ОБЗОР МЕТОДОВ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

REVIEW OF METHODS FOR FORECASTING TIME SERIES USING ARTIFICIAL NEURAL NETWORKS

УДК 004

Дауб Игорь Сергеевич, Магистрант кафедра - "Прикладная математика, информатика и информационные системы" СамГУПС

Научный руководитель: Фатеев Владимир Алексеевич, доцент, кандидат технических наук Кафедра - "Прикладная математика, информатика и информационные системы". СамГУПС

Daub Igor Sergeevich, Master's student of the department - "Applied Mathematics, Informatics and Information Systems" Samara State Transport University, igordaub@gmail.com

Academic Supervisor: Fateev Vladimir Alekseevich, Associate Professor, Candidate of Technical Sciences Department - "Applied Mathematics, Informatics and Information Systems" Samara State Transport, University v_a_fateev@mail.ru

Аннотация

Нейросети обладают широким спектром возможностей. Они способны решать задачи классификации, кластеризации, распознавания образов, анализ паттернов, принятие решений, прогнозирование. В данной статье будут рассмотрены некоторые способы прогнозирования временных рядов с помощью искусственных нейронных сетей.

Целью статьи является обзор методов прогнозирования временных рядов с помощью искусственных нейронных сетей (ИНС). В качестве исходных

данных был сгенерирован временной ряд, обладающий трендом, сезонностью и шумом. Это приводит к тому, что испытуемые методы будут поставлены в одинаковые условия. В данной статье нейросетевые модели прогнозирования будут применяться к одинаковому временному ряду без его предварительной обработки.

Summary

Neural networks have a wide range of capabilities. They are able to solve problems of classification, clustering, pattern recognition, pattern analysis, decision making, forecasting. This article will explore some of the ways to predict time series using artificial neural networks.

The purpose of the article is to review methods for forecasting time series using artificial neural networks (ANN). A time series with trend, seasonality and noise was generated as input data. This leads to the fact that the tested methods will be placed in the same conditions. In this article, neural network prediction models will be applied to the same time series without preliminary processing.

Ключевые слова: методы прогнозирования временных рядов, передача данных, нейронная сеть.

Key words: methods for forecasting time series, data transmission, neural network.

Для обзора методов прогнозирования с помощью ИНС необходимо разбить временной ряд на три части:

- Обучающий период
- Валидационный период
- Тестовый период

Если временной ряд содержит сезонность, необходимо, чтобы каждый период содержал целое число сезонов. После завершения подстройки модели, перед началом ее использования в производстве, необходимо обучить ее на всем временном ряду, включая тестовый период. Это отличается от обычной практики

машинного обучения, где мы никогда не обучаем модель на тестовом периоде, однако данный шаг необходим, поскольку самые поздние показания во временном ряду содержат самую важную информацию для прогнозирования.

Для передачи данных нейросетевой модели, необходимо разбить временной ряд на окна фиксированной длины и заставить модель предсказывать следующее значение после окончания окна. Обычно для передачи данных в таком формате они должны быть фиксированной длины, поэтому окна, в которых количество значений недостаточно, не используются. Для данной статьи длина окна будет выбрана равной тридцати временным шагам. Перемешанными временными окнами будет называться такой способ передачи временных окон, когда они передаются без сохранения изначального порядка.

1. Метрики производительности

Для оценки производительности моделей прогнозирования будут использованы следующие метрики:

- Errors = прогнозные значения – действительные значения
- MSE (Mean Squared Error) = $\frac{1}{n}\sum(\text{Прогноз} - \text{Действительныезначения})^2$
- RMSE (Root Mean Squared Error) = $\sqrt{MSE}$
- MAE (Mean Absolute Error) = $\frac{1}{n}\sum|\text{Прогноз} - \text{Действительныезначения}|$
- MAPE (Mean Absolute Percentage Error) = $\frac{1}{n}\sum\frac{|\text{Прогноз}-\text{Действительныезначения}|}{\text{Действительныезначения}}$

В случаях, когда большие ошибки важны и они могут привести к серьезным последствиям, следует уделять больше внимания MSE. В том случае, когда приоритетом обладают малые ошибки, следует уделить внимание MAE.

2. Линейная регрессия и полносвязные модели

В качестве модели линейной регрессии будет представлен 1 полносвязный слой с 1 нейроном без функции активации.

В модель передавались перемешанные окна фиксированной длины и были получены следующие результаты:

- MAE = 5.3433857
- MAPE = 6.8779497
- MSE = 53.180916
- RMSE = 7.292525

Для полносвязной модели будет использована последовательная модель с двумя полносвязными слоями, имеющими 10 нейронов в каждом и функцией активации relu, а также последний полносвязный слой с одним нейроном и без функции активации.

В модель передавались перемешанные окна фиксированной длины и были получены следующие результаты:

- MAE = 5.2994733
- MAPE = 6.9079204
- MSE = 52.995438
- RMSE = 7.2797966

3. Простая рекуррентная нейронная сеть

Рекуррентные нейросети – обобщение нейронной сети прямого распространения, которая имеет внутреннюю память. Такая нейросеть выполняет одну функцию для каждого входа данных, в то время как выходное значение текущего входа зависит от прошлых вычислений. После получения выходного результата, он копируется и отправляется назад в рекуррентную нейросеть. Для принятия решения она принимает во внимание текущие входные данные и выходные данные, которые она узнала на предыдущем входе.

В отличие от сетей прямого распространения, рекуррентные сети могут использовать их внутреннее состояние для обработки последовательности входов. Это делает их подходящими для таких задач как распознавание рукописного текста, распознавание речи. В других нейросетях все входы

независимы друг от друга, но в рекуррентных все сходы зависят между собой. [1]

Рекуррентные нейросети могут быть лучшим инструментом, когда имеется огромный объём данных и высокое соотношение сигнал/шум. Но данное утверждение не всегда верно. Одной из проблем данного типа нейросетей является «исчезающий» или «взрывающийся» градиент ошибки.

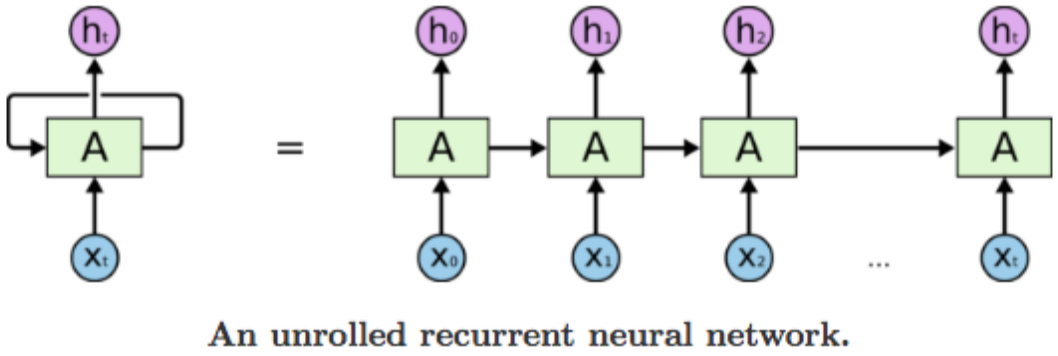


Рисунок 1 – «Развернутая» рекуррентная нейросеть

3.1. Рекуррентная нейронная сеть без сохранения состояния

Рекуррентная нейронная сеть без сохранения состояния - это тип рекуррентной нейронной сети, в которой окна с данными передаются смешанно и могут пересекаться друг с другом. В конце каждого окна будет образовываться вектор состояния, который будет сброшен и не получит дальнейшего применения. Из-за данной особенности будет проблематично заставить рекуррентную нейронную сеть без сохранения состояния запомнить длительные паттерны временного ряда (максимум – на несколько дюжин временных шагов).

Использовалась нейросеть с двумя рекуррентными слоями по 100 нейронов в каждом. В модель передавались перемешанные окна фиксированной длины и были получены следующие результаты:

- MAE = 5.13363
- MAPE = 6.5848627
- MSE = 49.19472
- RMSE = 7.013895

3.2. Рекуррентная нейронная сеть с сохранением состояния:

В рекуррентной нейронной сети с сохранением состояния окна с данными должны передаваться упорядоченно и не пересекаться. На выходе одного окна будет образовываться вектор состояния, который не будет сброшен сразу. Он будет передан следующей итерации и будет сброшен в конце эпохи. Это позволит сети получить имеющий смысл вектор состояний в начале каждой итерации и запомнить более долгие паттерны. Она сложнее в реализации и обучение проходит намного дольше.

Использовалась нейросеть с двумя рекуррентными слоями по 100 нейронов в каждом. В модель передавались не перемешанные, непересекающиеся окна фиксированной длины и были получены следующие результаты:

- MAE = 6.277829
- MAPE = 7.9098797
- MSE = 72.34238
- RMSE = 8.505432

4. LSTM

Long Short-Term Memory - техника для предоставления возможности рекуррентной нейросети запоминать долгосрочные паттерны. LSTM нейросеть имеет более долгую память – способна запомнить паттерны, длиной более 100 временных шагов. Тем не менее, с более долгими паттернами справиться ей будет сложно. Также была решена проблема исчезающих градиентов, которая является острой при работе с обычными рекуррентными нейронными сетями и требует для ее решения некоторых модификаций. LSTM сети хорошо подходят для задач классификации, обработки и прогнозирования временных рядов. [1] «Гейты» в LSTM:

- Forget gate – полносвязный слой с сигмоидной функцией активации. Решает, когда необходимо «забыть» содержимое долгосрочного вектора состояния.

- Input gate – Состоит из простого рекуррентного слоя, а также полносвязного слоя с сигмоидной функцией активации, который решает, когда в долгосрочный вектор должна быть добавлена информация из простой рекуррентной нейросети.
- Output gate – полносвязный слой с сигмоидной функцией активации, который решает, какая часть долгосрочного вектора состояний должна быть передана на следующий шаг.

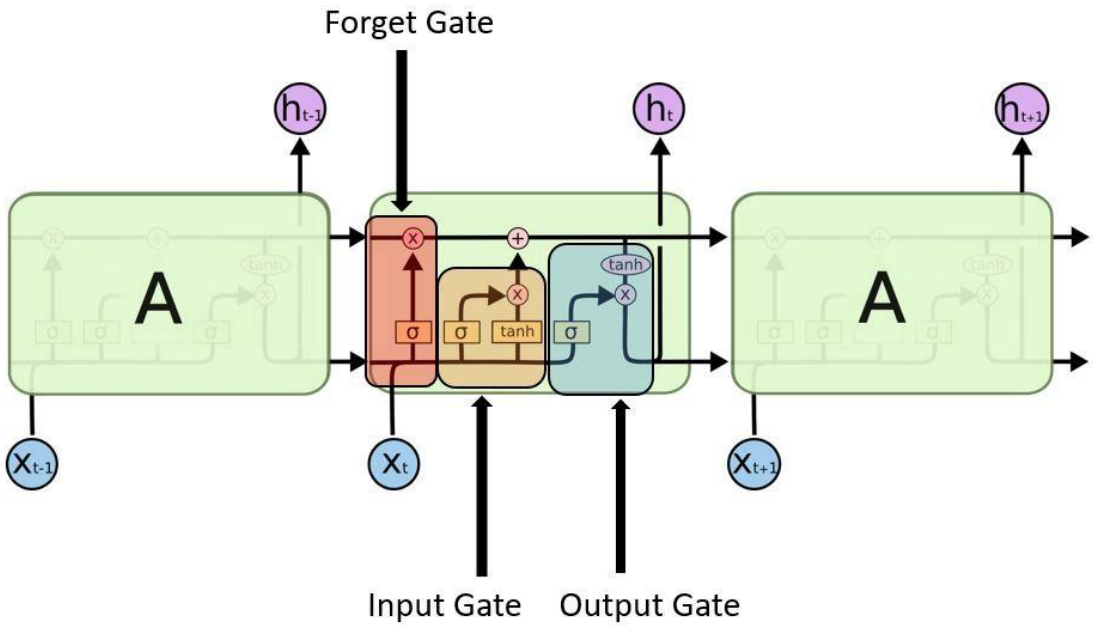


Рисунок 2 – Составляющие LSTM ячейки

Использовалась нейросеть с двумя рекуррентными слоями по 100 нейронов в каждом. В модель передавались не перемешанные и не пересекающиеся окна фиксированной длины и были получены следующие результаты:

- MAE = 5.788751
- MAPE = 7.1823955
- MSE = 63.2326
- RMSE = 7.951893

5. Сверточные нейронные сети

Сверточные нейронные сети могут быть крайне эффективны для решения задач на подобии классификации изображений, но также они эффективны при работе с последовательностями. В данных нейросетях происходит сдвиг фильтра по временному ряду, получая взвешенную сумму и применяя к ней функцию активации. Сверточные нейросети не обладают памятью, однако при составлении многих слоев получается, что верхние слои видят большую часть последовательности, поэтому сверточные нейросети могут охватывать длительные паттерны.

Можно попытаться с помощью них улучшить результат, используя для предобработки временного ряда сверточные слои. Например, можно попытаться улучшить LSTM сеть из прошлого примера, добавив входной сверточный слой и получить следующие результаты:

- MAE = 5.112641
- MAPE = 6.588705
- MSE = 47.91797
- RMSE = 6.922281

Пример использования сверточных нейросетей для обработки временных рядов – архитектура WaveNet, предложенная DeepMind в 2016 году. Это многослойная сверточная нейросеть, особенностью которой является применение коэффициента расширения. Расширенные свертки – свертки, когда фильтр применяется на площадь, большую, чем его размер из-за пропуска входящих значений с определенным шагом. Многослойные расширительные свертки позволяют сетям иметь очень большое поле восприятия при небольшом количестве слоев, сохраняя при этом разрешение входных данных на протяжении всех сверток, как и вычислительную эффективность. [2] WaveNet увеличивает коэффициент расширения с каждым слоем мультипликативно (т.е. 1,2,4,8…), тем самым мы достигаем экспоненциальной зависимости между

глубиной слоев и полем восприятия. [3] Использование подобных сверток позволяет выходу быть зависимым от большего числа элементов на входе. [4]

Сеть имела шесть слоев с мультипликативным увеличением расширительного коэффициента от 1 до 32х, 32мя фильтрами, размером фильтра = 2, сдвигами на единицу с применением функции активации relu, с применением каузальных отступов для сохранения размера выходной последовательности, при котором в начало последовательности недостающее количество данных заменяется нулями, а также выходным сверточным слоем из фильтра размером 1 и размером фильтра равным 1. В модель передавались перемешанные. Были получены следующие результаты:

- MAE = 4.529327
- MAPE = 5.9493346
- MSE = 38.715195
- RMSE = 6.2221537

Заключение

Данные результаты не являются самыми лучшими для всех рассмотренных архитектур. Все они были реализованы с помощью Keras, настроены на лучшую MAE и был выбран оптимальный коэффициент скорости обучения путем построения графика изменения ошибки от коэффициента скорости обучения. Тем не менее остается большое количество гиперпараметров, которые необходимо подстраивать вручную под каждый конкретный случай – количество слоев и нейронов, оптимизаторы, терпеливость алгоритма обучения, количество эпох и т.д. Необходимо проводить множество экспериментов, наблюдая за тем, как изменяются метрики, в зависимости от внесенных изменений. Также имеется возможность предварительной обработки временного ряда, убрав из него тренд и сезонность, что в данной статье продемонстрировано не было. Тем не менее для такого действия не подойдет обычная операция дифференцирования, поскольку в прогнозировании стационарного ряда нет никакого смысла. Для прогнозирования временного ряда с трендом и сезонностями, без его

предварительной обработки, по рассмотренным метрикам, лучше всего себя проявила сверточная нейросеть, построенная на подобии архитектуры WaveNet. Однако, для каждого конкретного случая, будет разумно провести эксперименты и с рекуррентными нейросетями.

Литература

1. Aditi Mittal Understanding RNN and LSTM [Электронный ресурс]. – Режим доступа :URL: https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e
2. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu WaveNet: A Generative Model for Raw Audio [Электронный ресурс]. – Режим доступа :URL: https://arxiv.org/abs/1609.03499
3. Joseph Eddy Time Series Forecasting with Convolutional Neural Networks - a Look at WaveNet [Электронный ресурс]. – Режим доступа :URL: https://jeddy92.github.io/JEddy92.github.io/ts_seq2seq_conv/
4. Anastasia Borovykh, Sander Bohte, Cornelis W. Oosterlee Conditional Time Series Forecasting with Convolutional Neural Networks [Электронный ресурс]. – Режим доступа :URL: https://arxiv.org/abs/1703.04691

Literature

1. Aditi Mittal Understanding RNN and LSTM [Electronic resource]. - Access mode: URL: https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e
2. Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu WaveNet: A Generative Model for Raw Audio [Electronic resource]. - Access mode: URL: https://arxiv.org/abs/1609.03499

3. Joseph Eddy Time Series Forecasting with Convolutional Neural Networks - a Look at WaveNet [Electronic resource]. - Access mode: URL: https://jeddy92.github.io/JEddy92.github.io/ts_seq2seq_conv/
4. Anastasia Borovykh, Sander Bohte, Cornelis W. Oosterlee Conditional Time Series Forecasting with Convolutional Neural Networks [Electronic resource]. - Access mode: URL: https://arxiv.org/abs/1703.04691