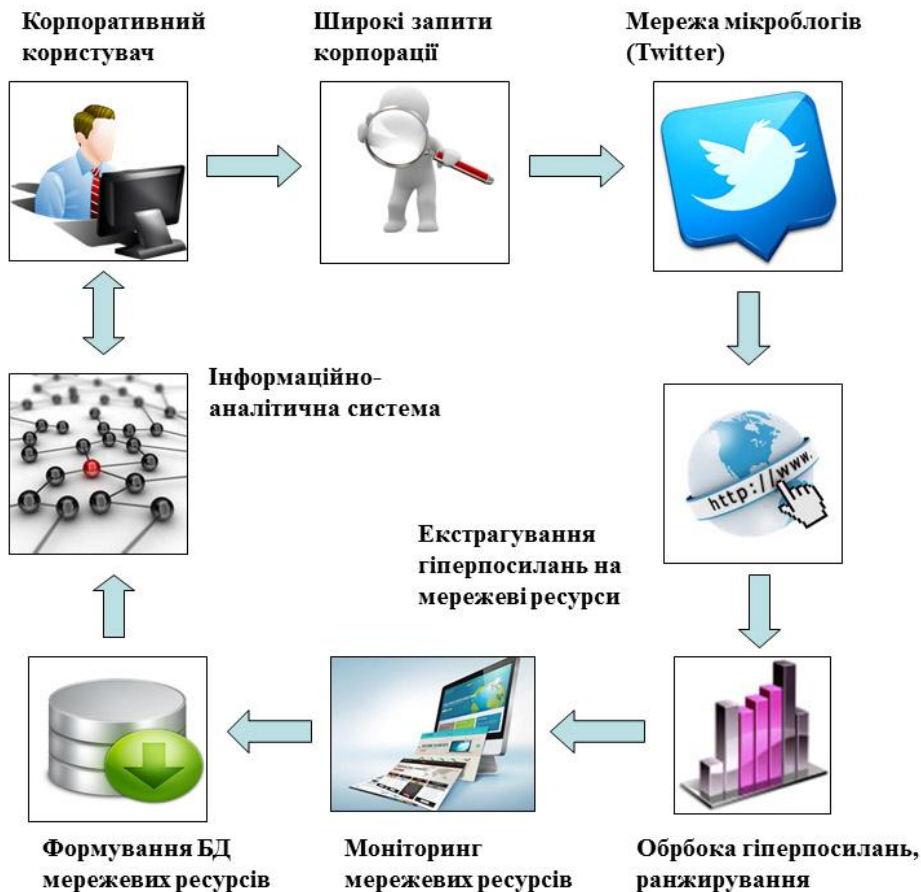


Рис. 2. Загальна схема формування бази даних корпоративної системи моніторингу мережевих інформаційних ресурсів

Наведемо основні етапи цієї процедури.

1. В інтересах корпоративного користувача створюються запити до мережі мікроблогів, наприклад:

Банки України

ДіамантБанк

Укрсиббанк

ПриватБанк

Банк Хрещатик

Платинум Банк

Кредит Дніпро

2. Сформований «широкий» пакет запитів передається програмі сканування мережі мікроблогів, у результаті чого на корпоративний сервер на регулярній надходять повідомлення, формально релевантні цим запитам.

3. Витяг з відсканованих повідомлень мережі мікроблогів гіперпосилань на зовнішні мережеві інформаційні ресурси.

4. Обробка гіперпосилань, розкриття «коротких адрес», сортування гіперпосилань, ранжирування окремих документів і зовнішніх джерел.

5. Сканування зовнішніх інформаційних ресурсів, відповідних виділеним гіперпосиланням, первинна обробка отриманих документів, приведення їх до вхідного формату використовуваної корпоративної інформаційно-аналітичної системи.

6. Завантаження сформованого інформаційного потоку в корпоративну інформаційно-аналітичну систему, надання в доступ корпоративним користувачам.

У рамках мультиагентних систем агенти розглядаються як автономні компоненти, що діють (інформують користувачів мережі, у нашому випадку) за певним сценарієм. Агенти безперервно взаємодіють один з одним, можуть передавати один одному деяку інформацію. У такій системі рішення задачі можуть формуватися за рахунок взаємодії великої кількості агентів, що безперервно взаємодіють один з одним, а також формується колективна поведінка.

Представлений підхід до формування баз даних на основі врахування посилань у мікроблогах на інформаційні ресурси забезпечує, поряд з істотним скороченням охоплення інформаційного простору, наступні переваги.

1. Оперативність — інформаційне повідомлення потрапляє в базу даних інформаційно-аналітичної системи в режимі реального часу в міру того, як на нього зробив посилання перший користувач.

2. Охоплення головних інформаційних матеріалів за темою. Врахування думки аудиторії зацікавлених користувачів, контент повідомлень яких задовольняє широке корпоративне запитам. Можливість ранжирування інформаційних матеріалів, виходячи з інтересів користувачів соціальних мереж.

3. Компактність баз даних, а, отже, зручність доступу кінцевих користувачів. Передбачуваність обсягів баз даних, динаміки їхнього наповнення.

4. Технологічна сумісність з існуючими інформаційно-аналітичними системами і системами контент-моніторингу.

5. Можливість виявлення інформаційних кампаній, операцій.

Висновки

У результаті описаних досліджень побудовано мультиагентну модель поширення інформаційних повідомлень, які містять посилання на інформаційні ресурси в мережі Інтернет. Результати моделювання перевірено шляхом дослідження реальної мережі мікроблогів Twitter.

Знайдені закономірності можуть використовуватися при формуванні баз даних інформаційно-аналітичних систем, при вивченні аномалій у статистиці посилань на окремі інформаційні матеріали, а відповідно, і у виявленні інформаційних операцій, штучно підтримуваних інформаційних кампаній [11].

1. Ringland G. Scenario planning: managing for the future / G. Ringland. — Chichester: Willey, 1998. — 422 с.

2. *Сценарный* подход к моделированию функциональных задач на компьютерном моделирующем комплексе / Додонов А.Г., Путьятин В.Г., Куценко С.А., Юрасов А.А. // Математические машины и системы. — 2015. — № 2. — С. 113–129.
3. *Морозов А.А.* Ситуационные центры. Информационные технологии будущего / А.А. Морозов, В.А. Яценко. — К.: СП «Интертехнодрук», 2008. — 332 с.
4. *Морозов А.А.* Информационно-аналитические технологии поддержки принятия решений на основе регионального социально-экономического мониторинга / А.А. Морозов, В.Л. Косолапов. — К.: Наук. думка, 2002. — 230 с.
5. *Ландэ Д.В.* Интернетика: Навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский, И.В. Безсуднов. — М.: Либроком (Editorial URSS), 2009. — 264 с.
6. *TEDAS: A Twitter-based Event Detection and Analysis System* / [Li R., Lei K.H., Khadiwala R., Chang K.C.C.] // Data Engineering (ICDE), 2012 IEEE 28th International Conference, 2012. — P. 1273–1276.
7. *Конкурентная разведка в компьютерных сетях* / [Додонов А.Г., Ландэ Д.В., Прищепя В.В., Путьятин В.Г.]. — К.: ИПРИ НАН Украины, 2013. — 248 с.
8. *Woo J.* Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog / J. Woo, H. Chen // Springerplus. — 2016. — N 22. — P. 5–66.
9. *Lerman K.* Information Is Not a Virus, and Other Consequences of Human Cognitive Limits. Future Internet / Lerman K. — 2016; 8(2):21.
10. *Ландэ Д.В.* Мультиагентная модель распространения информации в социальной сети / Д.В. Ландэ, А.Н. Грайворонская, Б.А. Березин // Реєстрація, зберігання і оброб. даних. — 2016. — Т. 18, № 1. — С. 70–77.
11. *Додонов А.Г.* Распознавание информационных операций: мультиагентный подход / А.Г. Додонов, Д.В. Ландэ, В.А. Додонов // Открытые семантические технологии проектирования интеллектуальных систем (OSTIS-2016): материалы VI междунар. науч.-техн. конф. (Минск 18–20 февраля 2016 года). — Минск: БГУИР, 2016. — С. 253–256.

Надійшла до редакції 10.06.2016