Neural SDE: phase trajectories of SDE in the action

A Preprint

Papay Ivan
papai.id@phystech.edu

Vladimirov Eduard
vladimirov.ea@phystech.edu

Strijov Vadim
strijov@phystech.edu

2024 год

Данная статья предлагает углубиться в математический аппарат, на котором строится модель Neural SDE. В ней будет рассмотрено, как вычисление фазовых траекторий СДУ обеспечивает качественный прогноз аномалий во временном ряду. Таким образом, это предоставит как возможность эффективнее бороться с шумами, так и, в частности, полезный инструмент для упреждения "чёрных лебедей"/антигауссиан. Подобного рода аномалии могли бы нарушить корректную работу Neural SDE в виду высокой корреляции элементов анализируемой выборки между собой.

Keywords SDE · Stratonovich integral · More

1 Введение

Сбор данных и подготовка их к последующей обработке являются одной из важнейших задач машинного обучения. К сожалению не всегда исследователь может гарантировать их целостность и корректность, ведь для тренировки модели чаще всего требуются выборки из тысяч, а то и десятков тысяч элементов - не удивительно, что в данных допускается наличие шума, влияющего на работу обученной модели. Эта задача остаётся актуальной и для временных рядов. Нужно, имея данные для начала временного ряда, проверить: возможно ли предсказать его на некотором горизонте так, что это предсказание будет точным, и не сломает ли это природу текущего временного ряда в вероятностном смысле?

В данной работе предполагается, что природа данных стохастическая. Иначе говоря, входная выборка представляет из себя данные о некоем дискретном случайном процессе. От дискретного процесса корректно будет перейти к непрерывному в силу того факта, что он порождается сигма-алгеброй из конечномерных распределений, которые реально апроксимировать с помощью данных, предоставленных для обучения модели.

Под данными имеются в виду значения n-мерного случайного процесса в фиксированном конечном множестве точек, которые необходимо будет сжать посредством CCA + CCM (Convergent-Cross Mapping). В таком случае в качестве варианта апроксимации полученного одномерного временного ряда предлагается взять метод приближения обыкновенными дифференциальными уравнениями, от которых затем уже перейти к стохастическим.

Сама идея использования обыкновенных дифференциальных уравнений("ОДУ") далеко не так нова[1]. Так, примерно с 2017-го года она была использована[2] для создания и теоретического обоснования корректности работы модели Neural ODE. Тем не менее, такой метод был всё ещё слаб в робастном смысле: был уязвим к состязательным атакам. Модель Neural SDE уже строилась на использовании стохастических дифференциальных уравнений("СДУ") и была в этом плане эффективнее своего предшественника.

Главной целью данного исследования является построение decision-rejection(принятие-отрицание) критерия корректности той или иной гипотезы о вероятностном распределении входных данных, как некоторого непрерывного случайного процесса. Для проверки фрагмента временного ряда на наличие аномалий достаточно применить этот критерий для проверки гипотезы о тождественности распределений его и всего остального ряда - разумно будет заключить, что в ряду происходят аномалии, если природа данных в стохастическом смысле резко поменялась.

Таким образом, полагая, что временной ряд порождается определенными конечномерными распределениями, мы сможем приблизить его с помощью стохастических дифференциальных уравнений. То же, разумеется, применимо и к анализируемому диапазону, который требуется проверить на наличие аномалий. Если фазовые траектории полученных дифференциальных уравнений различаются, то есть происходит резкое их возмущение, то очевидно, что в ряду произошла аномалия.

В прошлых работах, связанных с Neural SDE[2,3,4], СДУ использовались только для построения доверительных интервалов для элементов временного ряда. Этот подход в статье предлагается развить посредством использования фазовых траекторий полученных СДУ. Таким образом, можно будет проверять большие массивы данных на корреляцию между собой. В работе рассматривается задача - проверить, что два временных ряда обладают одинаковым вероятностным распределением. А именно проверяется соответствие видеоряда готовки еды и ряда данных, полученных с акселерометра, прикреплённого к его руке.

Ниже будет детально разобрано, как по СДУ, построенным по временным рядам, получить фазовые траектории путём зануления диффузии СДУ и его семплирования, как случайной величины. Тогда в качестве decision-rejection критерия как раз и будет использоваться сравнение полученных траекторий в плане их главных компонент.

2 Связанные работы

2.1 Neural ODE

Изначально Neural ODE был разработан как альтернатива методу остаточных нейронных сетей, состоящих из последовательности скрытых слоёв, значения на каждом из которых подчинялись следующей формуле:

$$h_{n+1} = h_n + f(h_n, w_n), \tag{1}$$

Где h_n - вход n-го слоя и f(h_n, w_n) - нелинейная функция, параметризованная по w_n - динамический параметр, задающийся непосредственно перед началом обучения модели

Было предложено[5] представление (1) в виде:

$$h_t = h_s + \int_s^t f(h_l, l; w) dl, \tag{2}$$

А вычисление такого дифференциального уравнения уже есть задача для Neural ODE. В данной работе в качестве числа слоев берется число элементов во временном ряду.

Algorithm 1 Neural ODE-solver

Require: динамические параметры w, начальное/конечное время t_0, t_1, конечное значение $z(t_1)$, градиент функции потерь в конечной точке $\frac{\delta L}{\delta z(t_1)}$

$s_0 = [z(t_1), \frac{\delta L}{\delta z(t_1)}, 0_{[w]}]$ ▷ Начальное состояние

$[z(t_0), \frac{\delta L}{\delta z(t_1)}, \frac{\delta L}{w}] = ODESolve(s_0, [f(z(t), t, w), -a(t)^T \frac{\delta f}{\delta z}, -a(t)^T \frac{\delta f}{\delta w}], t_1, t_0, w)$

$return \frac{\delta L}{\delta z(t_0)}, \frac{\delta L}{w}$ ▷ Возвращаем градиенты

Здесь ODESolve - это черный ящик, возвращающий нам решение ODE - в качестве него предлагается использовать метод Рунге-Кутта, а L - MSE

2.2 Neural SDE

Для учёта шума в наше дифференциальное уравнение следует добавить недетерменированную компоненту, диффузию. Получится следующее выражение, являющееся интегралом Стратоновича:

$$dX_t^w = h(t, X_t^w; w)dt + \sigma(X_t^w; w)dB_t, \tag{3}$$

Где $B_t = [B_t^1 ... B_t^n]$ - n-мерный Винеровский процесс, а $\sigma(X_t^w; w)$ - его матрица ковариаций в t-й момент времени

Таким образом выражение (1) для (n+1)-го слоя изменится

$$h_{n+1} = h_n + f(h_n, w_n) + \sigma(X_t^w; w)B_t, \tag{4}$$

Соответственно алгоритм остаётся тем же, что и для ODE с поправкой на вычисление матрицы ковариации путём семплирования входной выборки.

3 Постановка задачи

Давайте подытожим: в общем виде для решения поставленной задачи требуется осуществить следующие шаги - применить метод Neural ODE к временному ряду, учесть гауссовский шум и тем самым свести задачу к модели Neural SDE, вычислить фазовые траектории для временных рядов, предварительно свернув два многомерных временных ряда(из видео и из акселерометра) к минимально возможному размеру и, наконец, сравнить полученные фазовые траектории временных рядов по поведению. Далее предлагается вкратце их разобрать

Первый и второй временной ряд соответственно:

$$\{X_{t_i}\}_{i=1}^n, \{Y_{t_i}\}_{i=1}^n \tag{5}$$

, где t_i - временные метки, общие для обоих временных рядов. Предполагается, что траектории обоих рядов подчиняются некоторым двум СДУ(см. 2.2)

Решив СДУ с помощью SDE-solver-а мы получим точные значения $f(h_n, w_n) = C_n - const$(см. (4)) для $\forall n$. Тогда в той же формуле (4) приняв во внимание факт того, что $EB_t = 0$, занулим коэффицент диффузии и получим следующее выражение:

$$h_{n+1} = h_n + C_n, \tag{6}$$

Это оценки приращений элементов временного ряда. По ним строится матрица Ганкеля, описывающая фазовые траектории процесса, размера n/2 на n/2 вида:

C$_1 C_2 \cdots C_{n/2}$
$C_2 C_3 \cdots C_{n/2+1}$

$\vdots \vdots \ddots \vdots$
$C_{n/2} C_{n/2+1} \cdots C_n$

Матрицы от первого и второго процессов обозначим B и C

Наконец, данные матрицы проверяются с помощью метода Сугихары на элемент корреляции.

$$\rho = \frac{1}{N}\sum_{i=1}^{n} corr(B_{\sigma_i}, C_{\omega_i}) \tag{7}$$

Где corr - это коэффицент корреляции; считающийся по формуле Кэндалла или Спирмена, B_j - j-й столбец матрицы B; $\sigma_j, \omega_j \sim U[0, n]$ ограниченные на дискретном носителе; N - число итераций для подсчета корреляции, подбирается заранее.

Таким образом получим критерий для проверки гипотезы об однородности временных рядов. Если $\rho \notin [-0.95, 0.95]$ гипотеза отвергается.

Список литературы

[1] “Neural Ordinary Differential Equations Ricky” T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud

[2] “Neural SDE: Stabilizing Neural ODE Networks with Stochastic Noise” Xuanqing Liu, Tesi Xiao, Si Si, Qin Cao, Sanjiv Kumar, Cho-Jui Hsieh

[3] “Riemannian Neural SDE: Learning Stochastic Representations on Manifolds” Sung Woo Park , Hyomin Kim , Kyungjae Lee , Junseok Kwon

[4] "Riemannian Diffusion Models" Chin-Wei Huang, Milad Aghajohari, Avishek Joey Bose, Prakash Panangaden, Aaron Courville

[5] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In Advances in Neural Information Processing Systems, pages 6572–6583, 2018.