

*Двигун Д. І.,
Криворізький національний університет
Шаповалова Н. Н.,
ст. викладач, Криворізький національний університет*

ВИКОРИСТАННЯ ПАРАДИГМИ ЙМОВІРНІСНОГО ПРОГРАМУВАННЯ В МАШИННОМУ НАВЧАННІ

Досліджено використання імовірнісного програмування у машинному навчанні, як спосіб визначення найбільш інформативними ознак у наборі даних, на основі баєсівського виводу.

У сучасний час, коли використання машинного навчання все більше впроваджується в різні сфери нашого життя і постійно вимагає все кращих результатів, доводиться ретельніше вивчати наявні проблеми в алгоритмах навчання і шляхи їх уникнення. Кількість різних даних, на яких можна навчати алгоритми, зростає, як і складність цих даних. Зазвичай в даних можуть бути присутні шуми, викиди, може бути безліч ознак і може бути складно визначити чи буде обрана ознака інформативною або, навпаки, вона тільки погіршить якість навчання.

Історично одним із способів допомогти у прийнятті рішень в умовах невизначеності було використання імовірнісної системи міркувань. Імовірнісні міркування об'єднують наші знання про ситуацію з законами імовірності для визначення тих неспостережуваних чинників, які мають вирішальне значення для прийняття рішення. Імовірнісне програмування – це спосіб створення систем, які допомагають приймати рішення в умовах невизначеності. Багато повсякденних рішень включають судження у визначенні відповідних факторів, які ми не спостерігаємо безпосередньо. В імовірнісному міркуванні створюється модель, яка збирає всі відповідні загальні знання області в кількісному, імовірнісному вираженні [1].

Застосувати імовірнісне програмування в машинному навчанні можна за допомогою наївного баєсівського алгоритму. Наївний баєсівський алгоритм – це алгоритм класифікації, заснований на теоремі Баєса з припущенням про незалежність ознак. Тобто, він

передбачає, що будь-будь яка ознака в класі не пов'язана з наявністю якоїсь іншої ознаки. Наприклад, фрукт може вважатися яблуком, якщо він червоний, круглий і його діаметр становить близько 8 сантиметрів. Навіть, якщо ці ознаки є залежними одна від одної, або від будь-яких інших ознак, в будь-якому випадку вони вносять незалежний внесок у ймовірність того, що цей фрукт є яблуком. Саме тому алгоритм називається «наївним».

Моделі, засновані на наївному баєсівському алгоритмі, досить прості і корисні у випадку великого набору даних. Не дивлячись на свою простоту, цей метод здатен перевершити деякі складні алгоритми класифікації [2].

Переваги баєсівського підходу:

- наявність апріорного розподілу: можливість кількісної оцінки будь-яких попередніх знань, за відповідними апріорними розподілами;
- кількісна оцінка невизначеності: можливість отримувати повний апостеріорний розподіл про те, наскільки ймовірні різні значення.

Теорема Баєса надає спосіб перегляду існуючих передбачень або теорій (імовірностей поновлення) з урахуванням нових або додаткових доказів. У фінансах теорема Баєса може бути використана для оцінки ризику кредитування потенційних позичальників.

Теорема Баєса заснована на включенні попередніх імовірнісних розподілів для генерації апостеріорних ймовірностей. У Баєсовому статистичному виводі апріорна ймовірність – це ймовірність події до збору нових даних. Апостеріорна ймовірність – це переглянута ймовірність події, що відбулася після врахування нової інформації. Апостеріорна ймовірність розраховується шляхом поновлення апріорної ймовірності з використанням теореми Баєса:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}, \quad (1)$$

де $P(\theta)$ – апріорна ймовірність гіпотези θ , висловлює наші знання про значення параметрів, що передують побудові моделі; $P(y|\theta)$ – апостеріорна ймовірність, являє собою розподіл ймовірностей для параметра в нашій моделі, розраховане з урахуванням апріорних знань і правдоподібності нових; $P(\theta|y)$ – правдоподібність, ймовірність спостерігати дані у при істинності гіпотези θ ; $P(y)$ –

свідощтво, повна ймовірність гіпотези θ . Іншими словами, ймовірність спостереження даних, усереднених по всіх можливих значеннях, які можуть приймати параметри.

Суть формули в тому, що вона дозволяє переставити причину і наслідок: за відомим фактом події обчислити вірогідність того, що вона відбулася за певної причини. Цю формулу також називають формулою зворотної ймовірності [3].

ВИСНОВКИ

Використання статистичного баєсівського виводу дає можливість точно визначити які параметри будуть більш інформативними для алгоритму навчання, на основі ймовірностей кожної ознаки. Також перевагою цього методу є те, що він надає спосіб поновлення ймовірностей.

ЛІТЕРАТУРА

1. Pfeffer A. Practical Probabilistic Programming / Avi Pfeffer., 2016. – 456 с.
2. Bayes' Theorem Definition [Електронний ресурс] – Режим доступу до ресурсу: <https://www.investopedia.com/terms/b/bayes-theorem.asp>.
3. Байесовская статистика и вероятностное программирование на Python [Електронний ресурс] – Режим доступу до ресурсу: <https://nagornyy.me/courses/data-science/bayes/#vviedieniie-v-baiiesovskii-statistichieskii-vyvod>.

*Богун М. С.,
Криворізький національний університет
Шаповалова Н. Н.,
ст. викладач, Криворізький національний університет*

МОДЕЛЬ ВИЯВЛЕННЯ ШАХРАЙСТВА З КРЕДИТНИМИ КАРТКАМИ

Розроблено програмний модуль банківської системи безпеки, здатний виявляти шахрайство з кредитними картками. Модуль