

Орлов Є.В.

к. ф.-м. наук, доцент

Одеський національний економічний університет (Україна)

Шинкаренко В.М.

к. ф.-м. наук, доцент

Одеський національний економічний університет (Україна)

Долгозвяга А. М.

здобувачка першого (бакалаврського) рівня ВО

Одеський національний економічний університет (Україна)

ЗАСТОСУВАННЯ БАЄСОВИХ МЕТОДІВ ДЛЯ КЛАСИФІКАЦІЇ ЕЛЕКТРОННИХ ЛИСТІВ

Баєсові методи в машинному навчанні спираються на теорему Баєса, розроблену Томасом Баєсом у XVIII столітті. Ці методи широко застосовуються для класифікації електронних листів на спам та не спам, оскільки вони характеризуються динамічною адаптацією до потоків листів, стійкістю до невизначеності та широким спектром застосувань.

Одним з найпопулярніших Баєсівських методів класифікації електронних листів є модель наївного Баєса. Ця модель ґрунтується на припущенні про незалежність слів в електронному листі [1]. Тобто, ймовірність появи одного слова не впливає на ймовірність появи іншого. Проте, використання баєсових методів має певні обмеження. Зокрема, вони вимагають наявності достовірної апіорної інформації про ймовірності подій або параметрів моделі. Крім того, баєсові методи можуть бути чутливими до помилок у вхідних даних та мати труднощі з інтерпретацією складних моделей. Також, їх застосування може бути обмежене для великих наборів даних через високу обчислювальну складність.

Розглянемо приклад застосування баєсового класифікатора для визначення, чи є електронний лист спамом чи не спамом, на основі наявності слів "ліки" та "знижка". Нехай у вибірці з 45 листів, 20 класифіковані як спам (1), а 25 - як не спам (0). Кількість входжень слова "ліки" у спам-листах становить 15 з 20, а у не спам-листах - 5 з 25. Для слова "знижка" ці показники складають 5 з 25 у спам-листах та 20 з 25 у не спам-листах. Використаємо баєсовий класифікатор для розрахунку ймовірностей [2]. Нехай X - вектор, що представляє вміст електронного листа, а Y - випадкова змінна, яка визначає клас листа (спам або не спам).

- $P(X_1=\text{ліки} \mid Y=1)$ – ймовірність появи слова «ліки» у спам-листах;
- $P(X_2=\text{знижка} \mid Y=1)$ – ймовірність появи слова "знижка" у спам-листах;
- $P(X_1=\text{ліки} \mid Y=0)$ – ймовірність появи слова "ліки" в не спам-листах;
- $P(X_2=\text{знижка} \mid Y=0)$ – ймовірність появи слова "знижка" в не спам-листах.

Тому, розрахуємо ймовірність входження слів у листи. Воно обчислюються як кількість входжень слова у відповідний клас листів, нехай це буде n , поділена на загальну кількість слів у цьому класі, нехай це буде q .

$$P(X|Y) = \frac{n}{q} \quad (1)$$

Імовірність включення слова "ліки" у спам-листи становить 0,75, тоді як для слова "знижка" ця ймовірність дорівнює 0,25. У той же час, імовірність включення слова "ліки" у не спам-листи складає 0,2, а для слова "знижка" - 0,8. Таким чином, входження обох слів у спам листи ($P(X_1, X_2|Y=1)$) та не спам листи ($P(X_1, X_2|Y=0)$) за теоремою множення ймовірностей двох незалежних подій буде складати 0,15 і 0,2. У вищезгаданому випадку, усього 45 листів у вибірці, з яких 20 класифіковані як спам та 25 як не спам, апіорна ймовірність відповідного класу листів обчислюється як відношення кількості спам-листів до загальної кількості листів у вибірці. Отже, апіорна ймовірність того, що електронний лист у наданій вибірці даних є спамом ($P(Y=1)$), становить приблизно 0,444 або 44,4%. Відповідно, апіорна ймовірність того, що лист є "не спамом" ($P(Y=0)$), становить 0,556 або 55,6%.

Наступним кроком можна розрахувати загальну ймовірність появи вмісту X ($P(X)$), яка є необхідною для подальшої класифікації листа як спам або не спам.

$$\begin{aligned} P(X) &= P(X_1, X_2|Y=1) * P(Y=1) + P(X_1, X_2|Y=0) * P(Y=0) \\ P(X) &= 0,15 * 0,444 + 0,2 * 0,556 \end{aligned} \quad (2)$$

Загальна ймовірність появи вмісту X ($P(X)$) = 0,1778. Таким чином враховуючи ймовірності спостережуваних даних (X_1, X_2) для кожного класу ($Y=1$ та $Y=0$), ми можемо обчислити апостеріорні ймовірності, використовуючи формулу Баєса для розрахунку ймовірностей того, що електронний лист буде класифікований як спам та не спам.

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)} \quad (3)$$

Після обчислення ймовірностей за допомогою Баєсового класифікатора, отримали, що ймовірність класифікації електронного листа як спаму $P(Y=1|X_1, X_2) \approx 0,375$, а ймовірність класифікації як не спаму $P(Y=0|X_1, X_2) \approx 0,625$. Оскільки ймовірність класифікації як не спаму перевищує ймовірність класифікації як спаму, вміст електронного листа буде класифікований як не спам [3].

$$P(Y=1|X_1, X_2) < P(Y=0|X_1, X_2) \quad (4)$$

Наведемо відповідний код на Python.

```
import numpy as np
spam_words = input("Введіть слова, які зустрічаються у спам-листах (через пробіл):").split()
spam_counts = [int(count) for count in input(f"Введіть кількість входжень кожного слова '{', '.join(spam_words)}' у спам-листах (через пробіл): ").split())
```

```
non_spam_counts = [int(count) for count in input(f"Введіть кількість входжень  
кожного слова '{', ' '.join(spam_words)}' у не спам-листах (через пробіл): ").split())  
total_spam = int(input("Введіть загальну кількість спам-листів: "))  
total_non_spam = int(input("Введіть загальну кількість не спам-листів: "))  
total_emails = total_spam + total_non_spam  
p_spam = total_spam / total_emails  
p_non_spam = total_non_spam / total_emails  
p_word_given_spam = [count / total_spam for count in spam_counts]  
p_word_given_non_spam = [count / total_non_spam for count in non_spam_counts]  
email_words = input("Введіть слова, які містить електронний лист (через пробіл):  
").split()  
p_spam_given_email = p_spam  
p_non_spam_given_email = p_non_spam  
for word in email_words:  
    if word in spam_words:  
        index = spam_words.index(word)  
        p_spam_given_email *= p_word_given_spam[index]  
        p_non_spam_given_email *= p_word_given_non_spam[index]  
    if p_spam_given_email < p_non_spam_given_email:  
        print("Електронний лист класифікований як не спам")  
    else:  
        print("Електронний лист класифікований як спам")
```

Список використаних джерел

1. Бабенко І. М. Моделі і методи протидії системам фільтрації спаму на основі машинного навчання, 2021. URL: <https://ela.kpi.ua/handle/123456789/56743>.
2. Стеценко О. Б. Аналіз ефективності використання методів класифікації текстів для пошуку спама в смс-повідомленнях, 2020. С. 33 URL: <https://dspace.library.khai.edu/xmlui/bitstream/handle/123456789/986/Stetsenko.pdf?sequence=1>.
3. Кузьма К. Т., Зівенко В. О. Аналіз методів фільтрації електронної пошти від спаму. Геометричне моделювання та інформаційні технології, (1), 2017. С. 84-89.