Прогностические мультимодели разномасштабных временных рядов интернета вещей

Нейчев Радослав Георгиевич

МФТИ, ФУПМ
Кафедра «Интеллектуальные системы»
Научный руководитель д.ф.-м.н Стрижов В. В.

29 июня 2016

Цели работы

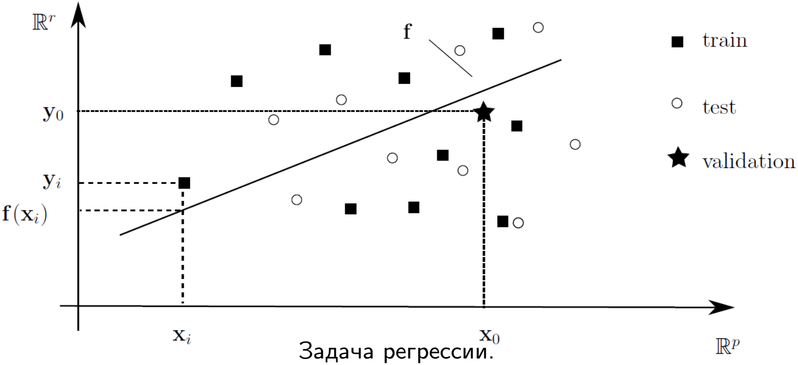
Решается проблема прогнозирования большого числа разнородных мультимасштабных временных рядов.

Проблемы

- Временные ряды, генерируемые устройствами, представлены с различной частотой дискретизации.
- Ввиду разнородности данных, генерируемых интернетом вещей, их описание с помощью единой модели затруднительно.

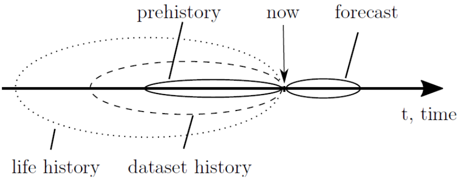
Метод

- Для построения авторегрессионной матрицы временные ряды оптимально ресемплируются. Отношение частот приводится к рациональным числам.
- Для повышения качества прогноза используется смесь моделей (экспертов).

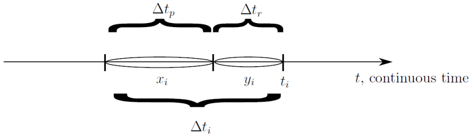


Задача регрессии.

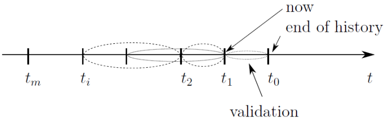
Необходимо найти отображение $\mathbf{f} : X \to Y$, доставляющее наименьшую ошибку.



Представление временного ряда.



Порождение пары объект-ответ.



Генерация объектов.

Литература

- Decision forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. *A. Criminisi, J. Shotton, E. Konucoglu* / Microsoft Research technical report TR-2011-114.
- Short-Term Load Forecasting Using Random Forests. *Gregor Dudec* / Intelligent systems, 2014.
- Kernel-based mixture of experts models for linear regression. *J. Santarcangelo, X. Zhang* / IEEE 2015.
- Stresstest procedure for feature selection algorithms. *A.M. Katrutsa, V.V. Strijov* / Chemometrics and Intelligent Laboratory Systems 2015.

Постановка задачи прогнозирования

Для рядов $\mathbf{s}_1, \ldots, \mathbf{s}_l$ строится матрица объект-признак:

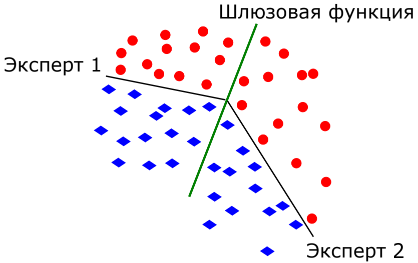
$$\mathbf{X}^* = \left[\begin{array}{c|ccc} \hat{\mathbf{y}}' & \mathbf{x}_0' & \mathbf{x}_0'' & \ldots \\ \hline \mathbf{y}_1' & \mathbf{x}_1' & \mathbf{x}_1'' & \ldots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_m' & \mathbf{x}_m' & \mathbf{x}_m'' & \ldots \end{array}\right] = \left[\begin{array}{c|c} \underset{1\times r}{\hat{\mathbf{y}}} & \underset{1\times n}{\mathbf{x}_0} \\ \hline \underset{m\times r}{\mathbf{Y}} & \underset{m\times n}{\mathbf{X}} \end{array}\right].$$

Для получения прогноза $\hat{\mathbf{y}} = \hat{\mathbf{f}}(\mathbf{x}_0)$ требуется найти отображение $\hat{\mathbf{f}} : \mathbf{X} \to \mathbf{Y}$, такое что

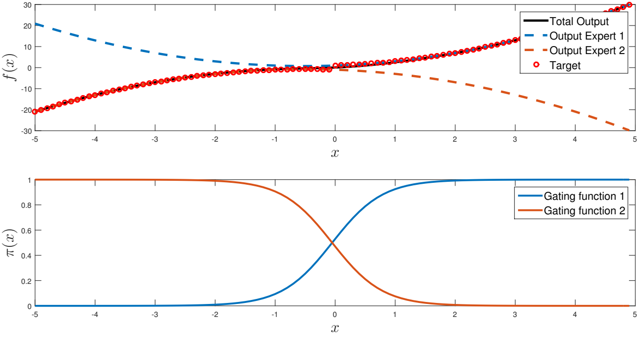
$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} S(\mathbf{f}, \mathbf{X}, \mathbf{Y}),$$

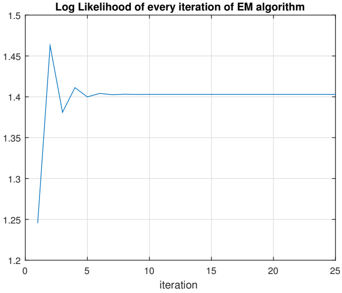
где $S(\mathbf{f}, \mathbf{X}, \mathbf{Y})$ — фиксированная функция ошибки.
В качестве $\hat{\mathbf{f}}$ выступает смесь моделей (экспертов).

Шлюзовая функция (англ. gating function) — отображение $\pi(\mathbf{x}, \mathbf{v}) : X \to [0, 1]$, определяющее вероятность того, что соотвествующая ей модель $\mathbf{f}$ описывает объект $\mathbf{x} \in X$, где X — пространство объектов.



Если в смеси моделей заменить веса моделей на шлюзовые функции $\pi_k(\mathbf{x}, \mathbf{v})$, получится **смесь экспертов**.





Смесь двух экспертов (полиномов второй степени) на синтетических данных.

Смесь экспертов

Пусть есть K линейных моделей $\mathbf{f}_1, \dots, \mathbf{f}_K$, пару $(\mathbf{x}, \mathbf{y})$ k-ая модель описывает с вероятностью $p(k|\mathbf{x}, \mathbf{w})$. С учетом всех моделей

$$p(y|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^{K} p(y, k|\mathbf{x}, \boldsymbol{\theta}) = \sum_{k=1}^{K} p(k|\mathbf{x}, \boldsymbol{\theta}) p(y|k, \mathbf{x}, \boldsymbol{\theta}) =$$

$$= \sum_{k=1}^{K} \frac{\exp(\mathbf{v}_k^\mathsf{T}\mathbf{x})}{\sum_{k'=1}^{K} \exp(\mathbf{v}_{k'}^\mathsf{T}\mathbf{x})} \exp\left((-\frac{1}{2\beta_k}(\mathbf{y} - \mathbf{w}_k^\mathsf{T} X)^\top (\mathbf{y} - \mathbf{w}_k^\mathsf{T} X) \right).$$

За $\boldsymbol{\theta}$ обозначен вектор гиперпараметров

$$\boldsymbol{\theta} = [\mathbf{w}_1, \dots, \mathbf{w}_k, \boldsymbol{\pi}(\mathbf{x}, \mathbf{v}), \beta]^\mathsf{T},$$

где $\mathbf{B} = \beta \mathbf{I}_m$ — ковариационная матрица $\mathbf{y}$.

Чтобы найти оценку максимального правдоподобия $\hat{\boldsymbol{\theta}}$

$$\hat{\boldsymbol{\theta}} = \mathrm{argmax}_{\boldsymbol{\theta}} \ln p(\mathbf{y}|\boldsymbol{\theta}),$$

введем скрытые индикаторные переменные
$Z = [\mathbf{z}_1, \dots, \mathbf{z}_m]\,, \quad z_{ik} \in \{0, 1\}$, такие что

$$z_{ik} = 1 \Leftrightarrow y_i \sim \mathcal{N}(\mathbf{w}_k^\mathsf{T}\mathbf{x}_i, \beta).$$

В таком случае логарифм правдоподобия имеет вид

$$p(\mathbf{y}|X, Z, \boldsymbol{\theta}) = \sum_{i=1}^{m}\sum_{k=1}^{K} z_{ik}\left(\ln \pi_k(\mathbf{x}_i, \mathbf{v}_k) + \ln \mathcal{N}(y_i|\mathbf{w}_k^\mathsf{T}\mathbf{x}_i, \beta)\right).$$

Перейдем к матожиданию по Z и применим **EM-алгоритм** для максимизации $\mathsf{E}_Z[p(\mathbf{y}, Z|X, \boldsymbol{\theta})]$.

EM-алгоритм для смеси экспертов

Е-шаг: По правилу Байеса

$$\gamma_{ik}^{r+1} = \mathsf{E}(z_{ik}) = p(k|\mathbf{x}_i, \boldsymbol{\theta}^r) = \frac{\pi_k(\mathbf{x}_i, \mathbf{v}_k)\mathcal{N}(\mathbf{y}_i|\mathbf{x}_i^\mathsf{T}\mathbf{w}_k^r, \beta^r)}{\sum_{k'=1}^K \pi_{k'}(\mathbf{x}_i, \mathbf{v}_{k'})\mathcal{N}(\mathbf{y}_i|\mathbf{x}_i^\mathsf{T}\mathbf{w}_{k'}^r, \beta^r)}.$$

М-шаг: Пересчитывается оценка параметров

$$\mathbf{v}_k^{r+1} = \mathrm{argmax}_{\mathbf{v}} \sum_{i=1}^m \gamma_{ik}^{r+1} \ln \pi_k(\mathbf{x}_i, \mathbf{v}),$$

$$\mathbf{w}_k^{r+1} = \mathrm{argmax}_{\mathbf{w}_k} \left[-\sum_{i=1}^m \gamma_{ik}^{r+1} \left(y_i - \mathbf{w}_k^\mathsf{T}\mathbf{x}_i\right)^2 \right],$$

$$\beta_k^{r+1} = \mathrm{argmax}_{\beta} \left[n \ln \beta - \frac{1}{\beta} \left(y_i - \mathbf{x}_i^\mathsf{T}\mathbf{w}_k^{r+1}\right)^2 \right].$$

В случае смеси моделей вместо выражения для $\mathbf{v}_k^{r+1}$ используется $\pi_k^{r+1} = \frac{1}{n}\sum_{i=1}^m \gamma_{ik}^{r+1}$

Вычислительный эксперимент

Цель эксперимента

Показать, что разбиение обучающей выборки и применение к подвыборкам простых моделей позволяет получать адекватные результаты.

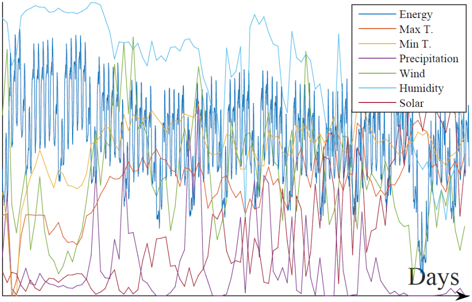
Сравнивались прогнозы, полученные с помощью следующих моделей:

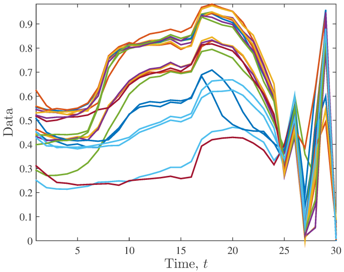
- MLR — Векторная линейная регрессия с l-2 регуляризацией.
- Нейронные сети.
- SVR — Support Vector Regression machine.
- Random Forest — простой вариант смеси моделей.

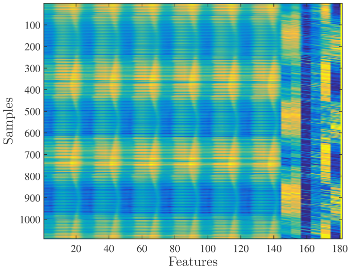
Описание данных

Данные по потреблению электроэнергии в Польше за период 1999-2004, включающие в себя:

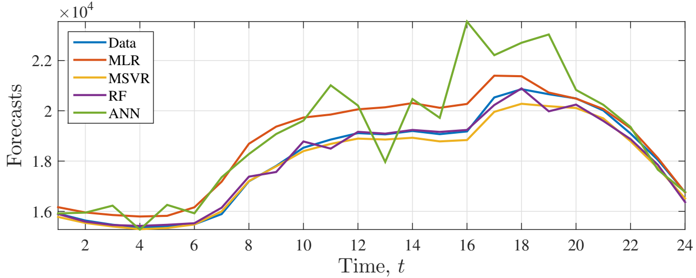
- Почасовые значения потребления электроэнергии (52512 наблюдений)
- Шесть временных рядов, фиксирующих природные показатели: максимальная и минимальная температуры, влажность, количество осадков, сила ветра, облачность (2188 наблюдений).

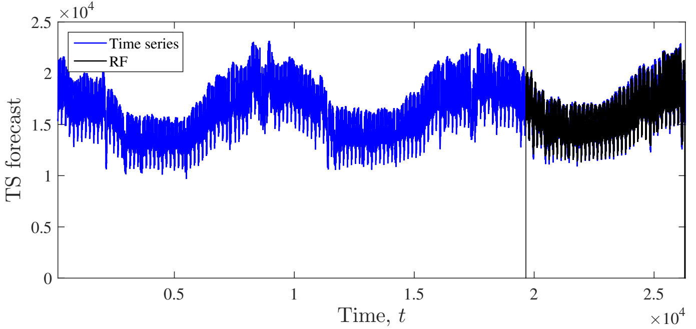






Модель	Train	Test
Baseline	0.213	0.211
MLR	0.120	0.126
MSVR	0.076	0.082
Random Forest	0.069	0.084
Neural network	0.176	0.184


$\times 10^4$
2.2
2
1.8
1.6
Forecasts
Data
MLR
MSVR
RF
ANN
2
4
6
8
10
12
14
16
18
20
22
24
Time, t


$\times 10^4$
2.5
2
1.5
1
0.5
0
TS forecast
Time series
RF
0.5
1
1.5
2
2.5
Time, t
$\times 10^4$

Основные результаты данной работы

- Предложен комплексный подход к прогнозированию временных рядов, позволяющий работать с разномасштабными данными.
- Теоретически обоснована применимость мультимоделей к задаче прогнозирования и произведена практическая проверка данного подхода в рамках вычислительного эксперимента.

Список публикаций:

1. Выбор оптимального набора признаков из мультикоррелирующего множества в задаче прогнозирования. *Нейчев Р.Г., Катруца А.М., Стрижов В.В.* / Заводская лаборатория. №3 2016. Том 2.
2. Отбор мультикоррелирующих признаков в задаче векторной авторегресии. *Ахтямов П.И., Нейчев Р.Г., Стрижов В.В* / готовится к подаче.
3. Feature generation for multiscale time series forecasting. *Мотренко А.П., Нейчев Р.Г., Исаченко Р.В., Попова М.С., Стрижов В.В.* / подана на ICDM 2016.