Вероятностное расширение рекуррентной нейронной сети

Маргасов Арсений Олегович

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: margasovarsenij@gmail.com

Научный руководитель — *Ульянов Владимир Васильевич*

В работе рассматривается идея применения аппарата дифференциальных уравнений в рамках глубокого обучения (нейронные обыкновенные дифференциальные уравнения или НОДУ) [1], а также предложена новая нейросетевая архитектура.

Предложенная модель p-ODE-RNN построена на вероятностном расширении рекуррентной сети с предобработкой скрытого состояния нейронными дифференциальными уравнениями (ODE-RNN [2]). Сравнение нейросетевых архитектур и эксперименты по оценке их обобщающей способности проводились на двух наборах данных: помесячная цена закрытия индекса Московской биржи и среднемесячная температура воздуха в городе Москве. Полученные результаты показывают, что p-ODE-RNN позволяет получать сопоставимое с ODE-RNN качество, затрачивая при этом значительно меньшее время на обучение. Также было показано, что можно рассматривать p-ODE-RNN как случайный процесс, что позволяет взглянуть на эту модель не только с точки зрения глубокого обучения, но и с точки зрения теории случайных процессов.

Определение 1. *Уравнение вида*

$$\frac{dh(t)}{dt} = f(h(t), t, \theta), \quad t \in [t_0, t_1]\,. \tag{1}$$

называется нейронным обыкновенным дифференциальным уравнением, в котором h —скрытое состояние сети, f — нейронная сеть, задающая динамику развития скрытого слоя, θ — её параметры, а t определяет нумерацию скрытых слоёв.

p-ODE-RNN. Рассмотрим новую архитектуру p-ODE-RNN, для которой скрытое состояние (слой сети) h_i является случайной величиной следующего вида:

RNNCell($\cdot$) — блок стандартной рекуррентной нейронной сети, ODESolve($\cdot$) – численный метод решения НОДУ. Параметр p был

$\mathbb{P}(h_i = x)$	p	$1 - p$
x	$\text{RNNCell}(\text{ODESolve}((t_{i-1}, t_i), h_{i-1}, \theta, f), x_i)$	$\text{RNNCell}(h_{i-1}, x_i)$

Таблица 1: Случайная величина h_i.

выбран равным $1/t$. В качестве меры ошибки моделей был использован коэффициент детерминации или $R^2 = 1 - SS_{res}/SS_{tot}$, где y – вектор реальных значений, $\hat{y}$ – вектор, предсказанных значений, $\bar{y} = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$, $SS_{tot} = \sum\limits_{i=1}^{n}(y_i - \bar{y})^2$, $SS_{res} = \sum\limits_{i=1}^{n}(y_i - \hat{y_i})^2$.

модель	ODE-RNN	p-ODE-RNN
R^2	0.89	0.83
время	1.86 сек.	0.20 сек.

Таблица 2: Итоговое значение R^2 и среднее время, затраченное на обучение на 1 эпоху (цены закрытия индекса Московской биржи).

модель	ODE-RNN	p-ODE-RNN
R^2	0.90	0.87
время	2.64 сек.	0.28 сек.

Таблица 3: Итоговое значение R^2 и среднее время, затраченное на обучение на 1 эпоху (среднемесячная температура в С в Москве).

Литература

1. Chen Ricky T. Q., Rubanova Yulia, Bettencourt Jesse, Duvenaud David. Neural Ordinary Differential Equations // Advances in Neural Information Processing Systems 31 (NIPS 2018), Montreal, Canada, 2018.
2. Rubanova Yulia, Chen Ricky T. Q., Duvenaud David. Latent ODEs for Irregularly-Sampled Time Series // Advances in Neural Information Processing Systems 32 (NIPS 2019), Vancouver, Canada, 2019.