Виявлення маніпуляцій суспільною думкою у соціальних інтернет-сервісах на основі аналізу текстового контенту

Катерина Молодецька-Гринчук

Кафедра комп'ютерних технологій і моделювання систем, Житомирський національний агроекологічний університет, УКРАЇНА, м. Житомир, бульвар Старий, 7, E-mail: kmolodetska@gmail.com

Коротка анотація – Social networking services represent an effective tool of social communication. However, the communication benefits of social networking services have become an effective tool for attackers impact on public consciousness members of virtual communities. The method for signs of manipulation of public opinion using text mining content. The methodology bases on calculation of uncertainty of information, allowing to automate and increase the effectiveness of the monitoring of information environment of social networking services. The experimental research developed technique which has shown its effectiveness in specific examples of information actions in social networking services. The results of the analysis of convergent researched text content specialized units.

Ключові слова – інформаційна безпека держави, соціальний інтернет-сервіс, маніпуляції, суспільна думка, контент, ентропія.

I. Вступ

Соціальні інтернет-сервіси (СІС) є ефективною платформою взаємодії учасників віртуальних спільнот, яких називають акторами [1, 2]. СІС активно використовуються громадянським суспільством для самоорганізації з метою впливу на політичні й суспільні процеси в державі, формування суспільної думки тощо. Проте, позитивні комунікаційні характеристики СІС перетворили віртуальні спільноти на ефективний інструмент проведення інформаційних операцій проти людини, суспільства держави.

Під час взаємодії акторів у СІС виникає низка психологічних явищ [3-6], використання яких зловмисниками для проведення інформаційних операцій створює передумови для маніпулювання суспільною думкою учасників віртуальних спільнот – акторів. Тому аналіз маніпулятивних технологій, які застосовуються для впливу на акторів, розробка методів своєчасного виявлення маніпулятивних ознак інформаційних операцій у СІС є актуальним теоретико-прикладним завданням на шляху забезпечення інформаційної безпеки держави.

II. Методи маніпуляцій суспільною думкою у СІС

Під *маніпулятивними ознаками інформаційних операцій* у СІС будемо розуміти наявність у контенті віртуальних спільнот прихованого впливу на акторів з метою зміни їх поведінки, цілей, намірів чи інших психологічних характеристик в інтересах суб'єкта впливу [6]. Серед найбільш дієвих технологій, які використовуються у віртуальних спільнотах СІС під час проведення інформаційних операцій, узагальнивши, виділимо наступні [3, 6]:

«Спіраль мовчання» – це модель комунікації, запропонована Е. Ноель-Нойманн, яка описує особливості процесів висловлювання та поширення громадської думки. Суть моделі полягає в приховуванні акторами своєї громадянської позиції, якщо вона не співпадає з точкою зору більшості.

Стадний інстинкт акторів СІС пов'язаний з колективною поведінкою особистості й полягає в тому, що більша увага приділяється публікаціям контенту або віртуальним спільнотам з великою кількістю коментарів, «лайків», репостів, учасників тощо. Такими діями зловмисники виконують соціалізацію контенту.

Лідери думок у СІС представляють собою акторів або віртуальні спільноти акторів, які обізнані в деякій галузі. Вони публікують контент з власною оцінкою, поясненнями і аргументацією подій, а менш активні актори сприймають його як пояснення явищ й фактів.

Посилання на анонімний авторитет зводиться до згадування в якості джерела контенту авторитетних осіб. З метою збільшення переконливості контенту наводяться оцінки експертів, свідчення учасників подій, документи. Однак, джерело фактів не ідентифіковане і відповідальність за поширення такого контенту ніхто не несе.

Емоційний резонанс в СІС використовується для створення у акторів віртуальних спільнот заданого емоційного стану і одночасної передачі контенту. Такий підхід забезпечує сприйняття контенту на рівні емоцій і вимкнення механізмів логіки та критичного мислення.

Відволікання уваги акторів спрямоване на їх перефокусування від першочергового контенту до другорядного, який поданий як сенсація. Таким чином створюється інформаційний шум в СІС, який приховує важливі події.

Міфи або *фейки* – це прийом поширення в СІС контенту, який містить викривлені, спотворені, вигадані факти про дійсність. Метою даної технології маніпуляції є забезпечення сприйняття акторами контенту як правди без критичного осмислення і перевірки фактів. Поширення фейків часто поєднується з іншими технологіями для досягнення бажаного ефекту суб'єктами маніпуляцій свідомістю.

Нейролінгвістичне програмування застосовується у СІС для управління свідомістю акторів з використанням спеціальних лінгвістичних конструкцій контенту, образів, зображень, відео тощо.

Узагальнюючи частинні ознаки використання технологій маніпулятивного впливу на акторів СІС для їх виявлення доцільно виділити наступні [6]: сумнівність викладених фактів; емоційне забарвлення контенту для відображення емоційного стану його

автора; тональність контенту по відношенню до деякого об'єкту чи події, яка відображає оцінювальні судження; сенсаційність контенту, яка має на меті привернути увагу акторів; прихований (імпліцитний) зміст контенту пов'язаний з його глибинним змістом, отриманим в результаті розумової діяльності на основі співвідношення системи знань і цінностей актора з мовними одиницями й конструкціями.

III.Методика виявлення маніпуляцій суспільною думкою у СІС

Розроблено методику виявлення маніпуляцій суспільною думкою акторів у СІС в результаті аналізу текстового контенту, яка ґрунтується на сучасних методах обробки даних – контент-аналізу [5] і машинного навчання та зводиться до такого [6].

Крок 1. Встановлення ознак сумнівності викладених у контенті СІС фактів. Виявляються ознаки недостовірності контенту віртуальних спільнот СІС, а саме: посилання на суб'єктивну точку зору; відсутність аргументації; частка запитальних речень; вживання числових даних; сумнівні висловлювання.

Крок 2. Визначення емоційного забарвлення контенту. Даний крок має на меті встановлення наявності у текстовому контенті проявів індивідуального настрою чи почуттів актора щодо досліджуваних об'єктів чи подій. Суть кроку полягає у встановленні таких ознак: окличні речення у текстовому контенті; вигуки; прислівники; використання емоційного словника.

Крок 3. Оцінка тональності контенту. Метою є визначення позиції актора відносно досліджуваних об'єктів або подій. Задача оцінки тональності контенту віртуальних спільнот розв'язується шляхом застосування методів машинного навчання та інформаційного пошуку. Даний крок [6] зводиться до віднесення тональності публікації, до попередньо визначеної категорії – негативна, позитивна, нейтральна тощо. Оцінка тональності виконується з використанням одного з таких підходів: на основі правил; на основі словників; машинне навчання з учителем; машинне навчання без учителя.

Крок 4. Сенсаційність контенту. На цьому кроці оцінюється здатність текстового контенту своїм змістом зацікавити, вразити і привернути увагу акторів СІС. Даний крок зводиться до виявлення таких ознак: підвищення уваги акторів; створення відчуття оперативності при подачі контенту.

Крок 5. Виявлення прихованої теми контенту. Під темою контенту СІС будемо розуміти його основний зміст, який автор доносить до читача. Для автоматизації процедур встановлення прихованої теми текстового контенту найбільш ефективними є методи ймовірнісного тематичного моделювання [6]. Такі методи використовують для аналізу колекції документів і вилучають з них теми, зв'язки між темами та їх зміни у часі. При цьому досліджувані документи розглядаються як набір не пов'язаних між собою слів або *Bag of words*. Для кожної публікації у СІС розраховуються ймовірності її належності до набору тем. Для розв’язку поставленої задачі доцільно застосувати один з таких методів: ймовірнісне латентно-семантичне індексування; приховане розміщення Діріхле; робастна тематична модель.

Крок 6. Розрахунок інформаційної ентропії маніпуляції суспільною думкою в СІС. Розглянуті частинні ознаки маніпуляцій суспільною думкою в СІС пов’язані між собою у вигляді ієрархічногодерева прийняття рішень [6]. На даному кроці виконується оцінка інформаційної ентропії текстового контенту віртуальних спільнот, тобто встановлення рівня невизначеності щодо наявності у контенті прихованого впливу на акторів та порівняння його числового значення із допустимим граничним. Значення інформаційної ентропії зменшується при зростанні частот появи ознак маніпуляцій суспільною думкою акторів СІС. У випадку малих частот прояву ознак маніпуляцій у текстовому контенті СІС інформаційна невизначеність зростає.

Висновок

Запропонована методика виявлення маніпуляцій суспільною думкою у СІС ґрунтується на сучасних методах інтелектуального аналізу текстового контенту – контент-аналізі та методах машинного навчання. Рішення про наявність маніпуляцій у текстовому контенті СІС приймається на основі розрахованого значення ентропії, що забезпечує автоматизацію процедур прийняття рішень, підвищення ефективності та швидкодії процесів моніторингу інформаційного середовища.

Література

[1] Fuchs Chr. Social Media: Critical Introdaction / Christian Fuchs. – Sage, 2013. – 304 p.

[2] Молодецька К. В. Соціальні інтернет-сервіси як суб’єкт інформаційної безпеки держави / К. В. Молодецька // Information technology and security. – 2016. – Vol. 4, Iss. 1(6). – С. 13–20.

[3] Поліщук Ю. Я. Мас медіа як канал маніпулятивного впливу на суспільство / Ю. Я. Поліщук, С. О. Гнатюк, Н. А. Сєйлова // Інформаційна безпека. – 2015. – Т. 21, ч. 3. – С. 301–308.

[4] Грищук Р. В. Основи кібернетичної безпеки : моногр. / Р. В. Грищук, Ю. Г. Даник ; під заг. ред. проф. Ю. Г. Даника. – Житомир : ЖНАЕУ, 2016. – 636 с.

[5] Панченко В. М. Лінгвостатистичні ознаки маніпулювання суспільною свідомістю в засобах масової комунікації / В. М. Панченко // Сучасні інформаційні технології у сфері безпеки та оборони. – 2009. – № 1(4). – С. 81–85.

[6] Молодецька-Гринчук К. В. Методика виявлення маніпуляцій суспільною думкою у соціальних інтернет-сервісах / К. В. Молодецька-Гринчук // Інформаційна безпека. – 2016. – № 4(24). – С. 80–93.