Final Project Proposal

Cory Schillaci and Ana Rufino Ferreira

April 9, 2014

The goal of our project is to use machine learning techniques to provide computationally inexpensive near-real-time updates to improve the wind forecasts created by Mike Dvorak of Sail Tactics, LLC. These forecasts are used by recreational sailors to provide tactical advantages during races. Our initial goal is to provide these enhanced forecasts for wind flow in the Gulf of the Farallones.

We will be submitting this project for both CS289A and Info 290T.

1 Motivation

Sailboat speeds depend strongly on the local wind strength and relative direction, as can be seen in Figure 1. In sailboat racing, leveraging differences in wind direction or wind speed over the race course is crucial to sailing the shortest possible course at the greatest possible speed. Accurate wind-forecasts give sailors clues to what the wind is doing far away on the course and how it might change in the future, building this information into their strategy.

On multi-day offshore races, use of weather forecasts and routing software has been commonplace for some time. Until recently, however, the best forecasts available had insufficient resolution and accuracy to be helpful for short day races. Sail Tactics has just introduced a wind forecast for the San Francisco Bay intended specifically for sailors which provides sufficient detail and reliability to be used in racing situations. In the near future, they also hop to begin providing forecasts in the Gulf of the Farallones (roughly the area between the Golden Gate Bridge and the Farallon Islands), where many races also take place.

Previous studies have evaluated whether machine learning techniques could further improve the forecast. A linear support vector regression model was trained on the forecast and on publicly available weather observations made before the forecast day. With this approach, the researchers were able to reduce the RMS error of wind readings taken at the Berkeley Yacht Club by as much as 20%[1].

2 The Sail Tactics Model

The Sail Tactics forecast provides the best over-the-water wind forecast for San Francisco Bay. This is achieved by running the forecasts the morning of the forecast day, and by running the

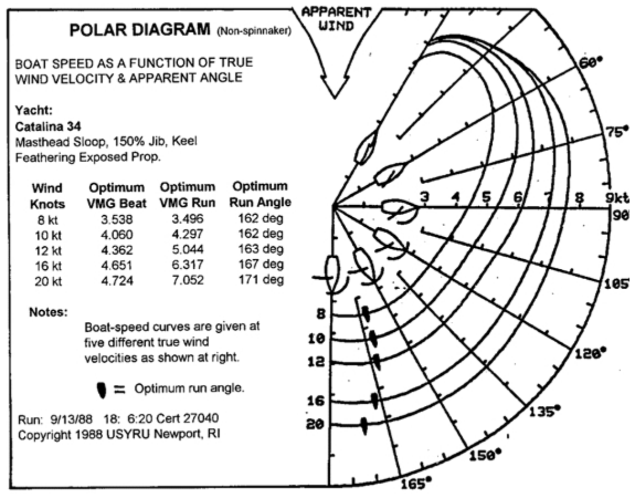

Figure 1: An example polar digram showing the dependence of a sailboat's speed through the water on wind speed and direction.

forecast on a fine-meshed grid. Information intended for a non-technical audience is available at http://www.sailtactics.com/forecast-accuracy.

The Weather Research and Forecasting (WRF) model is used to generate the forecast. WRF combines solutions of physics-based nonlinear differential equation models with data assimilation. Essentially, this process assumes that both forecasts and observations are uncertain and tries to produce a better forecast by combining them.

3 Training Data

We will have access to the following data sets for our training data:

- Publicly available weather buoy data
- Public data from Oakland Airport, SFO and other land based observation stations
- Two weeks of "one day ahead" WRF forecast data from Sail Tactics, LLC
- Two weeks of "morning of" WRF forecast data from Sail Tactics, LLC

The primary challenge of working with this data is the limited availability of WRF training data. Each day of forecast data includes forecasts from 8:00 until 19:30 local time at half hour intervals, for a total 252 forecasts. However, if our proof-of-concept work is promising, more data can be obtained for training once the WRF model is being run daily.

Observational data presents the opposite problem, where a variety of features could be extracted from the multitude of weather stations and buoys. We expect real time wind observations, the temperature profile and some barometric pressure information to be crucial, but experimentation and dimensionality reduction will likely be necessary here.

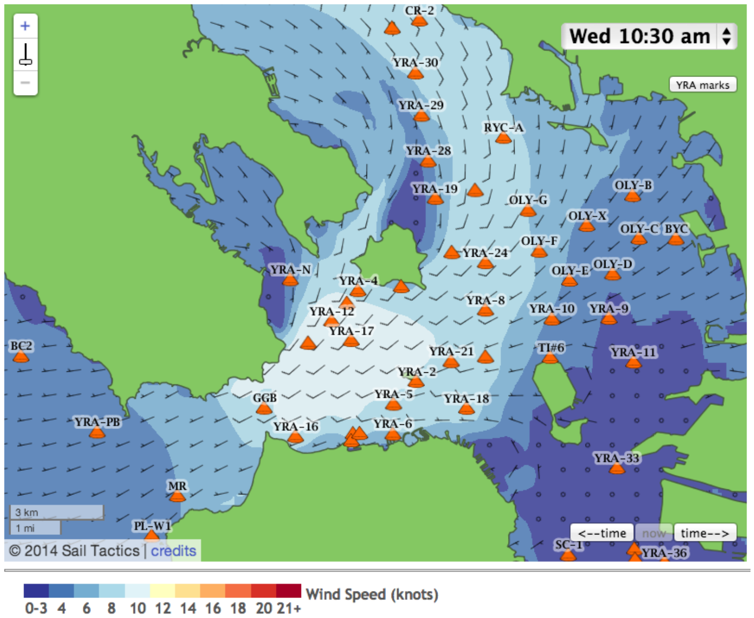

Figure 2: A sample forecast of the SF Bay winds for 10:30AM on April 9, 2014

Finally, the observational data which will be used as training labels is sparse compared to the WRF grid. Observations are not be aligned with the WRF grid. We may also get access to wind speed and direction data recorded by the instruments of boats racing in this area for a few days, which could help in making an informed interpolation.

4 Proposed Work

To reiterate, the goal is to create an enhanced forecast for the Gulf of the Farallones. Our first goal is to provide an improved hybrid forecast combining the WRF and machine learning based on data available at the time the WRF computation completes (approximately 8am each morning). An additional objective would be to produce an updated forecast every thirty minutes (or some other reasonable time interval) based on recent observational data.

This is a regression problem. We need to forecast the wind speed and direction. An initial goal is to do so at the weather buoys, where training labels are available. Beyond this, an interpolation technique to extend the corrections over the entire forecast grid would be a major accomplishment but is likely beyond the scope of this project.

A brief glance at the literature indicates that support vector regression (SVR) and neural networks are popular choices for weather forecasting. The underlying physics is known to be strongly nonlinear. Neural networks are good at modeling nonlinearity but would also be prone to overfitting with such a small data set, although ensemble methods using shallow networks might be applicable. We will certainly experiment with various kernels and SVR.

5 Timeline

This project must be completed in five weeks. Here is an estimated timeline

Week 1: Data Exploration and Preprocessing
Make sure we understand how to access data and what it contains. Identify any gaping holes in the feature set. Associate training labels with features.

Week 2: SVR training for Daily Enhanced Buoy Forecasts
Experimenting with SVR for enhancing daily forecasts with morning of observations.

Week 3: SVR training for Real Time Enhanced Buoy Forecasts
Experiment with SVR for continually updated enhanced forecasts with real-time observations.

Week 4: Experiments with Other Statistical Learning Models
Experiment with Neural Networks, Decision Tree Regression, etc.

Week 5: Write Report
Finalize results and write a report.

References

[1] Daniel Bejarano and Alejandro Quiroga. Wind prediction: Physical model improvement through support vector regression, December 2013.