УДК 004.056:004.738.5

В.О.Дамаскін

ЗАСІБ ДЛЯ ВИЯВЛЕННЯ ДЕЗІНФОРМАЦІЇ НА ОСНОВІ ЕКСПЕРТНИХ ОЦІНОК

Вінницький національний технічний університет

Анотація

Розглянуто підхід до створення програмного засобу для виявлення дезінформації на основі експертних оцінок та методів класифікації інформації. Виконано аналіз сучасних методів протидії інформаційним впливам. Запропоновано структуру програмного засобу, алгоритм агрегування експертних оцінок та використання байєсівського підходу для підвищення достовірності результатів класифікації.

Ключові слова: дезінформація, експертна оцінка, класифікація, інформаційна безпека, достовірність інформації.

Abstract

The paper considers an approach to developing a software tool for detecting disinformation based on expert assessments and classification methods. Modern methods of counteracting information influence are analyzed. The structure of the software tool, the aggregation algorithm of expert assessments, and the application of the Bayesian approach to improve classification reliability are proposed.

Keywords: disinformation, expert assessment, classification, information security, information reliability.

Вступ

У сучасному інформаційному просторі проблема поширення дезінформації набуває критичного значення. Соціальні мережі та електронні медіа забезпечують високу швидкість поширення інформації, що ускладнює її перевірку та підвищує ризики інформаційних впливів.

Існуючі методи виявлення дезінформації базуються переважно на алгоритмах машинного навчання, аналізі текстових характеристик і перевірці джерел. Однак такі підходи не завжди забезпечують достатню точність через складність контекстного аналізу та неоднозначність інформації.

Метою роботи є підвищення достовірності визначення правдивості інформації шляхом поєднання експертного оцінювання та методів класифікації.

Результати дослідження

Запропонований підхід передбачає використання експертних оцінок для аналізу інформаційних повідомлень за рядом критеріїв, зокрема: достовірність джерела, логічна узгодженість, емоційна нейтральність, відповідність перевіреним фактам, рівень сенсаційності та суб'єктивності. Застосування цих критеріїв дозволяє здійснити багатофакторний аналіз інформації та зменшити вплив окремих суб'єктивних оцінок.

Для агрегування оцінок використовується інтегральний показник достовірності:

$$R = \Sigma (w_i \cdot x_i),$$

де w_i — вагові коефіцієнти, що визначають важливість окремих критеріїв, x_i — оцінки за відповідними критеріями. Використання вагових коефіцієнтів дозволяє враховувати різний вплив ознак на кінцевий результат.

На основі методів класифікації інформації додатково застосовано байєсівський підхід, що дозволяє визначити ймовірність належності повідомлення до класу правдивої або неправдивої інформації:

$$P(y|x) = (P(x|y) \cdot P(y)) / P(x).$$

Використання цього підходу дає змогу враховувати апріорні ймовірності та підвищує точність прийняття рішень у випадках невизначеності.

Для підвищення обґрунтованості прийняття рішень використовується система нормованих ознак, що дозволяє кількісно оцінити характеристики інформаційного повідомлення. Кожна ознака приводиться до інтервалу [0;1], що забезпечує можливість їх порівняння та подальшого агрегування.

До основних ознак оцінювання належать: сумнівність викладених фактів, емоційне забарвлення

контенту, тональність, сенсаційність, наявність прихованого змісту, авторитетність джерела, якість інформації та її актуальність. Формування такого набору ознак дозволяє створити універсальний підхід до аналізу різних типів інформаційних повідомлень.

У межах дослідження прийнято, що середнє значення інтегрального показника для правдивої інформації становить приблизно 0,55, тоді як для дезінформації — близько 0,4. Це дозволяє встановити порогове значення (0,5), при перевищенні якого інформація вважається достовірною, а при меншому значенні — недостовірною.

Додатково враховується, що розподіл значень ознак може підпорядковуватися нормальному закону, що дозволяє використовувати статистичні методи для оцінки точності класифікації. Це створює передумови для подальшого вдосконалення алгоритму шляхом навчання на накопичених даних.

Використання байєсівського підходу також дозволяє адаптувати систему до змін інформаційного середовища, зокрема в умовах активних інформаційних впливів та маніпуляцій. Це забезпечує гнучкість і підвищує ефективність виявлення дезінформації.

Запропонований підхід передбачає можливість накопичення статистичних даних та подальшого навчання системи, що дозволяє підвищувати точність класифікації з часом. Таким чином, поєднання експертного оцінювання та ймовірнісних методів забезпечує більш надійне виявлення дезінформації порівняно з використанням лише одного підходу.

Розроблений програмний засіб реалізує зазначений підхід та включає модулі введення даних, експертного оцінювання, класифікації, розрахунку інтегральних показників і формування звітності, що забезпечує комплексний аналіз інформаційних повідомлень.

Висновки

У роботі запропоновано підхід до виявлення дезінформації, що поєднує експертне оцінювання та статистичні методи класифікації. Використання байєсівського підходу дозволяє підвищити точність прийняття рішень щодо достовірності інформації.

Розроблений програмний засіб забезпечує комплексний аналіз інформаційних повідомлень та може бути використаний у системах моніторингу інформаційного простору.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Литвиненко О. В. Інформаційна безпека держави : підручник. Київ : Центр учбової літератури, 2020.

2. Почепцов Г. Г. Інформаційні війни : монографія. Київ : ВД «Києво-Могилянська академія», 2019.

3. ISO/IEC 27001:2022 Information security management systems.

Дамаскін Володимир Олександрович — студент, Вінницький національний технічний університет, м. Вінниця, e-mail: vovadamaskin4@gmail.com

Науковий керівник: ***Дудатьєв А.В.*** — К.т.н., доцент, Вінницький національний технічний університет, м. Вінниця

Volodymyr Damaskin — Student, Vinnytsia National Technical University, Vinnytsia, e-mail: vovadamaskin4@gmail.com

Supervisor: ***Dudatyev A.V.*** — **PhD (Eng.), Associate Professor, Vinnytsia National Technical University, Vinnytsia.**