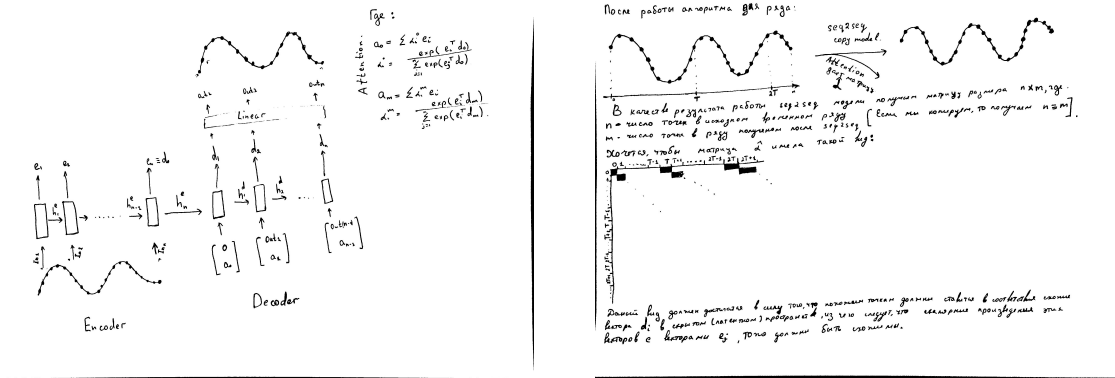


Attention для аппроксимации временных рядов

1 Некоторое введение

Под временным рядом будет предполагать набор упорядоченных точек, которые получены путем наблюдения за некоторым непрерывным процессом с некоторой фиксированной частотой.

2 Модель seq2seq и Attention



(a) Описание seq2seq

(b) Описание предполагаемых результатов

Рис. 1: Основные сведения об seq2seq + Attention

Используем LSTM с простейшим Attention без дополнительных параметров. Пример, принципа работы seq2seq с Attention показан на рис. 1а.

3 Предположения

Введя предположения о том, что скрытые вектора LSTM имеет похожие вектора для похожих кусков временного ряда можно предположить, что матрица Attention будет иметь вид показан на рис. 1b. Под матрицей Attention подразумевается матрица размера $n \times m$, где n, m — длины входного и выходного сигналов соответственно, в ячейках которой показан уровень схожести (либо можно сказать уровень зависимости) между куском временного ряда на выходе и куском временного ряда на входе.

4 Проблемы

Предлагается использовать seq2seq, который просто копирует временной ряд. В этом методе есть ряд вопросов.

- Первый косяк состоит в том, что если мы учим копировать временной ряд, и берем слишком сложный Attention с большим количеством параметров, то он начинает переобучаться, и в итоге получается матрица Attention диагональной. Это следует из того, что модель начинает просто учить друг за другом числа. Поэтому предлагается использовать просто скалярное произведение векторов (самую простую модель Attention, Dot метрика из 1).
- Даже если взять простую модель Attention, все равно можно получить диагональную матрицу Attention, если мы просто будем копировать временной ряд, поэтому предлагается не просто учить все ряды из обучающего множества рядов, а на вход Encoder'а давать некоторый кусок (например из 100 сигналов ряда дать только первые 20 сигналов), и сравнивать как Decoder восстановит весь сигнал. Данный дополнительный финт немного улучшает качество.
- Второй вопрос состоит в том, что нужно придумать на каких данных учить данную модель (нужно подумать над тем какие сигналы давать на вход для копирования). Этот вопрос возникает в синтетических данных, но и в реальных данных. Я думаю, что можно модель учить не только на реальных данных, а и на синтетических, но для этого нужно придумать максимально правдоподобный сигнал акселерометра.

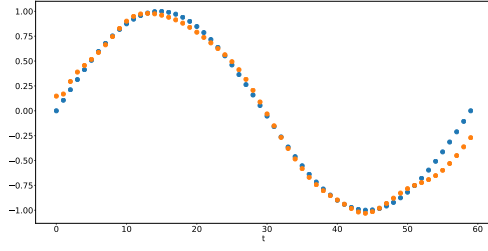
Таблица 1: Описание разных Attention

Type	formula
Additive	$score_{i,j} = \mathbf{w}_o \tanh(\mathbf{W}_e \mathbf{e}_i + \mathbf{W}_d \mathbf{d}_j)$
Dot	$score_{i,j} = \mathbf{e}_i^\top \mathbf{d}_j$
General	$score_{i,j} = \mathbf{e}_i^\top \mathbf{W} \mathbf{d}_j$

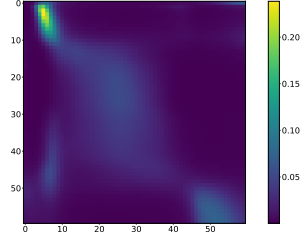
5 Эксперимент

5.1 Эксперимент с простыми периодическими структурами

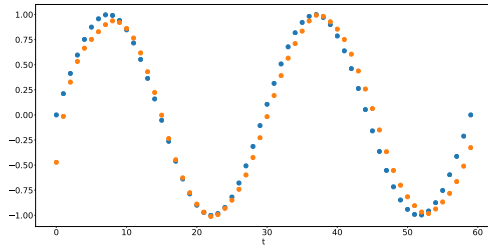
Пусть модель обучена на синусоидальных сигналах с произвольной частотой, произвольной амплитудой и произвольной начальной фазой. Как видно из результатов на рис. 3,



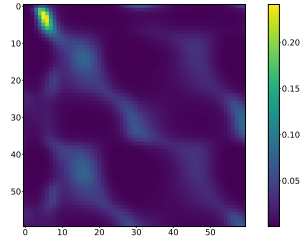
(a) $\sin(x)$



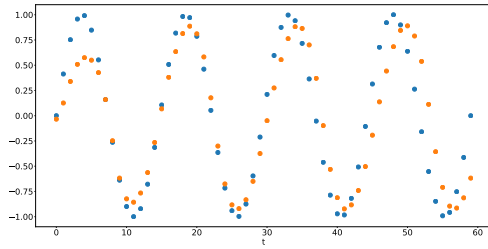
(b) $\sin(x)$



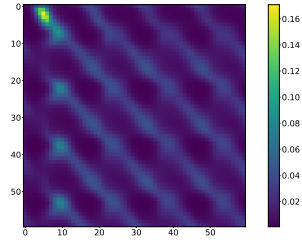
(c) $\sin(2x)$



(d) $\sin(2x)$



(e) $\sin(8x)$



(f) $\sin(8x)$

Рис. 2: Результаты для модели с рис. 1a

предположение о виде матрицы Attention подтверждается. На рис 3 показано как меняется вид матрицы Attention в зависимости от частоты синуса. Видно что количество диагоналей в матрице Attention соответствует частоте синусоидального сигнала.

В работе [1] показано, что для нахождения периодов во временном ряде нельзя использовать простой Attention. Был предложен модифицированный Attention который выглядит в следующем виде:

$$e_{ij} = \mathbf{w}_o \tanh(\mathbf{W}_e(\boldsymbol{\pi}_{i-j}\mathbf{e}_i) + \mathbf{W}_d\mathbf{d}_j) \cdot [i - j \leq T], \quad (1)$$

где $\boldsymbol{\pi}$ — еще один обучаемый вектор параметров, который указывает на степень зависимости i -го и j -го элемента.

Мои эксперименты показали, что прирост качества в двух разных Attention не очень большой, поэтому можно использовать любой из них.

5.2 Эксперимент с простыми переодическими структурами в случае размерности скрытого пространства 2

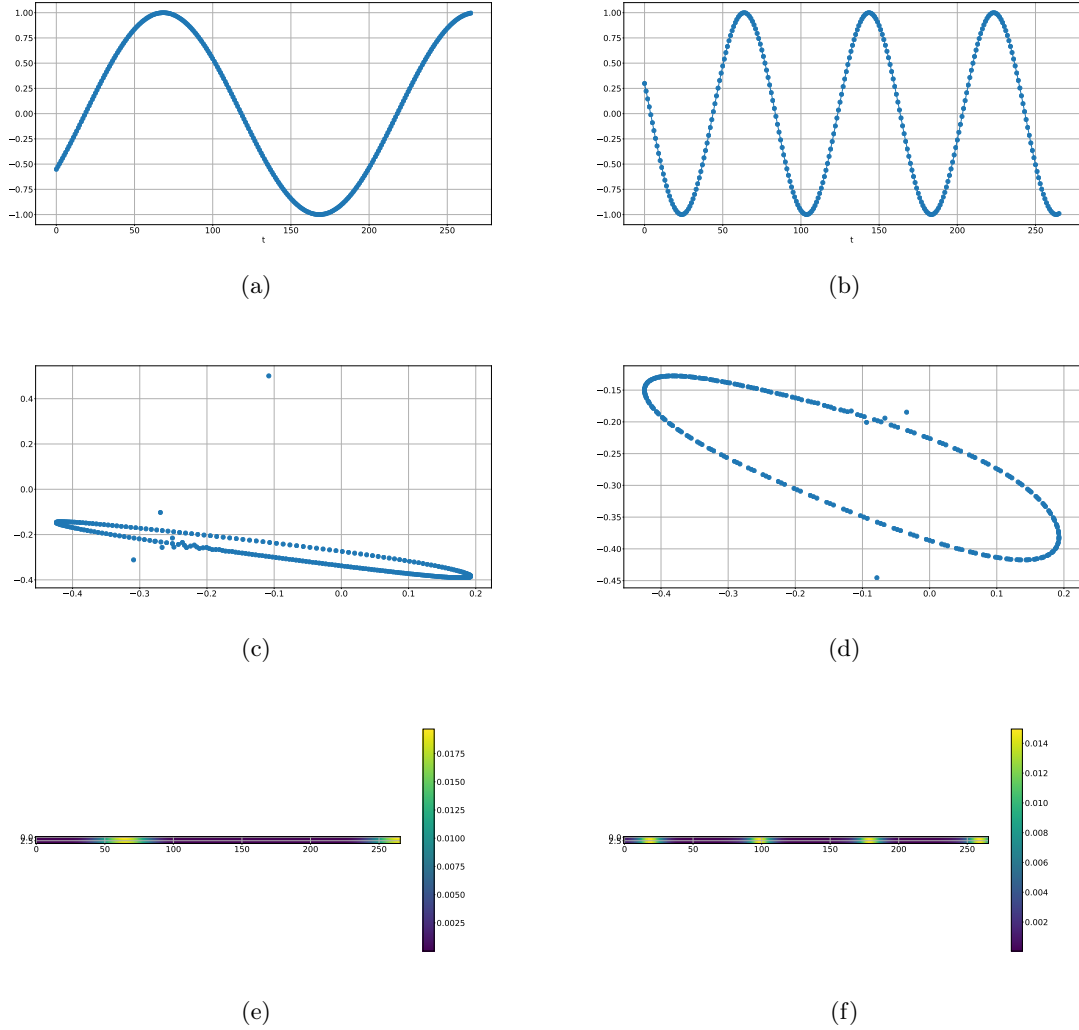


Рис. 3: Результаты для модели с рис. 1a

5.3 Эксперимент с простыми двумя простыми переодическими структурами внутри одного ряда

Задача состоит в том, чтобы при помощи механизма Attention с временного ряда, который состоит из двух переодических сигналов (к примеру человек по очереди идет некоторое время и бежит некоторое время) выделить моменты времени где наблюдается первое действие и выделить моменты времени где наблюдается второе действие.

Предлагается использовать seq2seq модель с Attention.

Сама структура LSTM строит посути фазовое пространство большой размерности. Попробуем обучить seq2seq модель, которая на вход принимает сигнал, который состоит

из двух периодических сигналов, выдать на выход два сигнала, первый сигнал это чистый первый периодический сигнал со входа, а второй сигнал это второй периодический сигнал со входа. То есть другими словами, попробуем, сделать так, чтобы LSTM смогла из фазовому пространству вычленить два разных периодических сигнала. Для определенности будем считать первым периодическим сигналом тот сигнал который во временном ряде встретился раньше.

Как сделать чтобы LSTM по одному входу дала два сигнала???? У этого вопроса есть два возможных ответа:

- первый случай простой, а давайте выход декода будет не число а два числа — значения первого сигнала и второго сигнала в один тик времени (простыми словами у нас есть 2 линейные модели которые по фазовому пространству делают некоторый предикт сигнала)
- второе решение немного сложнее (сейчас пробую сделать его). У нас есть как и раньше один выход LSTM. Но мы можем варировать начальную точку LSTM с которой она начинает генерировать ряд. К примеру если мы хотим получить первый сигнал, что в качестве начальной точки дадим 0 а если хотим сгенерировать второй сигнал, то начальной точкой LSTM дадим 1.

Ничего не получилось.

Список литературы

- [1] *Y. G. Cinar and H. Mirisae* Period-aware content attention RNNs for time series forecasting with missing values // *Neurocomputing*, 2018. Vol. 312. P. 177–186.