

Metalearning методы для выбора лучших конфигов во временных рядах

Для подбора лучшей модели и лучших гиперпараметров для нее на каком-то конкретном наборе данных существует 3 популярных метода: полный перебор, случайный (или константный) выбор и ансамбль. Первый метод очень хорош с точки зрения качества, однако для перебора всех моделей и всех гиперпараметров к ним нужно много времени и ресурсов. Для второго метода не нужно много времени, однако качество оставляет желать лучшего. Существует метод, который будет описан ниже, работающий в онлайне быстро и очень близко к наилучшему возможному результату.

1 Metalearning методы

Имея множество различных рядов можно подобрать (перебором) для них лучшие модели и лучшие гиперпараметры. После обучить модель (классификатор для случая подбора модели и регрессор или классификатор для поиска гиперпараметров), входом которой является какое-то признаковое описание ряда, а таргетом номер модели или же значение какого-либо гиперпараметра. Обучаемся на фичах всего ряда => нужно их придумать. Они должны отражать все параметры от временной структуры данных (Примеры: количество точек смен тренда, "мощность" тренда и сезонности, коэффициент в преобразовании Бокса-Кокса, горизонт прогноза и так далее) (Во 2 статье есть очень много различных признаков). Для нового ряда мы вычисляем все признаки, прогоняем их через обученный классификатор и получаем прогнозируемо-наилучшую модель для предсказания ряда.

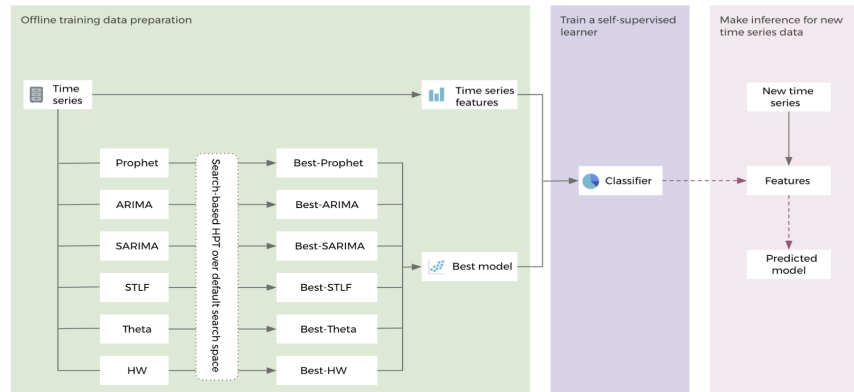


Figure 1: Workflow of Self-Supervised Learning for Model Selection (SSL-MS)

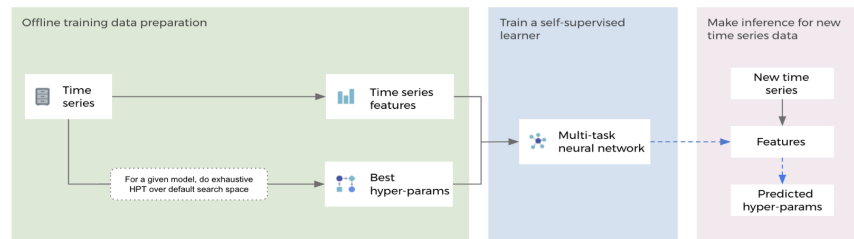


Figure 2: Workflow of Self-Supervised Learning for Hyper-Parameter Tuning (SSL-HPT)

Для наилучшего эффекта задачи поиска лучшей модели и гиперпараметров для данного ряда могут решаться вместе:

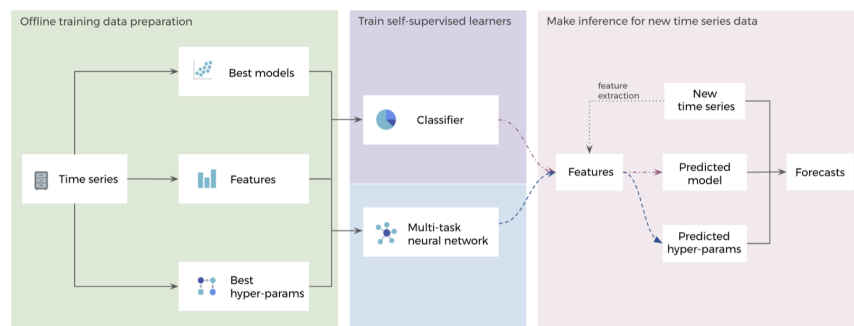


Figure 4: Workflow of SSL-MS with SSL-HPT

Очевидно, для обучения метамоделей нужно очень много рядов (при чем желательно с различными длинами, трендами, шумами и сезонностями). Тут может помочь аугментация (добавление/вычитание трендов/сезонностей, добавление шума, изменение горизонта и так далее).

Во всех статьях на эту тему все делают ровно так, но есть 3 вопроса, от которых качество работы и зависит.

1.1 Какие признаки с рядов собирать?

У разных разработчиков они разные, очень много вариаций, нужно смотреть/думать, что работает лучше для конкретных нужд и задач (взять много признаков не вариант: классификатор не сможет обучиться). Предлагают также снижать размерность за счет отбора признаков или PCA.

1.2 Сколько и какие модели выбрать?

Нигде сложных моделей не встречал в статьях. Самое сложное из всех моделей это prophet. По количеству тоже не сильно много, 5-10 моделей.

1.3 Что с метамоделями?

Для метамоделей, выбирающей прогнозирующую модель, можно выбирать любой классификатор (нейронки, логистическая регрессия, случайный лес, дерево, бустинги, SVM, kNN) (дерево хорошо интерпретируемо; легко посмотреть, в каких ситуациях какая модель лучше работает). После экспериментов из одной статьи лучшими (почти) всегда оказались случайный

лес или бустинг (XGBoost), в других статьях тоже отдавали предпочтение деревянным моделям.

Для гиперпараметров: очевидно, что для разных параметров нужны разные модели (если признак категориальный, то классификатор, если вещественный, то регрессор). Чтобы не плодить кучу моделей для каждого параметра в первой статье предложен следующий способ: обучается нейронка с общими параметрами. Для каждого конкретного параметра на выходы с общих слоев навешивают несколько слоев. По задумке общие слои сохраняют всю нужную информацию о поведении ряда, что помогает предсказать все лучшие параметры.

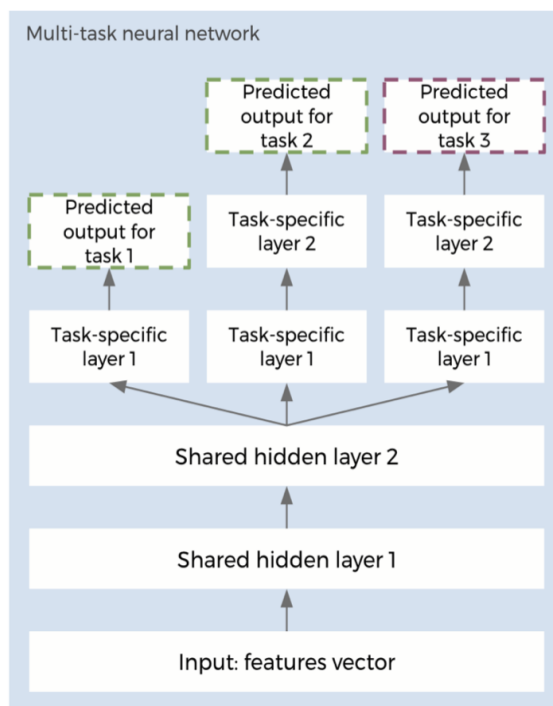


Figure 3: Illustration of a sample multi-task neural network

Можно брать не один лучший конфиг с лучшими гиперпараметрами, а например топ15. Во первых скорее всего между этими конфигами результат будет почти одинаковый, во вторых сразу увеличивается обучающих набор данных.

Список литературы

- [1] Self-supervised learning for fast and scalable time series hyper-parameter tuning, <https://arxiv.org/pdf/2102.05740.pdf>
- [2] Meta-learning how to forecast time series, <https://www.monash.edu/business/ebs/our-research/publications/ebs/wp06-2018.pdf>
- [3] Meta-learning for time series forecasting and forecast combination, http://eprints.bournemouth.ac.uk/11186/1/Neurocomputing_Lemke_Gabrys_2009_accepted.pdf
- [4] Time series model selection with a meta-learning approach; evidence from a pool of forecasting algorithms, <https://arxiv.org/ftp/arxiv/papers/1908/1908.08489.pdf>