Підхід до виявлення маніпуляцій суспільною думкою у соціальних інтернет-сервісах

Молодецька-Гринчук Катерина
Кафедра КТІМС
Житомирський національний агроекологічний університет
Житомир, Україна
kmolodetska@gmail.com

The approach to detect manipulation of public opinion actors of social networking services based on modern methods of content analysis and machine learning. The developed approach differs from the known information into account uncertainty arising in the case of manipulation and technologies including suggestive. The result is greater efficiency and system performance information security of the state in social networking services.

Ключові слова: соціальні інтернет-сервіси, загрози, інформаційна безпека, суспільна думка, ентропія.

Вступ

Соціальні інтернет-сервіси (СІС) є платформою взаємодії учасників віртуальних спільнот, яких називають акторами [1, 2]. Проте, позитивні комунікаційні характеристики перетворили СІС на інструмент проведення інформаційних операцій проти людини, суспільства держави [1, 3]. Встановлено, що під час взаємодії акторів у СІС виникає ряд психологічних явищ, які створюють передумови для маніпулювання суспільною думкою. Маніпуляції представляють собою спосіб інформаційно-психологічного впливу на акторів у прихованому вигляді для спонукання об’єкта впливу до реалізації заданих дій і досягнення суб’єктом впливу однобічних переваг [2, 3].

Зростання кількості загроз інформаційній безпеці держави у СІС і недостатній рівень розвитку наукових методів автоматизованого їх виявлення призвели до появи протиріччя між проблемами практики і науки. Тому розробка методів своєчасного виявлення ознак інформаційного впливу на акторів СІС є актуальним теоретико-прикладним завданням на шляху забезпечення інформаційної безпеки держави.

Зміст підходу

Розроблений підхід до виявлення маніпуляцій суспільною думкою акторів у СІС не суперечить дослідженням В. М. Панченко [4], ґрунтується на методах контент-аналізу та машинного навчання [3] й полягає в такому.

Крок 1. Виявлення ознак сумнівності викладених у контенті СІС фактів Q_1, який зводиться до наявності посилання на суб'єктивну точку зору F_1, відсутності аргументації F_2, розрахунку частки числових даних F_3, запитальних речень F_4, сумнівних висловлювань F_5.

Крок 2. Визначення емоційного забарвлення контенту СІС Q_2, що полягає у встановленні наявності окличних речень F_6, вигуків F_7, прислівників F_8, вживання лексем емоційного характеру F_9.

Крок 3. Оцінка тональності контенту Q_3 на основі сучасних методів машинного навчання з учителем, без учителя, використанням правил або словників. Вибір конкретного методу зумовлюється складністю поставлених завдань.

Крок 4. Встановлення сенсаційності контенту Q_4 внаслідок підвищення уваги акторів СІС F_{10}, оперативності контенту F_{11} в результаті використання слів для створення атмосфери швидкоплинності й терміновості явищ.

Крок 5. Виявлення прихованої теми контенту Q_5. Для автоматизації процедур встановлення прихованої теми текстового контенту СІС доцільно використати ймовірнісне латентно-семантичне індексування, приховане розміщення Діріхле або робастну тематичну модель. Для подальшої оцінки використовується максимальне значення ймовірності належності документа до однієї із тематик досліджуваних набору контенту СІС.

Крок 6. Розрахунок інформаційної ентропії H_n маніпуляції суспільною думкою в СІС. Зв'язок між частинними ознаками маніпуляцій суспільною думкою в СІС, розглянутими на попередніх кроках, зображено у вигляді ієрархії (рис. 1).

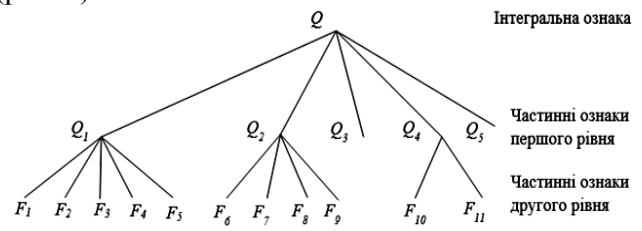

Рис. 1. Дерево рішень

Суть кроку 6 полягає у встановленні рівня невизначеності щодо наявності у контенті прихованого впливу на акторів

$$H = -\sum_{\nu=1}^{k}\sum_{l=1}^{g} Q_l^{\nu} \log_2 Q_l^{\nu} \ .$$

Для зручності інтерпретації розрахованих значень використано нормоване значення ентропії H_n. Його числове значення порівнюється із шкалою оцінки [2] для прийняття рішення про рівень загрози.

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ

Проведено експериментальне дослідження запропонованого підходу виявлення маніпуляцій суспільною думкою у СІС. Для аналізу використано текстовий контент соціальної мережі *ВКонтакте.* Визначення тональності текстового контенту реалізовано на основі мультиноміального наївного методу Байєса, а детектування прихованої тематики – ймовірнісного латентно-семантичного індексування. У результаті розрахунку ентропії частинних ознак (рис. 1) отримано такі числові значення: $H_{Q1}=0,30$, $H_{Q2}=0,40$, $H_{Q3}=0,34$, $H_{Q4}=0,52$, $H_{Q5}=0,52$. Відповідно до запропонованого підходу нормоване значення ентропії дорівнює $H_n=0,67$. Отже, у текстовому контенті виявлено істотні прояви прихованої тематики і сенсаційності, присутні емоційна лексика й тональність контенту. Досліджуваний контент СІС містить загрозу інформаційній безпеці значного рівня, тому вимагає заходів захисту інформаційного середовища [5].

Таким чином, запропонований підхід до детектування маніпуляцій суспільною думкою у СІС відрізняється від відомих використанням сучасних методів аналізу текстового контенту і врахуванням інформаційної невизначеності, яка виникає у разі застосування маніпуляцій і сугестивних технологій зокрема. Розроблений підхід покладено в основу функціонування системи забезпечення інформаційної безпеки держави у СІС, що дозволило підвищити її ефективність і дієвість.

ЛІТЕРАТУРА

[1] Р. Грищук and Ю. Даник, Основи кібернетичної безпеки, 1st ed. Житомир: ЖНАЕУ, 2016.

[2] К. Молодецька, "Методика виявлення маніпуляцій суспільною думкою у соціальних інтернет-сервісах", Інформаційна безпека, no. 4(24), pp. 80–93, 2016.

[3] В. Петрик, М. Присяжнюк, Л. Компанцева, Є. Скулиш, О. Бойко and В. Остроухов, Сугестивні технології маніпулятивного впливу, 2nd ed. Київ: ЗАТ "ВІПОЛ", 2011.

[4] В. Панченко, "Лінгвостатистичні ознаки маніпулювання суспільною свідомістю в засобах масової комунікації", Сучасні інформаційні технології у сфері безпеки та оборони, no. 1(4), pp. 81–85, 2009.

[5] R. Hryshchuk and K. Molodetska, "Synergetic Control of Social Networking Services Actors' Interactions", Recent Advances in Systems, Control and Information Technology, pp. 34-42, 2016.