ISSN: 2181-4171

Engineering Problems and Innovations

Vol. 02 Iss. 04 || Dec 27th, 2024

Том - 02 || Вып. - 04

Проблемы инженерии и Инновации



SCAN ME

Fergana - 2024

ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ С ИСПОЛЬЗОВАНИЕМ БАЙЕСОВСКИМ АЛГОРИТМОМ КЛАССИФИКАЦИИ И ПРИМЕНЕНИЕ В МЕДИЦИНСКИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ.

Абдуманонов Ахроржон Адхамжонович

международный медицинский университет “Central Asian Medical University”, доцент

Адхамжонов Мухаммадсолих Ахроржон угли

Ташкентский государственный Транспортный университет, студент

Аннотация: *В статье приведено описание алгоритм Наивного классификатора Байеса и его типы для применения решение медицинских задач диагностики заболевании. Рассмотрены три основных типа классификатора: Gaussian Naive Bayes, Multinomial Naive Bayes и Bernoulli Naive Bayes. Gaussian Naive Bayes подходит для анализа непрерывных физиологических данных, таких как уровень сахара в крови или артериальное давление, и эффективен для обработки данных с нормальным распределением. Multinomial Naive Bayes используется для классификации текстовых данных, таких как описания симптомов и истории болезни, что способствует автоматизации диагностических процессов и улучшению качества медицинских услуг. Bernoulli Naive Bayes оптимален для работы с бинарными данными, например, при анализе наличия или отсутствия симптомов и интерпретации результатов лабораторных тестов. Выбор подходящего типа классификатора зависит от природы данных (непрерывные или дискретные) и специфики задачи, что подчеркивает необходимость предварительного анализа данных. Результаты исследования показывают, что применение Наивного Байеса в медицинских информационных системах может существенно повысить точность диагностики и оптимизировать процесс принятия решений в здравоохранении.*

Ключевые слова: *алгоритм, классификация, машинное обучения, Naive Bayes, медицина*

TRAINING OF INTELLIGENT SYSTEMS USING BAYESIAN CLASSIFICATION ALGORITHMS AND APPLICATION IN MEDICAL INFORMATION SYSTEMS.

Abdumananov Ahrorjon Adkhamjonovich

International Medical University “Central Asian Medical University”, Associate Professor

Adkhamjonov Mukhammadsolih Ahrorjon ugli

Tashkent State Transport University, student

Abstract: *The article describes the Naive Bayes classifier algorithm and its types for application in solving medical problems of disease diagnostics. Three main types of classifier are considered: Gaussian Naive Bayes, Multinomial Naive Bayes and Bernoulli Naive Bayes. Gaussian Naive Bayes is suitable for analyzing continuous physiological*

data, such as blood sugar or blood pressure, and is effective for processing data with a normal distribution. Multinomial Naive Bayes is used to classify text data, such as descriptions of symptoms and medical history, which helps to automate diagnostic processes and improve the quality of medical services. Bernoulli Naive Bayes is optimal for working with binary data, for example, when analyzing the presence or absence of symptoms and interpreting laboratory test results. The choice of an appropriate classifier type depends on the nature of the data (continuous or discrete) and the specifics of the problem, which emphasizes the need for preliminary data analysis. The results of the study show that the use of Naive Bayes in medical information systems can significantly improve the accuracy of diagnosis and optimize the decision-making process in healthcare..

Keywords: *algorithm, classification, machine learning, Naive Bayes, medicine.*

Введение. Байесовский алгоритм классификации — это один из методов машинного обучения, основанный на применении теоремы Байеса. В её основе лежит вероятностная модель, которая определяет вероятность принадлежности объекта к одному из классов на основе известных данных. Один из наиболее популярных вариантов байесовских классификаторов — это Наивный Байесовский классификатор.

Теорема Байеса выражает связь между априорной вероятностью события и его апостериорной вероятностью с учётом наблюдений:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Где:

- P(C|X) — апостериорная вероятность того, что объект принадлежит классу C, при условии, что наблюдается X.
- P(X|C) — вероятность наблюдения признаков X при условии, что объект принадлежит классу C.
- P(C) — априорная вероятность класса C.
- P(X) — полная вероятность наблюдения признаков X (нормализующий коэффициент).

Алгоритм называется "наивным", потому что он предполагает независимость признаков друг от друга, что в реальных данных не всегда соответствует действительности. Однако, несмотря на эту "наивность", алгоритм показывает хорошие результаты на многих задачах классификации [1-3].

Шаги работы наивного Байесовского классификатора:

Обучение:

- Для каждого класса C вычисляется априорная вероятность P(C), то есть доля объектов, принадлежащих этому классу.

- Для каждого признака Xi и для каждого класса C вычисляется условная вероятность P(Xi|C).

Классификация:

- Для каждого нового объекта вычисляется вероятность того, что объект принадлежит каждому из классов Cj с учётом признаков объекта:

$$P(C_j|X) \propto P(C_j) \prod_i P(X_i|C_j)$$

Класс с максимальной апостериорной вероятностью выбирается как результат классификации.

При решение интеллектуальных задач используется следующие типи наивных байесовских классификаторов: Gaussian Naive Bayes — используется для данных с нормальным распределением. Multinomial Naive Bayes — подходит для дискретных данных (например, для текстовой классификации). Bernoulli Naive Bayes — применяется для бинарных данных. Этот метод, несмотря на свою простоту, может быть очень эффективен для начальных этапов решения задач классификации и часто используется как базовая модель.

Сравнение различных типов Наивных Байесовских классификаторов (Gaussian, Multinomial и Bernoulli) для решения задач медицинской классификации заболеваний важно для выбора наиболее подходящего подхода в зависимости от характера данных. Эти алгоритмы широко применяются в медицине для задач классификации заболеваний, предсказания рисков и диагностики на основе симптомов или биомедицинских показателей [4-7].

Gaussian Naive Bayes (Гауссовский наивный Байес) предполагает, что признаки имеют нормальное (гауссово) распределение. Он используется для данных, которые являются непрерывными, такими как физиологические показатели: температура тела, уровень сахара в крови, артериальное давление. Пример использования в медицинских данных:

- Классификация уровня риска сердечно-сосудистых заболеваний на основе физиологических показателей (возраст, артериальное давление, уровень холестерина и т.д.).

- Диагностика диабета или гипертонии.

Преимущества:

- Хорошо работает на непрерывных данных с нормальным распределением.

- Быстрая производительность и низкая сложность.

Недостатки:

- При нарушении предположения о нормальном распределении может значительно снижать точность.

Multinomial Naive Bayes используется для дискретных или счетных данных. Чаще всего его применяют для текстовых данных, таких как анализ медицинских записей пациентов, или при работе с категорическими признаками (например, частота симптомов, оценки в баллах). Пример использования в медицинских данных:

- Классификация текстовых медицинских записей для автоматического выделения пациентов с определёнными заболеваниями.

- Анализ частоты симптомов у пациентов для предсказания заболевания (например, классификация по анамнезу).

Преимущества:

- Подходит для задач, связанных с дискретными данными, особенно в медицинских текстах или диагнозах на основе категориальных данных.

- Эффективен в задачах медицинской обработки естественного языка (NLP).

Недостатки:

- Модель плохо работает на непрерывных данных без предварительной обработки (требуется дискретизация).

- Чувствителен к редким и неравномерным данным.

Bernoulli Naive Bayes применяется к бинарным данным, где каждый признак может принимать значение «0» или «1» (например, наличие или отсутствие определённых симптомов). Этот метод полезен для медицинских данных, где важно учитывать присутствие или отсутствие признаков. Пример использования в медицинских данных:

- Классификация наличия/отсутствия симптомов (например, кашель, лихорадка, боль в горле) для предсказания гриппа или COVID-19.

- Диагностика болезней на основе лабораторных тестов, где результаты могут быть бинарными (положительный/отрицательный).

Преимущества:

- Подходит для данных с бинарными признаками (например, присутствие симптомов).

- Хорошо работает при анализе данных с отсутствующими значениями или шумом.

Недостатки:

- Может быть менее эффективным для данных с непрерывными или многозначными признаками без предварительной бинаризации.

Сравнение моделей для решения медицинских задач

Критерий	**Gaussian Naive Bayes**	**Multinomial Naive Bayes**	**Bernoulli Naive Bayes**
Тип данных	Непрерывные данные (физиологические показатели)	Дискретные или частотные данные (тексты, категории)	Бинарные данные (симптомы: есть/нет)
Пример задачи	Диагностика на основе физиологических показателей	Текстовая классификация медицинских записей	Предсказание наличия/отсутствия болезни по симптомам
Точность	Высокая для нормально распределённых данных	Высокая для дискретных данных, например, тексты	Высокая для бинарных признаков
Предвари-тельная обработка данных	Не требуется, если данные имеют нормальное распределение	Требует дискретизации непрерывных данных	Требуется бинаризация данных
Применимость	Кардиология, диабет, анализ биохимии крови	Анализ медицинских текстов, истории болезни	Диагностика на основе симптомов и тестов
Проблемы	Нарушение нормальности снижает эффективность	Неэффективен для непрерывных данных без преобразования	Ограничен бинарными данными

Пример кода для сравнения моделей на медицинских данных (данные по диабету)

python

```
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB, MultinomialNB, BernoulliNB
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import Binarizer, MinMaxScaler
from sklearn.datasets import load_diabetes
# Загрузка данных по диабету (непрерывные признаки)
```

```
data = load_diabetes()
X, y = data.data, data.target
y = (y > y.mean()).astype(int)  # Бинаризация метки (диабет: да/нет)
# Разделение на тренировочную и тестовую выборки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
# Масштабирование для MultinomialNB и BernoulliNB
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
# Бинаризация для BernoulliNB
binarizer = Binarizer()
X_train_binarized = binarizer.fit_transform(X_train_scaled)
X_test_binarized = binarizer.transform(X_test_scaled)
# GaussianNB
gnb = GaussianNB()
gnb.fit(X_train, y_train)
y_pred_gnb = gnb.predict(X_test)
print(f"GaussianNB Accuracy: {accuracy_score(y_test, y_pred_gnb):.2f}")
# MultinomialNB
mnb = MultinomialNB()
mnb.fit(X_train_scaled, y_train)
y_pred_mnb = mnb.predict(X_test_scaled)
print(f"MultinomialNB Accuracy: {accuracy_score(y_test,
y_pred_mnb):.2f}")
# BernoulliNB
bnb = BernoulliNB()
bnb.fit(X_train_binarized, y_train)
y_pred_bnb = bnb.predict(X_test_binarized)
print(f"BernoulliNB Accuracy: {accuracy_score(y_test, y_pred_bnb):.2f}")
```

Выводы:

- Gaussian Naive Bayes — лучший выбор для медицинских данных, связанных с непрерывными физиологическими показателями (например, уровень сахара, артериальное давление).
- Multinomial Naive Bayes — применим для текстовых медицинских данных, таких как классификация симптомов или истории болезни.

- Bernoulli Naive Bayes — подходит для бинарных данных, таких как наличие/отсутствие симптомов или результаты лабораторных тестов.

Выбор модели зависит от природы данных: непрерывные или дискретные, а также от задачи классификации в медицинском контексте.

Литературы

1. Золина, Е. В., & Гамова, Н. А. (2019). Наивный классификатор Байеса для решения задачи сентимент-анализа тестов. Шаг в науку, (4), 140-142
2. Сысоев, Ю. И., Шиц, Д. Д., Пучик, М. М., Приходько, В. А., Идиятуллин, Р. Д., Котельникова, А. А., & Оковитый, С. В. (2022). Применение наивного байесовского классификатора для оценки влияния антипсихотических средств на параметры биоэлектрической активности головного мозга у крыс. Российский физиологический журнал им. ИМ Сеченова, 108(7), 874-889
3. Розыходжаева, Д. А. (2017). Алгоритм модели Наивного Байесовского классификатора для диагностики диастолической дисфункции левого желудочка. Наука и образование сегодня, 6 (17), 27-28
4. Абдуманонов А.А. (2020). Касалликларни ташхислашда билимларни шаклланитиришни продукцион модели, Scienceweb academic papers collection, Scienceweb-National database of scientific research of Uzbekistan
5. Abdumananov A.A. (2022). Intellectualization of medical diagnostics // Ахборот-коммуникация технологиялари ва телекоммуникациялари сохасида замонавий муаммолар ва ечимлар республика илмий-техник анжумани, 1(1), 47-49
6. Abdumananov A.A. (2022). Yurak-qon tomir kasalliklarni tibbiy tashxislashda sun’iy neyron tarmoqlardan foydalanish // Klinik va profilaktik tibbiyot jurnali, 1(1), 211-220
7. Abdumananov A.A. (2022). Artificial intelligence in medical diagnostics. Acta CAMU, (1), 89-105

Научно-технический журнал “Проблемы инженерии и Инновации”

Адрес редакции:
150100, Ферганская област, г. Фергана, улица Солим, дом 23/4, кв. 26

E-mail:
chief_editor@epai.uz

Наш сайт:
https://epai.uz/

Учредителем научно-технического журнала
«Engineering problems and innovations» является ООО “Fer-Teach”.
Журнал издается с 2023 года, зарегистрирован 28 декабря 2022 года Агентством информации и массовых коммуникаций при Администрации Президента Республики Узбекистан (Сертификат № 055940).

ISSN: 2181-4171.

«Инженерные проблемы и инновации» является специализированным журналом в области техники и охватывает все основные направления исследований и разработок в различных областях технических наук: физике, математике, информационных технологиях. В журнале также публикуются научные стати по лингвистике, методике преподавания и истории развития технических наук.

Языки издания: Русский, английский