An Intelligent Weather Prediction System

For the McMaster Engineering Competition 2017

Kelvin Lin
Joe Lin
Karen Wang
Ming Su

Contents

Introduction

A farmer's livelihood depends on many factors, but there are some factors that are beyond a farmer's control. In those instances, advance notice to adverse events is beneficial to the farmer as the farmer will have time to prepare. This report documents an Intelligent Weather Prediction System that can be used by farmers to get personalized weather predictions based on current weather information. Specifically, the system will use machine learning to predict dangerous weather, and it will let the farmer know if there is dangerous weather. At every opportunity possible, this report will highlight certain decisions made for the prototype and how these decisions may change in the final product.

Current Solutions

Brief research on available solutions for farmers to predict weather have resulted in 2 main approaches: looking outside (perhaps with a thermometer or other measurement devices), or consulting the weather station. The problem with the looking outside is that by the time adverse weather conditions are observed, it is too late for the farmer to prepare. Accurate intuition about future weather conditions takes years to develop, and it is mistakes can be detrimental to a farmer's livelihood. Weather stations, on the other hand, offer predictions many days in advance; however, farms are often large, and located far away from weather stations. In these instances, the predictions may not be as relevant for the farmer. The Intelligent Weather Prediction System aims to address this gap between accurate but untimely information, and timely but inaccurate information.

Our Solution

Aims and Overview

The Weather Prediction System is an application that uses current and previous data to make predictions about the weather. In particular, it performs two functions: one, it tells farmers about projected weather 24 hours in advance, and two, it tells the farmer whether the anticipated weather is dangerous to their crops. The Weather Prediction System uses a state vector machine in order to process information about weather in the previous time step, and uses this data to make a real time prediction for the future time step. This prediction is then conveyed to the farmer in a simple, large font format, perfect for people who are in a hurry, and hard to misread reducing the chance of human error involved with using the system.

The design of the system aims to be adaptive, flexible, and sensitive to the farmer's needs. With these considerations in mind, it was decided that a machine learning algorithm be used to predict the weather. Machine learning provides several advantages compared with other techniques. Machine learning algorithms can automatically learn structure from the data, alleviating the need for a comprehensive model of the weather. In other words, machine learning can be used to uncover underlying relationships undetectable to humans. As the weather system involves, perhaps from global warming, a machine learning system can respond to these changes by intelligently changing internal parameters: a new model does not need to be redefined. Moreover, machine learning allows for the system to adapt to different kinds of crops and their needs. For instance, for some crops, not having rain for several days could be detrimental, whereas for other crops, they can withstand dryness for days on end. Finally, the performance of machine learning is bound not by the programmer, but instead by the hardware that it runs on, which can be assumed (by factors such as Moore's law) to be continuously

improving. Accordingly, because machine learning provides the ultimate adaptivity, flexibility, and sensitivity to the farmer's needs, it was chosen to drive the system.

Data

From an interview with a Saskatchewan farmer, it was determined that weather was the most stressful factor beyond a farmer's control. Upon further research, it was determined that adverse weather conditions could be determined from six features: temperature, humidity, precipitation, wind, snow, and ice. This data can be gathered using sensors such as thermometers, humidity sensors, rain meters, and rulers for snow. Such data can be obtained from the weather station, or by local equipment that farmers have. For the prototype, data from a weather archive from the North Carolina Weather Archive will be used.

In order to effectively use the dataset, it first had to be processed. Extraneous columns were omitted as well as rows with missing data points. This allowed the state vector machine to train using all of the available data. In production, this would change as the farmer's local weather data would be used to train the system instead of North Carolina data.

Moreover, since state vector machines only work with labelled data, an equation was used to determine whether or not a condition was dangerous. The equation took into account the level of participation, the level of snow, the maximum temperature, the presence of ice, and the intensity of wind. These values were derived from the interview with the Saskatchewan farmer. In practice, the derivation of the danger labels will depend on the farmer's environment and the crops.

In the prototype, currently weather predictions are made using the

Algorithm

The backbone of the Intelligent Weather Prediction System is its machine learning algorithm. In particular, a state vector machine (SVM) was used to determine whether or not the weather conditions would pose a threat to farmers. The decision to use a SVM stemmed from the fact that, compared with traditional logistic regression, a SVM can be used to create non-linear decision boundaries. The SVM would take the level of precipitation, the level of snow, the maximum temperature, the presence of ice, and the intensity of wind from the previous day as input, and it will return as output whether or not the weather is expected to be dangerous.

The farmer can set 6 different parameters: whether the farmer wants the algorithm to consider the presence of ice, wind, cold, humidity, snow, and rain. It was decided that farmers should get the option to enable and disable certain factors because different crops require different temperatures to do well. This also illustrates the advantage of SVMs: instead of following one strict set of rules, it can learn new rules based on the environment.

Feedback from the prediction can, in theory, be fed back into the SVM as input in the next computation. This data will allow for more accurate models, as the state vector machine may be able to determine based on previous data whether or not the next time step will also be dangerous.
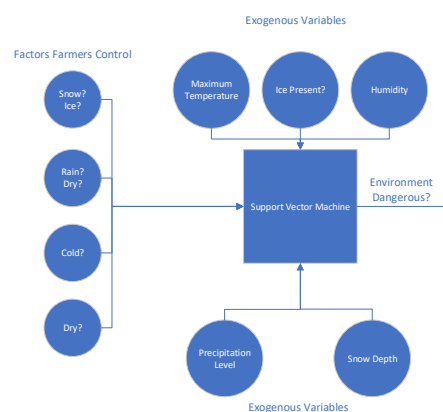

Figure 1: A diagram of the State Vector Machine as a black box

Kelvin and Friends

Module Internal Specification

Module: weather.java

Description: create a java gui for user interface

Functions:

weather(): initialize the interface

SubmitActionPerformed(): submits the farmer's preferences and test data to the SVM. Receives data from the SVM and updates the GUI using the data it obtained.

Module: prediction.py

Description: predict future data for farmer using SVM and returns whether or not the conditions are dangerous

Limitations

The Intelligent Weather Prediction System currently has several limitations. First, since it is data driven, it is crucial that the data used in training the SVM be accurate. In other words, the system is only as good as the data. This limitation can be addressed if multiple farmers in the same region use the system, and crowdsource knowledge and data. By taking the average of what all farmers within a region say about the weather, the system could produce more accurate models with the data.

Moreover, since farmers have control over what features to use in the model, inexperienced farmers may deselect relevant features that could have adverse effects to their crops. A more sophisticated system may ask the farmer for their location and their crop, and suggest settings that other farmers

Kelvin and Friends

have used as default or as "beginner's mode". This will alleviate the learning curve for some farmers, and it will attract others to use the system.

Finally, since this is a machine learning algorithm, further tests with other methods should be conducted to verify the accuracy and the optimality of the selected algorithm. The prototype trained the SVM using all of the points available, and then tested the SVM using "common-sense" data points. However, it would be more effective if the data was split into a training, validation, and testing set, and then multiple models using different algorithms and features were created and compared against. This way, it can be ensured that the most accurate model is chosen. Moreover, with multiple models, a "vote-based" mechanism can be used to poll each model and the danger level can be extrapolated by majority basis.

Future Works

Weather is clearly a time-dependent factor; however, SVMs do not exploit that in its prediction. A better approach, although it would take more time to implement, is a recurrent neural network (RNN). Recurrent neural networks inherently support time series data. With the use of a recurrent neural network, the system can more accurately exploit the order of the data as well as the features within the data. However, training a RNN takes time and lots of data points which would not have been feasible to create in the allotted time. It would be beneficial in the future to recreate this process replicating the system using RNN and other forms of neural networks instead of SVM and more traditional forms of machine learning.

Conclusion

This report details a system to warn farmers of dangerous weather using support vector machines. The farmer is allowed to choose features that they believe are important, and the support vector machine is trained on those features, and delivers a prediction to the farmer based on the current findings. Using a support vector machine allows the system to adapt to different kinds of environments and crops, as well as changes in the environment. In the future, this product will be adapted to use current weather from weather stations, as well as more modern forms of machine learning like recurrent neural networks. This will improve the performance of the system as well as the relevance for farmers.