

Предсказание температуры во времени и пространстве

Дробин М.Е. (МФТИ ГУ)

<2019-04-08>

Contents

1 Введение	1
2 Обзор литературы	2
3 Ссылки	3

Abstract

В этой работе исследуется качество прогнозирования температуры для одной из метеорологической станций, обучая нейросетевые алгоритмы(RBF, MLP) и классические(KNN, spatial averaging, inverse distance methods) по данных от других станций в пределах Англии.

1 Введение

Температура воздуха - временные ряды с высоким временным разрешением - измеряют только в немногочисленных, далеко разнесенных друг от друга метеорологических станциях. Поэтому появляется необходимость в пространственной интерполяции этих значений на области, где температура. Кроме пространственной интерполяции появляется необходимость в прогнозировании будущей температуры воздуха, т.е. временной интерполяции. Такие задачи появляются в сельском хозяйстве, прогнозировании погоды в городских условиях и т.д.

В качестве данных используется почасовая информация(влажность, ...) с 30 метеорологических станций Англии(каждая в отдельном городе) за 2 года. И для некоторых из них предсказывается температура. Данные взяты из сервиса darksky.

2 Обзор литературы

В статье Spatial Interpolation of Surface Air Temperatures... авторы сравнили качество MLP и spatial average, knn, inverse distance methods для задачи интерполяции температуры на 11 NOAA станций, на которых измеряют температуру. Эти станции расположены примерно в одном штате. Температура с этих станций - это множество ответов алгоритма, а данные с GCM(general circulation model) - численное модели - это мн-возможным объектов. Такая задача называется down-scaling GCM. Трудность заключается в том, что пространственное разрешение выходных данных GCM - от $2.5^\circ \times 2.5^\circ$ до $8^\circ \times 10^\circ$ в долготу и ширину - это слишком грубо для предсказания - поэтому обучают по такой сетке нейросеть и предсказывают температуру для точек между сетки - выхода GCM. Авторы предсказывали максимальную температуру за день T_{\max} и посчитали качество нейросети(архите- ктура нейросети: 4-30-11 и 16-54-11 с сигмоидной функцией активацией). Для сетки с 4мя входами R^2 и rmse были в среднем по 11 станциям 5.69 и 0.93; для сетки с 16ю входами - 5.12 и 0.94.

В статье Application of neural networks for the prediction of hourly mean surface temperatures in Saudi Arabia авторы использовали почасовые данные с шести метеорологических станций за 2 года с равнины, расположенной в Китае и ограниченной горами. 5 из них входят в мн-во объектов алгоритмов, для шестой предсказывается температу. 60% данных использовалось для обучения, остальное для теста. Сравнили MLP 5-17-1 с tanh активацией и RBF с скрытым слоем из 7450 нейронов и гауссовой функцией активацией. Качество на тестовой выборке: R^2 и rmse($^\circ\text{C}$) для MLP равны 0.96 и 1.067, а для RBF - 0.95 и 1.12 соответственно.

Данные по нескольким городам - погодные данные, взятые из измерителей в аэропортах этих городов. Задача - предсказать температуру(или прочие погодные показатели, напр., влажность..) для города, не использованного в обучении, используя нейросетевые алгоритмы.

3 Ссылки

Основные понятия и обозначения в машинном обучении. Воронцов К.В.