

СЕМАНТИЧНІ МЕТОДИ ФІЛЬТРАЦІЇ СПАМУ

UDC 004.658.114

D. Ilchenko, R. Zharovskyi, Ph.D.

SEMANTIC METHODS OF SPAM FILTRATION

Задача фільтрації спаму, являє собою задачу класифікації – визначення належності об'єкта (електронного повідомлення) до одного з заздалегідь виділених класів (спам і «не спам») на підставі аналізу сукупності ознак, що характеризують електронне повідомлення.

Семантичні методи передбачають розпізнавання спам листів за змістом (словосполучення, евристики, статистика) або розпізнавання за зразками листів (за сигнатурами). Більшість семантичних методів це методи з попереднім навчанням. Тому необхідно провести їх початкове навчання, тобто задати базу нормальних і спам листів.

Теорема Байєса покладена в основі багатьох сучасних систем, призначених для роботи в умовах невизначеності. Такі системи дають ймовірнісну оцінку, тому звичайно не замінюють експерта, а лише забезпечують підтримку прийняття рішення.

Віднесення листа до спаму або корисних листів виконується з врахуванням заданого програмістом, адміністратором, користувачем параметру «спамерності» електронного листа. Після ухвалення рішення щодо класифікації листа в базі даних обновляються ймовірнісні бази для слів, які входять до нього, тобто з кожним новим листом фільтр вдосконалюється. В основі фільтра лежить список ознак, за якими проводиться аналіз повідомлення і обчислюється умовна ймовірність спамності за кожного ознакою. Загальна ймовірність спаму повідомлення визначається за одним з методів:

- 1) об'єднуються всі ймовірності за теоремою Байєса;
- 2) ймовірності комбінуються і перевіряються на скільки отримана множина схожа з випадковою (метод Фішера).

Основним недоліком Байєсівського фільтра є припущення, що події, які відповідають наявності того чи іншого слова в електронному листі або повідомленні, є незалежними по відношенню один до одного, тобто всі слова статистично незалежні. Максимальний результат, досягнутий байєсовськими фільтрами складає близько 95% відфільтрованого спаму.

Існують модифікації, які дозволяють збільшити ефективність фільтра:

- метод градуйованої фільтрації «спаму», який забезпечує підвищення якості оцінок даних за рахунок врахування додаткових параметрів – кількості листів, в яких зустрічалися слова певної категорії, використання слів, що вперше зустрілися в листі і не існували до цього в базі, частоти використання слів у листах певної категорії,;
- побудова фільтра на основі багатошарового персептрона, що дозволяє враховувати семантичні зв'язки автоматично.

Перевага нейромережевого підходу перед Байєсовським класифікатором полягає в тому, що не робиться ніяких попередніх припущень про характер небажаних повідомлень, а семантичні зв'язки враховуються автоматично. Малодослідженим залишається питання використання нейромереж, що добре зарекомендували себе в задачах розпізнавання образів, окремим випадком яких є фільтрація спаму.

Таким чином, вважаю розвиток методів спам фільтрації буде в напрямку ймовірнісних методів і штучного інтелекту.