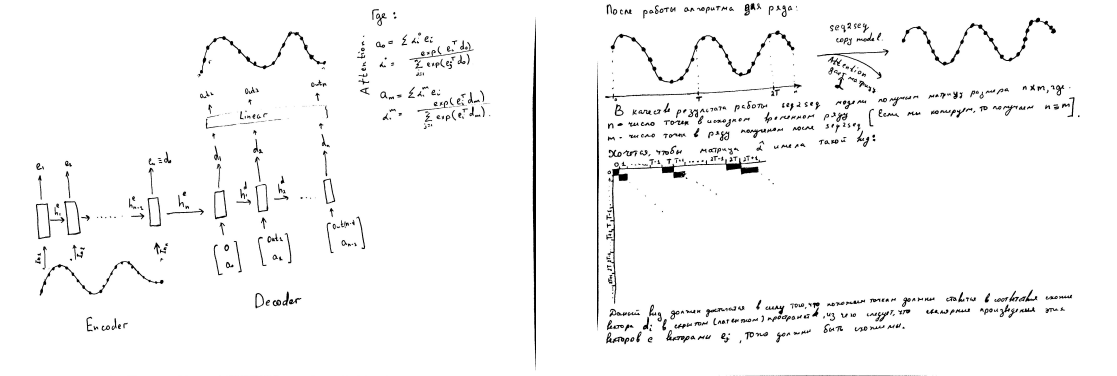


Attention для аппроксимации временных рядов

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.

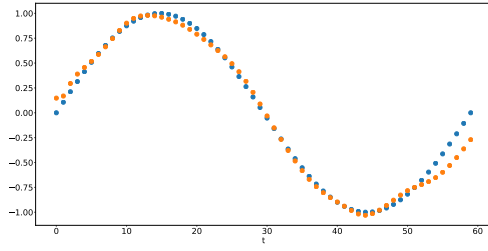
Таблица 1: Описание разных Attention

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

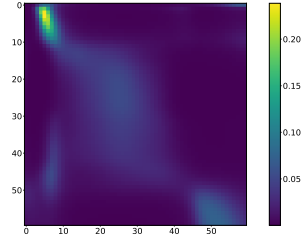
5 Эксперимент

5.1 Эксперимент с простыми периодическими структурами

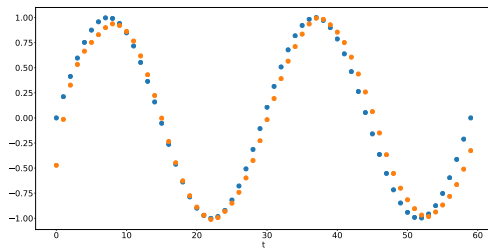
Пусть модель обучена на синусоидальных сигналах с произвольной частотой, произвольной амплитудой и произвольной начальной фазой. Как видно из результатов на рис. 2,



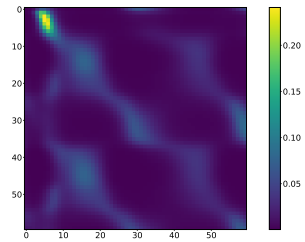
(a) $\sin(x)$



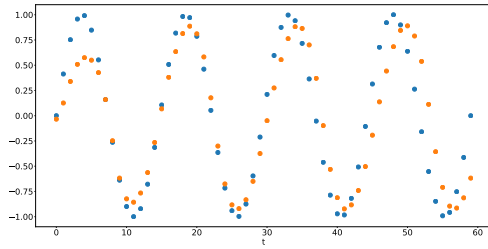
(b) $\sin(x)$



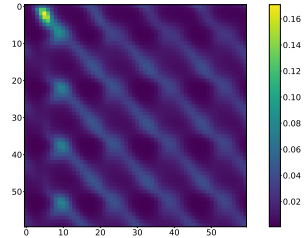
(c) $\sin(2x)$



(d) $\sin(2x)$



(e) $\sin(8x)$



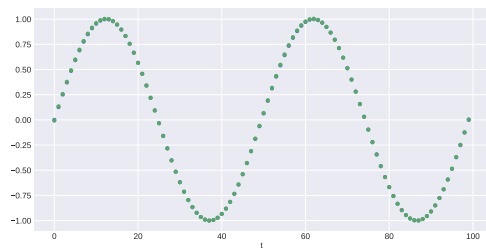
(f) $\sin(8x)$

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

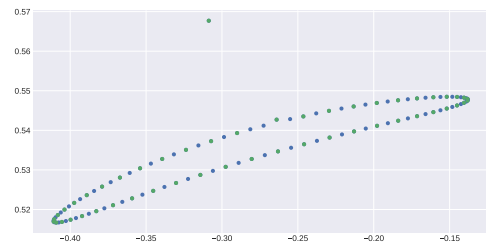
предположение о виде матрицы Attention подтверждается. На рис 2 показано как меняется вид матрицы Attention в зависимости от частоты синуса. Видно что количество диагоналей в матрице Attention соответствует частоте синусоидального сигнала.

5.2 Фазовое пространство, построенное при помощи LSTM

Как видно на рис. 3 в случае, когда размерность скрытого пространства 2, то получаем, что LSTM строит такое же пространство, как и Настя при помощи SVD разложения.



(a) $\sin(x)$



(b) $\sin(x)$

Рис. 3: Результаты для модели с рис. 1а без декодера