УДК 004.738.5:004.056.5(045)

Молодецька-Гринчук К.В.
Житомирський національний агроекологічний університет

МЕТОД ПОБУДОВИ ПРОФІЛІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ АКТОРІВ СОЦІАЛЬНИХ ІНТЕРНЕТ-СЕРВІСІВ

У статті обгрунтовано атрибути і характеристики профіля актора соціальних інтернет-сервісів, які використовуються для оцінки його рівня загрози і підвищення ефективності функціонування системи забезпечення інформаційної безпеки держави у соціальних інтернет-сервісах. Запропонований метод дозволяє автоматизувати процедури раннього виявлення загроз інформаційній безпеці держави у соціальних інтернет-сервісах завдяки градієнтному бустингу на бінарних деревах.

Ключові слова: профіль актора, градієнтний бустинг, оцінка загроз, класифікація.

Постановка проблеми у загальному вигляді та її зв'язок із важливими практичними завданнями. Сьогодні соціальні інтернет-сервіси (СІС) представляють собою дієвий інструмент групової взаємодії учасників віртуальних спільнот, яких називають акторами. Особливістю сучасних СІС є доступність для користувачів, оперативність публікації контенту, наявність засобів для організації акторів у групи за інтересами,

транскордонність процесів взаємодії тощо [1, 2]. В свою чергу, актори виступають джерелом контенту СІС, за допомогою якого інформують інших учасників віртуальних спільнот, виражають власну точку зору на актуальні події. Однак, такий контент може бути недостовірним, неповним або мати упереджений характер, створюючи передумови для поширення у суспільстві соціальної напруженості, протестних настроїв, маніпулювання суспільною думкою тощо. Тому СІС є ефективним засобом проведення інформаційних операцій проти людини, суспільства, держави [3]. У працях [3, 4] показано, що найпоширеніщими способами дії загроз інформаційній безпеці держави у СІС є інформаційний вплив на акторів віртуальних спільнот для прихованих деструктивних змін чи корекції їх поведінки та розвідувальна діяльність, направлена на цілеспрямоване добування інформації. Такі загрози інформаційній безпеці держави у СІС реалізовуються за безпосередньої участі спеціально залучених акторів.

У процесі створення профілів акторів СІС користувачі вводять персональну інформацію, сформульовану в анкеті. Тому використання узагальнених даних профіля актора, його публікацій у віртуальних спільнотах, особливості взаємодії з іншими акторами [5–8] є перспективним напрямком наукових досліджень для побудови їх профілів інформаційної безпеки. Під *профілем інформаційної безпеки актора СІС* у статті будемо розуміти набір агрегованих характеристик профіля актора у СІС, які дозволяють визначити рівень його загрози як можливого учасника інформаційних акцій, направлених проти інформаційної безпеки людини, суспільства, держави. Проблема прийняття рішення щодо рівня загрози акторів пов'язана з недостатньою кількістю атрибутів профілів та їх низькою інформативністю, складністю процедур автоматизованого аналізу змісту контенту СІС. Також актори часто вказують неповну або недостовірну інформацію про себе для анонімності взаємодії з іншими суб'єктами СІС, що додатково ускладнює процес побудови профіля інформаційної безпеки. Тому розробка методів автоматизованої побудови профілів інформаційної безпеки акторів і моделей прийняття рішень щодо їх залучення до інформаційних акцій є актуальним теоретико-прикладним завданням на шляху вирішення проблеми розроблення ефективної системи забезпечення інформаційної безпеки держави у СІС.

Аналіз останніх досліджень і публікацій [5–8] показав, що атрибути профілів акторів у СІС прийнято поділяти на числові – вік, рівень доходів та категоріальні – сімейний стан, професія, життєві цінності та інші, а набір конкретних значень атрибутів формує його загальну характеристику. Встановлено, що для цього використовуються методи машинного навчання, зокрема бінарна класифікація на основі атрибутів профілів акторів СІС. В задачах гендерної класифікації також застосовують структурні ознаки, які ґрунтуються на інтелектуальному аналізі контенту, який публікує актор.

Колективом авторів у публікаціях [5, 6] показано, що в загальному випадку задача встановлення прихованих атрибутів профілів акторів належить до класичних задач соціолінгвістики – визначення характерних особливостей мови різних соціальних груп. Для цього використовують методи машинного навчання з учителем для класифікації за лінгвістичними та іншими ознаками акторів у попередньо визначені класи, які відповідають заданим значенням наборів атрибутів. Встановлено, що більшість публікацій щодо виявлення прихованих атрибутів акторів [9, 10] присвячені визначенню гендерної належності. Наукові дослідження [11, 12] спрямовані на визначення віку акторів СІС як неперервної та дискретної величини. У загальному випадку для встановлення віку актора використовують текстовий контент, який він генерує, і стилістичні ознаки цього контенту. У публікації [7] запропоновано визначати політичну приналежність, відношення до мережі швидкого обслуговування *Starbacks* та етнічність акторів на основі атрибутів профіля у СІС, особливостей поведінки, змісту повідомлень і зв'язків з іншими акторами. Для визначення геолокації акторів СІС у [13, 14] використано відповідно методи тематичного моделювання текстового контенту і розподіл слів залежно від географічного місцеположення актора.

Отже, у загальному випадку профіль актора у СІС містить в явному або прихованому вигляді інформацію, достатню для прийняття рішення щодо його залучення до інформаційних акцій. Таким чином, існує об'єктивне протиріччя між рівнем розвитку інформаційних технологій та науковим базисом автоматизованого виявлення загроз в СІС, а відсутність дієвих методик аналізу профілів акторів для раннього попередження інформаційних впливів додатково актуалізує обраний напрямок досліджень.

Метою статті є підвищення ефективності функціонування системи забезпечення інформаційної безпеки держави у СІС шляхом автоматизації процедур завчасного виявлення акторів, які є суб'єктами інформаційних акцій у віртуальних спільнотах.

Виклад основного матеріалу дослідження. На сучасному етапі серед форм інформаційного протистояння в СІС виділяють такі [3]: розвідувальна, спрямована на приховане добування даних та інформації про управління в системах протиборчої сторони; наступальна, метою якої є викривлення, блокування, знищення інформації; оборонна, що проводиться державою для захисту власних інтересів. Для реалізації таких форм протистояння у СІС актори використовують відповідні моделі поведінки і взаємодії у віртуальних спільнотах.

Досвід гібридної війни з Російською Федерацією показав, що до проведення інформаційних операцій у СІС залучається спеціалізоване програмне забезпечення [2, 3], яке використовує соціальних ботів для поширення заданого контенту з метою деструктивного впливу на акторів. Також встановлено, що тролі є найбільш агресивним типом соціальних ботів. Головна функція тролів полягає у публікації образливих і ворожих коментарів, створенні суперечок між акторами, підтримці інформаційного фону для поширення заданого контенту. Однак, в якості троля можуть виступати реальні люди під впливом маніпулятивних технологій або особи, які на платній основі коментують і поширюють задані публікації контенту в СІС – так звані «тролі з Ольгіно». У результаті попередніх досліджень автором розроблено технологію виявлення соціальних ботів, запропоновану в публікації [2]. У свою чергу, автоматизація процедур аналізу профілів акторів у розрізі загрози інформаційній безпеці держави зводиться до побудови профіля інформаційної безпеки актора. Метод побудови профіля інформаційної безпеки актора СІС на основі аналізу його атрибутів грунтується на дослідженнях M. Pennacchiotti та A.-M. Popescu й зводиться до такого.

Етап 1. Аналіз атрибутів профіля актора СІС. Атрибути профіля актора СІС є опосередкованим джерелом інформації про його особу та інтереси. Так, часто актори вказують неповну чи недостовірну інформацію з різною метою – створення позитивного іміджу, приховування даних про особу тощо. Суттєве значення для розв'язку задачі побудови профіля інформаційної безпеки актора мають наступні його атрибути:

Ім'я актора. Дослідження показують, що імена штучно створених аккаунтів часто повторюються. Також важливе значення має інформація про акторів, які вже були ідентифіковані як учасники інформаційних акцій у СІС;

Місця народження та проживання актора використовуються для встановлення його належності до географічної області та подальшого виявлення закономірностей поведінки й взаємодії у СІС;

Навчальний заклад і місце роботи є джерелами інформації про кваліфікацію актора і додатково містять інформацію про його геолокацію.

Етап 2. Визначення показників активності публікації контенту. Встановлено, що серед акторів СІС, залежно від особливостей публікації контенту у своєму профілі, виділяють дві категорії. До першої з них зазвичай відносяться актори, які нечасто публікують контент, мають велику кількість друзів і схильні до інформаційного пошуку, коментують публікації інших акторів й вступають з ними в діалог – так звані *споживачі контенту*. Друга категорія акторів часто публікує власний контент або гіперпосилання на сторонні інформаційні ресурси у СІС і називається *постачальниками контенту*. Таким чином, встановлення показників частоти публікації контенту різного походження, геш-тегів, гіперпосилань тощо, їх аналіз і

узагальнення дозволяють віднести актора СІС до однієї з визначених категорій. Тому для оцінки показників активності акторів у СІС використаємо такі:

Загальна кількість публікацій актора у профілі СІС;

Загальна кількість і частка публікацій, які є репостами контенту профілів інших акторів;

Загальна кількість і частка коментарів публікацій або відповідей на них;

Середня кількість геш-тегів і гіперпосилань на публікацію;

Середній час між публікаціями контенту і стандартне відхилення;

Середня кількість публікацій в день і стандартне відхилення;

Частка публікацій контенту актором за останні 24 години.

Етап 3. Встановлення ознак, властивих контенту профіля актора. Особливості мови актора, його життєві цінності, сфера інтересів проявляються вживанням відповідних лінгвістичних одиниць у текстових публікаціях. Виявлення таких характерних ознак контенту, який публікує актор у СІС, є джерелом даних для побудови профіля інформаційної безпеки актора. При цьому текстовий контент представляється у вигляді непов'язаного набору слів або *Bag of Words* і встановлюються такі атрибути профіля актора:

Характерні слова, які вживає актор у свої публікаціях. Такі слова є лексичними одиницями для ідентифікації особливих ознак актора і подальшого віднесення його до заданого класу загроз. Для детектування характерних слів використано ймовірнісну модель їх автоматизованого вилучення [7, 15], яка грунтується на використанні даних кількох базових акторів B_i для кожного із заданих класів c_i і полягає у наступному:

– оцінка належності характерного слова до заданого класу акторів

$$char(w,c_i)=\frac{|w,B_i|}{\sum_{j=1}^{n}|w,B_j|}, \qquad (1)$$

де w – слово, яке вживається хоча б один раз базовими акторами класу c_i;

$|w,B_i|$ – кількість вживань слова w усіма акторами класу c_i;

n – кількість класів.

Для кожного класу обираються f характерних слів довжиною більше трьох символів, які мають найвищу оцінку.

– розрахунок коефіцієнта, який пов'язує кожне характерне слово chw з актором a

$$score\_chw(a)=\frac{|a,chw|}{\sum_{w\in W_a}|a,w|}, \qquad (2)$$

де $|a,chw|$ – кількість випадків вживання актором a характерного слова chw;

W_a – загальна кількість слів, які вживає актор a;

– розрахунок для кожного класу коефіцієнта, який ставить у відповідність актору функцію

$$score\_c(a)=\frac{\sum_{chw\in ChW}|a,chw|}{\sum_{w\in W_a}|a,w|}, \qquad (3)$$

де ChW – множина характерних слів для класу c;

Виявлення спільних інтересів базових акторів одного класу на основі прихованої тематики опублікованого контенту. Для вилучення прихованої у контенті профіля актора інформації доцільно використати методи ймовірнісного тематичного моделювання і ймовірнісного латентно-семантичного індексування (*PLSI*) зокрема. Даний метод грунтується на твердженні, що поява слів t у публікаціях d зумовлена змінними z, які представляють собою латентну (приховану) тематику контенту [16]. В нашому випадку латентна тематика контенту акторів представляє собою їх інтереси, а ймовірнісна модель появи пари (d,t) приймає такий вигляд [17]

$$P(d_m, t_n) = \sum_{k=1}^{J} P(z_k) P(d_m \mid z_k) P(t_n \mid z_k), \qquad (4)$$

де $P(z_k)$ – розподіл тем по колекції публікацій;

$P(d_m \mid z_k)$ – ймовірність, що публікація d_m належить до групи публікацій тематики z_k;

$P(t_n \mid z_k)$ – ймовірність, що слово t_n належить до групи слів тематики z_k;

Характерні геш-теґи, які використовує актор. Геш-теґи використовують у СІС для об'єднання публікацій акторів у групи, спрощення пошуку контенту на задану тематику і складаються з одного чи декількох слів та позначаються символом #. Актори СІС, які мають спільні інтереси, використовують аналогічні геш-теґи, тому доцільно сформувати набір характерних геш-теґів для базових акторів аналогічно характерним словам. Для цього використовують вирази (1)–(3).

Тональність опублікованого актором контенту. В окремих випадках індикатором належності актора до деякого класу загроз є тональність контенту відносно заданого набору слів. Серед таких слів виділяють пов'язані з питаннями національної безпеки, наприклад, політичного устрою, територіальної цілісності країни тощо. Тональність контенту профіля актора приймає значення «позитивна», «негативна» і «нейтральна». Для аналізу тональності доцільно використати методи машинного навчання з учителем [18–21], перевагами яких є висока точність і швидкодія, ефективність автоматизації процедур аналізу контенту, віднесення контенту до попередньо заданих класів тональності, наявність засобів оцінки точності. Після аналізу тональності контенту розраховуються такі показники:

Частки контенту позитивної, негативної та нейтральної тональності;

Середнє значення і стандартне відхилення тональності для усього заданого набору слів;

Число заданих слів, відносно яких актор має негативну, позитивну чи відсутню точку зору.

Етап 4. Аналіз зв'язків актора у СІС. Інформативними атрибутами профіля актора у СІС є його зв'язки з іншими акторами і віртуальними спільнотами, згадування їх у дописах чи поширення контенту. Тому для аналізу зв'язків актора у СІС врахуємо такі атрибути його профіля:

Загальна кількість друзів, підписників і віртуальних спільнот актора. Ці показники дозволяють зробити висновки про мету користування СІС у розрізі інформаційного обміну. Якщо актор має велику кількість друзів і є учасником багатьох віртуальних спільнот, то він є споживачем контенту СІС. Якщо у актора велика кількість підписників і він часто публікує контент, то він є постачальником контенту;

Характеристика профілів друзів і віртуальних спільнот актора. Інтереси актора проявляються у виборі друзів у СІС і віртуальних спільнотах, в яких він бере участь. Відомо, що за допомогою таких механізмів у СІС реалізовано організацію взаємодії між акторами зі спільними інтересами, а також можливості для їх самоорганізації у реальному житті. Для аналізу особливостей зв'язків актора використаємо принцип детектування характерних слів етапу 2 цього методу і рівняння (1)–(3). Для цього оберемо профілі популярних акторів, які є лідерами думок, відомими діячами тощо і використаємо їх в якості базових;

Поширення контенту друзів і віртуальних спільнот. Аналогічним чином до аналізу особливостей зв'язків актора у попередньому пункті дослідимо джерела, контент яких він цитує. Залежно від своїх інтересів, актор поширює контент, який відповідає його інтересам. Тому доцільно визначити базових акторів і віртуальні спільноти для кожного з класів загроз і використати рівняння (1)–(3) з метою віднесення актора до однієї з цих груп.

Етап 5. Визначення класу загрози. Після оцінки усіх характеристик профіля актора СІС необхідно віднести його до одного з класів загроз:

$$Y = \{\text{дуже високий}; \text{високий}; \text{значний}; \text{допустимий}; \text{низький}\}.$$

Для цього використано методи машинного навчання з учителем, які дозволяють виконати класифікацію акторів у попередньо задані класи Y. Ефективними для розв'язку задачі класифікації акторів за рівнем загроз є методи, які грунтуються на процедурах бустингу

бінарних класифікаторів. Суть бустингу зводиться до побудови композиції алгоритмів машинного навчання, коли кожен наступний алгоритм компенсує недоліки композиції усіх попередніх алгоритмів. У випадку мультикласової градієнтної бінарної класифікації на деревах рішень на Q класів задаємо функцію втрат як [22, 23]

$$L = -\sum_{q=1}^{K} y_q \log p_q(x), \qquad (5)$$

де y_q відображає належність об'єкта класу q;

p_q – ймовірність належності об'єкта класу q у результаті логістичної регресії.

Тоді підсумковий класифікатор приймає вигляд [22]

$$q(x) = \arg\min_{q\in[1,Q]} \sum_{\tilde{q}=1}^{K} c(q,\tilde{q}) p_{\tilde{q}M}(x), \qquad (6)$$

де $p_{\tilde{q}M}(x)$ – ймовірність належності актора класу загроз Q після проведення M-го циклу бустингу;

$c(q,\tilde{q})$ – вартість помилкового віднесення актора до класу загроз q, коли він належить до класу $\tilde{q}$.

Таким чином, перевагою запропонованого методу побудови профіля інформаційної безпеки у СІС є автоматизація процедур обробки даних профілів і віднесення актора до заданого класу загроз. Перевагою запропонованого підходу до класифікації є врахування вартості помилок класифікації, що дозволяє реалізувати ефективний моніторинг контенту СІС.

Висновки та перспективи подальших досліджень. Запропонований у статті метод побудови інформаційних профілів акторів СІС дозволяє автоматизувати процедури раннього виявлення загроз інформаційній безпеці держави. За даними побудованого профіля системою забезпечення інформаційної безпеки держави приймається щодо залучення акторів до інформаційних операцій у СІС. Розроблений метод відрізняється від відомих застосуванням сучасних методів інтелектуального аналізу даних, зокрема методів машинного навчання з учителем, і його узагальненням для використання у різних видах СІС. Таким чином, метод побудови профілів інформаційної безпеки акторів підвищує ефективність і швидкодію системи забезпечення інформаційної безпеки держави у СІС.

Література:

1. Панченко В. М. Соціальні інтернет-сервіси як засіб прихованого інформаційного впливу // Інформаційна безпека людини, суспільства, держави. – 2012. – № 1(8).– С. 65–69.
2. Молодецька-Гринчук К. В. Підхід до виявлення організаційних ознак інформаційних операцій у соціальних інтернет-сервісах / К. В. Молодецька-Гринчук // Пріоритетні напрямки розвитку телекомунікаційних систем та мереж спеціального призначення. Застосування підрозділів, комплексів, засобів зв'язку та автоматизації в АТО : збірн. матер. IX наук.-практ. конф., 25 листоп. 2016 р. – Київ : ВІТІ, 2016. – С. 130–131.
3. Грищук Р. В. Основи кібернетичної безпеки : моногр. / Р. В. Грищук, Ю. Г. Даник ; за заг. ред. проф. Ю. Г. Даника. – Житомир : ЖНАЕУ, 2016. – 636 с.
4. Молодецька К. В. Узагальнена класифікація загроз інформаційній безпеці держави в соціальних інтернет-сервісах / К. В. Молодецька // Защита информации : сб. науч. трудов. – 2016. – Вып. 23. – С. 75–87.
5. Определение демографических атрибутов пользователей микроблогов / А. Коршунов, И. Белобородов, А. Гомзин [и др.] // Труды Института системного программирования РАН. – 2013. – Т. 25. – С. 179–194.
6. Гомзин А. Г. Методы построения социо-демографических профилей пользователей сети Интернет / А. Г. Гомзин, С. Д. Кузнецов // Труды Института системного программирования РАН. – 2015. – Т. 27. – Вып. 4. – С. 129–143.
7. Pennacchiotti M. Democrats, republicans and Starbucks afficionados: user classification in Twitter / M. Pennacchiotti, A. M. Popescu // Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. – ACM, 2011. – С. 430–438.
8. Beller C. I'm a Belieber: Social Roles via Self-identification and Conceptual Attributes / C. Beller et al. // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. – 2014. – С. 181-186.
9. Schwartz H. A. Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach / H. A. Schwartz [et al.]. – PloS One. – 2013. – Т. 8. – №. 9. – P. 73791.
10. Deitrick W. Gender identification on twitter using the modified balanced winnow / W. Deitrick [et al.] // Communications and Network. – 2012. – Т. 4. – №. 3. – PP. 189–195.

11. Sun J. Applying stylometric analysis techniques to counter anonymity in Cyberspace / J. Sun [et al.] // Journal of Networks. – 2012. – Т. 7. – №. 2. – С. 259-266.

12. Nguyen D. Author age prediction from text using linear regression / D. Nguyen, N. A. Smith, C. P. Rosé – Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. – Association for Computational Linguistics, 2011. – PP. 115–123.

13. Eisenstein J. A latent variable model for geographic lexical variation / J. Eisenstein [et al.] // Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. – Association for Computational Linguistics, 2010. – PP. 1277–1287.

14. Gore R. J. You are what you Tweet: Connecting the geographic variation in America's obesity rate to Twitter content / R. J. Gore, S. Diallo, J. Padilla // Plos One. – 2015. – 10(9). – PP. 0133505.

15. Pasca M. What you seek is what you get: Extraction of class attributes from query logs / M. Pasca, B. Van Durme // In Proceedings of IJCAI. – 2007. – Vol. 7. – PP. 2832–2837.

16. Машечкин И. В. Методы вычисления релевантности фрагментов текста на основе тематических моделей в задаче автоматического аннотирования / И. В. Машечкин, М. И. Петровский, Д. В. Царёв // Вычислительные методы и программирование. – 2013. – Т. 14. – С. 91–102.

17. Steyvers M. Probabilistic topic models / M. Steyvers, T. Griffiths // Handbook of Latent Semantic Analysis. – Philadelphia: Psychology Press, 2007. – Vol. 427. – PP. 414–440.

18. Faraz A. A comparison of text Categorization methods / A. Faraz // International Journal on Natural Language Computing. – 2016. – 5(1). – PP. 31–44.

19. Fernández-Martínez F. Text categorization methods for automatic estimation of verbal intelligence / F. Fernández-Martínez, K. Zablotskaya, W. Minker // Expert Systems with Applications. – 2012. – 39(10). – PP. 9807–9820.

20. Sebastiani F. Machine learning in automated text categorization / ACM Computing Surveys (CSUR) // F. Sebastiani. – 2002. – 34(1). – PP. 1–47.

21. Воронина И. Е. Анализ эмоциональной окраски сообщений в социальных сетях (на примере сети «Вконтакте») / И. Е. Воронина, В. А. Гончаров // Вестник ВГУ, серия: системный анализ и информационные технологии. – 2015. – №4. – С. 151–158.

22. Natekin A. Gradient boosting machines, a tutorial / A. Natekin, A. Knoll // Frontiers in Neurorobotics. – 2013. – Vol. 7. – 21. – PP. 1–21.

23. Freund Y. A decision-theoretic generalization of on-line learning and an application to boosting / Y. Freund, R. Schapire // J. Comput. Syst. Sci. – 1997. – 55. – PP. 119–139.

Рецензент: д.т.н., проф. Марченко Д.М. Надійшла 23.02.2017

Молодецкая-Гринчук К. В.

МЕТОД ПОСТРОЕНИЯ ПРОФИЛЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ АКТОРОВ СОЦИАЛЬНЫХ ИНТЕРНЕТ-СЕРВИСОВ

В статье обоснованы атрибуты и характеристики профиля актора социальных интернет-сервисов, которые используются для оценки его уровня угрозы и повышения эффективности функционирования системы обеспечения информационной безопасности государства в социальных интернет-сервисах. Предложенный метод позволяет автоматизировать процедуры раннего выявления угроз информационной безопасности государства благодаря градиентном бустингу на бинарных деревьях.

Ключевые слова: профиль актора, градиентный бустинг, оценка угроз, классификация.

Molodetska-Hrynchuk K.

METHOD OF CONSTRUCTION ACTOR'S INFORMATION SECURITY PROFILES IN SOCIAL NETWORKING SERVICES

In the article the attributes and characteristics of the profile actor of social networking services, which are used to evaluate its threat level and efficiency of the system of information security of the state in social networking services. The proposed method automates procedures for early detection of threats to information security of the state through gradient boosting on binary trees.

Keywords: actor's profile, gradient boosting, threat assessment, classification.