Analysis of the Spatial Covariance Shift in Weather Inference

Simone PAPICCHIO, Federico TIBLIAS, Massimo PRONESTI, Daniele FALCETTA
Department of Data Science
Eurecom
Sophia Antipolis, France
{simone.papicchio, federico.tiblias, massimiliano.pronesti, daniele.falcetta}@eurecom.fr

I. INTRODUCTION

When the distribution of the train and test data differs, this is known as dataset shifting. This may produce several problems because the model is trained on one distribution but it is used to predict different data distributions, resulting in poorer results. There are different types of data shifting such as Covariance Shift[1], Probability Shift[2] and Concept Shift.[3] In this work we analyze the presence of the Spatial Covariance Shift and try to address its main challenges.

The data consists of pairs of meteorological features and target values at a particular latitude/longitude and time. The Regression task consist in predict the target value which is an air temperature measurements at 2 metres above the ground. The features are:

- *weather-related features:* sun evaluation at the current location, climate values of temperature, pressure and topography, and meteorological parameters on different pressure
- *surface levels* from weather forecast model predictions.

Weather forecast model predictions are values produced by the following weather forecast models:

- Global Forecast System *(GFS)*
- Global Deterministic Forecast System from the Canadian Meteorological Center *(CMC)*
- Weather Research and Forecasting *(WRF)*

Each model returns the following predicted values: wind, humidity, pressure, clouds, precipitation, dew point, snow depth, air and soil temperature characteristics. Where applicable, the predictions are given at different isobaric levels from 50 hPa (20 km above ground) to the ground level.

Altogether, there are 111 features in total. It is important to note that the features are highly heterogeneous, i.e., they are of different types and scales.

The main challenge of this dataset is the Spatial Covariance Shift from train to test as is visible in Fig. 1

[1]Changes in the independent variables or features of the dataset
[2]Changes in the target variable or the dependent variable in the dataset
[3]Change in the connection between the independent and the target variable across datasets

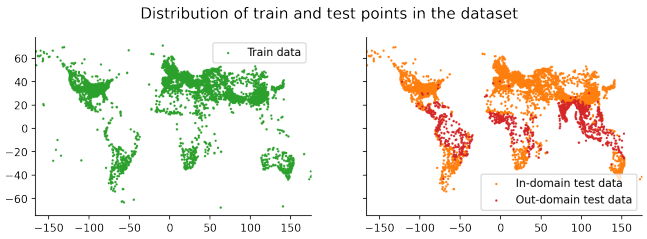

Fig. 1. The figure clearly shows that the regions in the range of latitude from -18 to 10 are not present in the train dataset. This phenomenon is called *Spatial Covariance Shift*. [best visualization in colors]

II. DATA ANALYSES

A. Outliers

B. Correlations

C. Feature Shift

III. PRE-PROCESSING

IV. MODELS

V. HYPER-PARAMETER TUNING

VI. RESULT

VII. CONCLUSION