

Документация

Проект: Угадывание случайных человеческих чисел алгоритмами

Автор проекта: Аль-Гураби Тимур

Цифровая школа МГПУ

Оглавление

1. Введение

1.1. Актуальность работы

1.2. Цель и задачи проекта

1.3. Практическое значение проекта

2. Выполнение работы

2.1. Что такое Перцептрон

2.2. Что такое RNN

2.3. Что такое LSTM

2.4. Что такое ARIMA

2.5. Что такое Random Forest

2.6. Что такое XGBoost

2.7. Что такое Цепи Маркова

3. Выводы

4. Список литературы

Введение

1.1. Актуальность работы

Проект направлен на исследование алгоритмов машинного обучения для решения задачи угадывания случайных чисел, которые вводятся пользователями. Несмотря на то, что человеческий выбор чисел считается случайным, он зачастую подчиняется определённым паттернам. Исследование таких паттернов важно для улучшения алгоритмов предсказания в различных сферах, таких как безопасность, маркетинг и анализ пользовательского поведения.

1.2. Цель проекта

Исследовать способность различных алгоритмов машинного обучения угадывать числа, вводимые человеком, с помощью анализа их возможных закономерностей.

1.3. Задачи проекта

1. Изучить работу ключевых алгоритмов машинного обучения: Перцептрон, RNN, LSTM, ARIMA, Random Forest, XGBoost и Цепи Маркова.
2. Реализовать модели предсказания случайных чисел.
3. Сравнить эффективность разных алгоритмов по точности предсказаний.
4. Проанализировать поведение пользователей и выявить закономерности в их выборе.

1.4. Практическое значение проекта

Результаты проекта могут быть использованы для:

- создания более точных предсказательных моделей;
- улучшения пользовательского опыта в системах рекомендаций;
- оптимизации алгоритмов безопасности, связанных с поведением человека.

Выполнение работы

2.1. Что такое Перцептрон

Перцептрон — это простейшая нейронная сеть, способная решать линейно разделимые задачи. Он состоит из одного слоя и применяется для базового анализа данных. В данном проекте перцептрон используется для начального тестирования предсказаний случайных чисел.

2.2. Что такое RNN

Recurrent Neural Network (RNN) — рекуррентная нейронная сеть, способная учитывать последовательности данных. Она хорошо подходит для задач анализа временных рядов, что полезно при предсказании последовательности введённых пользователем чисел.

2.3. Что такое LSTM

Long Short-Term Memory (LSTM) — разновидность RNN, которая устраняет проблему затухания градиента. Она способна запоминать долгосрочные зависимости в данных, что делает её эффективной для сложных последовательностей.

2.4. Что такое ARIMA

Автокорреляционная интегрированная модель скользящего среднего (ARIMA) — статистический метод анализа временных рядов. Применяется для прогнозирования чисел на основе трендов и сезонности.

2.5. Что такое Random Forest

Random Forest — ансамблевый метод машинного обучения, использующий множество решающих деревьев. В проекте он используется для анализа закономерностей в больших массивах данных.

2.6. Что такое XGBoost

XGBoost — улучшенный алгоритм градиентного бустинга, который оптимизирует производительность моделей. В проекте применяется для построения мощных предсказательных моделей.

2.7. Что такое Цепи Маркова

Цепи Маркова — это стохастические процессы, где вероятность следующего состояния зависит только от текущего состояния. Они используются для анализа вероятностных паттернов в последовательностях.

Выводы

В ходе работы были исследованы и протестированы алгоритмы машинного обучения для предсказания случайных чисел, вводимых пользователями. Было выявлено, что:

- LSTM и XGBoost показали наилучшую точность предсказания.
- Перцептрон и RNN продемонстрировали базовый уровень эффективности, но не справились с более сложными задачами.
- Цепи Маркова и ARIMA лучше всего подходят для анализа последовательных паттернов.

Данные результаты позволяют заключить, что человеческий выбор чисел не является полностью случайным, и машинное обучение может успешно выявлять определённые закономерности.

Список источников

1. Goodfellow I., Bengio Y., Courville A. Deep Learning. — MIT Press, 2016.
2. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. — Springer, 2009.
3. https://github.com/Ttimurttt/Random_numbers
4. Bishop C. Pattern Recognition and Machine Learning. — Springer, 2006.
5. Маккинли Д., Чодори М. Цепи Маркова: Теория и применение. — Москва: Физматлит, 2014.