

ПРОГНОЗИРОВАНИЕ ТРЕНДОВ В РАЗРЕЖЕННЫХ ВРЕМЕННЫХ РЯДАХ С ПРИМЕНЕНИЕМ НЕЙРОННЫХ СЕТЕЙ

Улизько М.В. (Университет ИТМО), Добренко Д.А. (ГУП «Петербургский Метрополитен»)
Научный руководитель – доцент, кандидат технических наук, Гусарова Н.Ф.
(Университет ИТМО)

Введение. Потребность в прогнозе разреженных временных рядов имеется во всех сферах научных исследований. При этом, традиционные модели и инструменты показывают плохой результат. Существует ряд возможных причин низкой производительности нейронных сетей в задаче прогнозирования временных рядов. В рассматриваемых работах авторы сводят задачу прогнозирования к линейной регрессии. Классические модели, такие как ARIMA доказали свою эффективность, но не могут быть использованы на малом количестве данных. В том числе, стоит проблема предобработки данных и перестройки модели под новые данные [4].

Основная часть. Для прогноза разреженных временных рядов предлагается использовать Байесовскую нейронную сеть. Байесовская сеть – модель нейронной сети, где связь между парой нейронов задается вероятностью по формуле Байеса [3]. Байесовские нейронные сети используют вероятностные слои, которые фиксируют неопределенность в отношении весов и активаций, и обучаются с использованием байесовского вывода. Байесовские сети относятся к группе алгоритмов, описывающих формализованные взаимосвязи [2]. В рамках интеллектуального анализа данных с помощью формулы Байеса можно более точно рассчитать вероятность, используя как ранее известную информацию, так и данные новых наблюдений. Рассмотрим главные преимущества Байесовской сети:

1. Значения неизвестных переменных обрабатываются значительно быстрее при наличии заданных зависимостей в модели между переменными;
2. Байесовские нейронные сети являются легко интерпретируемыми на этапе прогностического моделирования.

Для оценки качества работы и сравнения производительности с другими моделями был использован датасет с Kaggle «Student Grade Prediction». В нем собраны данные об учениках двух португальских школах. В качестве целевой переменной предлагается использовать G3 – оценку за финальный экзамен. Был произведен предварительный анализ данных и обработка датасета. Исходный набор данных содержал 398 записей. Далее набор данных был разрежен. Крайним значением для получения релевантных результатов стало 39 записей. В качестве оценки было произведено тестирование на базовых алгоритмах а затем на Байесовской сети. В качестве оценки использовались метрики MAE, RMSE и log like.

Выводы. На разреженном наборе данных была применена и настроена Байесовская нейронная сеть. Размерность датасета была уменьшена в 10 раз, при этом потеря в качестве прогнозирования составила 7%.

Список использованных источников:

1. Chimmula, Vinay Kumar Reddy, and Lei Zhang. "Time series forecasting of COVID-19 transmission in Canada using LSTM networks." // Chaos, Solitons & Fractals 135 (2020): 109864.
2. Кустицкая, Т. А. (2019, September). Прогнозирование успешности обучения студента с помощью байесовской сети. // Информатизация образования и методика электронного обучения: Материалы III Междунар. науч. конф.–Красноярск: Сибирский федеральный университет (pp. 257-262).
3. Wang, R., Ye, S., Li, K. and Kwong, S., 2021. Bayesian network-based label correlation analysis for multi-label classifier chain. // Information Sciences 554 (pp. 256-275).

4. Kareem, S.W., Alyousuf, F.Q.A., Ahmad, K., Hawezi, R. and Awla, H.Q., 2022. Structure Learning of Bayesian Network: A Review // QALAAI ZANIST JOURNAL, 7 (pp.956-975)

Улизько М.В. (автор)

Подпись

Гусарова Н.Ф. (научный руководитель)

Подпись