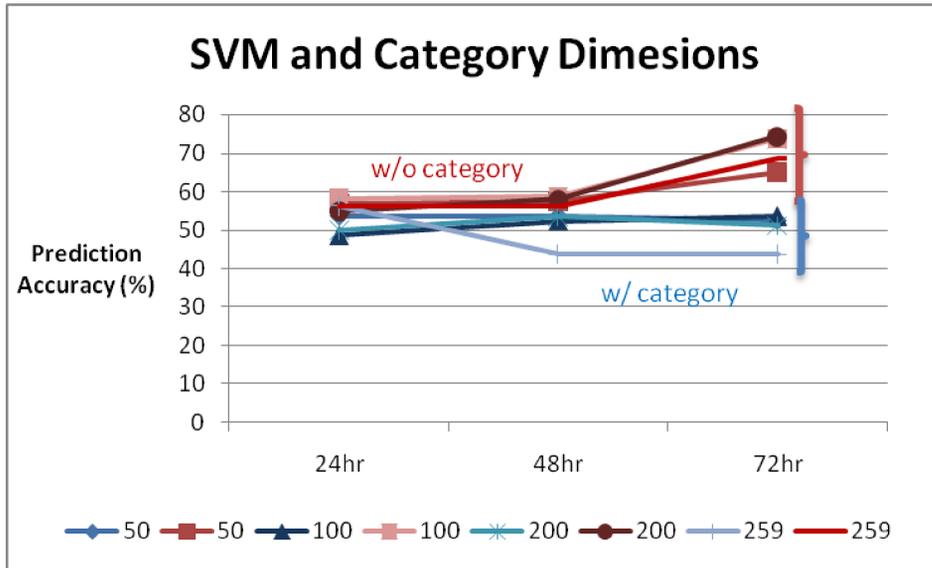


Predicting Atlantic basin tropical cyclone landfalls

Shaddi Hasan.

Support Vector Machine

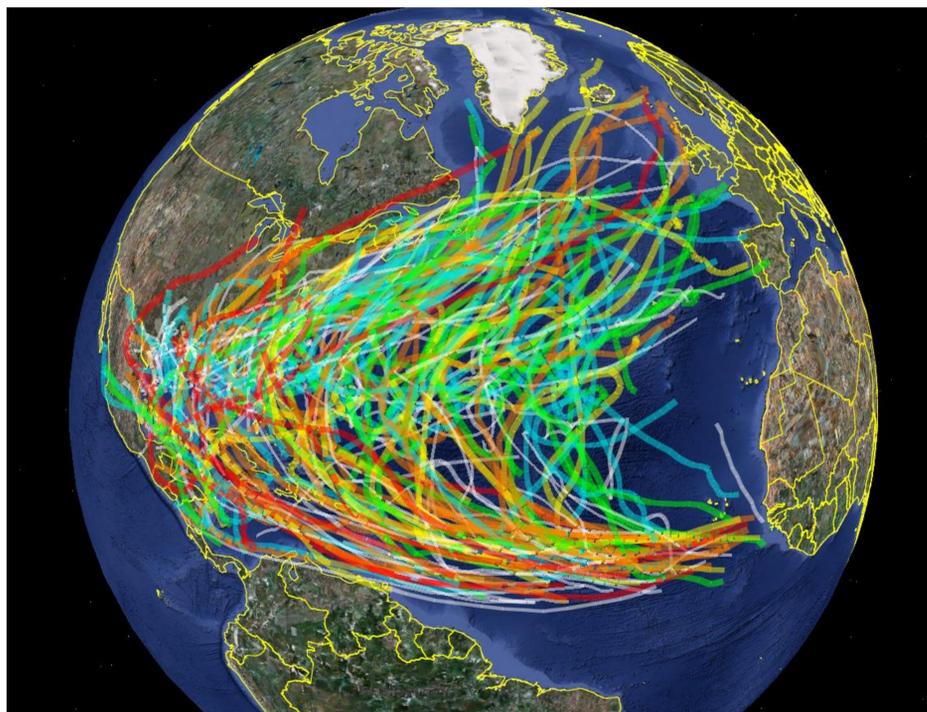


Average performance of SVM classifiers using RBF kernels given 24, 48, and 72hrs of storm track data. Performance generally improved with increased data, reaching over 70%.

Overview

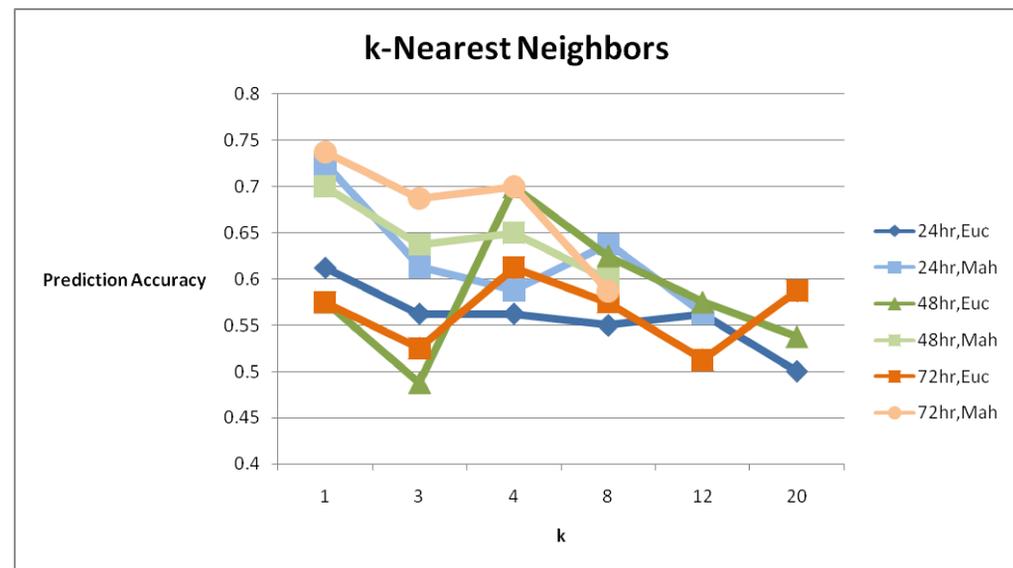
Tropical cyclones in the North Atlantic basin have a powerful impact on the east coast of North America and the Caribbean every year. The ability to predict if and where these storms will make landfall could mitigate their damage and reduce loss of life. Using historical hurricane track and North Atlantic Oscillation (NAO) index data, I predicted whether storms would make landfall in the US. I compared the performance of Support Vector Machines against naive k-NN classification (both Euclidean and Mahalanobis distance). I then created a hybrid k-NN/SVM classifier that was able to produce more accurate predictions even when combining low performing SVM and k-NN classifiers. Using high performing SVM and k-NN, I was able to predict US landfall with over 80% accuracy.

Datasets



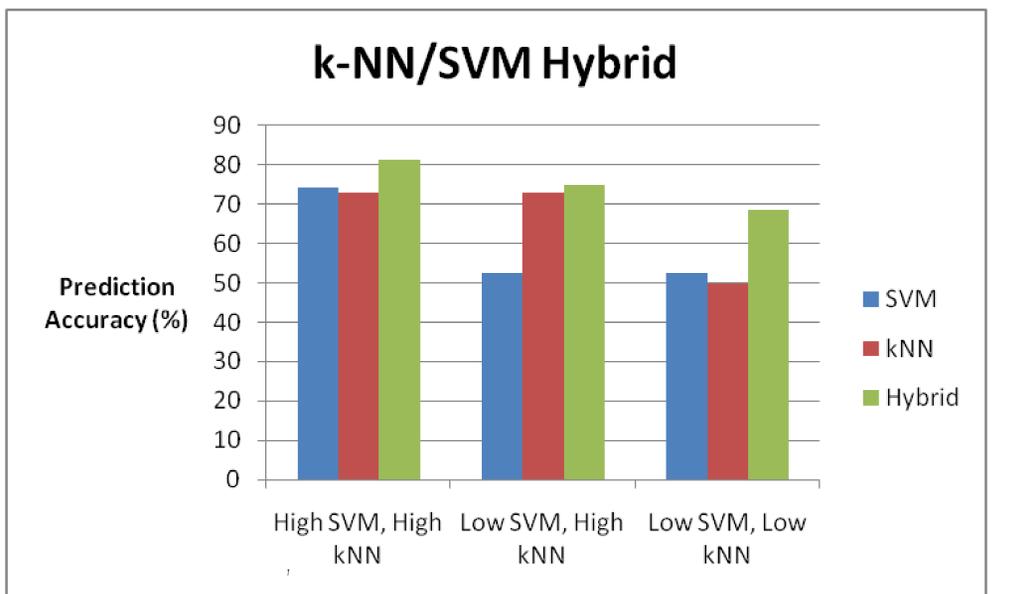
- NOAA HURDAT Best-Track analysis of NAB storms. 6-hour tracks of hurricane location, intensity, and heading since 1851. Training data from 1988-2007, verification data from 2008.
- NOAA North Atlantic Oscillation data, monthly 1988-2008.

k-Nearest Neighbors



k-Nearest Neighbor classifier performance for various values of k. Classifiers using Mahalanobis distance generally outperformed those using Euclidean distance and showed more consistent performance.

Hybrid Classification



The Hybrid classifier outperforms every combination of classifier applied to it. The hybrid classifier works by using an SVM classifier and a k-NN classifier to independently develop prediction sets for a set of storms. Several k-NN prediction sets are generated and scaled by the accuracy of each. These sets are averaged and added to a scaled prediction set of an SVM classifier; the sum is the hybrid's prediction set.

Future Work

Future work should include applying domain knowledge to improve the parameters of the training model (in particular, sea-surface temperatures). I made the assumption that include NAO would improve the quality of my classification but in some cases removing it actually improved accuracy. I would also like to profile the sensitivity of each classifier type to in order to build a better hybrid classifier; for instance, while SVM performance improved with the inclusion of category variables, k-NN performance actually fell by including them.

References

Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>

NOAA HURDAT: <http://www.aoml.noaa.gov/hrd/hurdat/>