River Runoff Forecasting

Brian Bell, Brian Wallace, Du Zhang
California State University, Sacramento
Department of Computer Science Sacramento, California
Email: bryan.w.bell@gmail.com, bwtech@gmail.com, zhangd@ecs.csus.edu

***Abstract*—How ”wet” or ”dry” a year is predicted to be has many impacts. Public utilities need to determine what percentage of their electric energy generation will be hydro power. Good water years enable the utilities to use more hydro power and, consequently, save oil. Conversely, in a dry year, the utilities must depend more on steam generation and therefore use more oil, coal, and atomic fuel. Agricultural interest use the information to determine crop planting patterns, ground water pumping needs, and irrigation schedules. Operators of flood control projects determine how much water can safely be stored in a reservoir while reserving space for predicted inflows. Municipalities use the information to evaluate their water supply and determine whether (in a dry year) water rationing my be needed.**

I. Introduction

Water forecasts lead to better planning and management of the State’s water resources – which benefit all Californians. The Cooperative Snow Surveys Program is an important part of this effort. Thus, Californians are dependent upon snow . . . and the snow surveyor.

Today in California more than 50 state, national, and private agencies pool their efforts in collecting snow data. Over three hundred snow courses (http:/ /cdec.water.ca.gov/cgi-progs/snowsurvey_p/SNOWTAB6) are sampled each winter.

One of the forecasting products produced by Snow Surveys is the Bulletin 120 (http://cdec.water.ca.gov/snow/bulletin120/index2.html . Bulletin 120 is a publication issued four times a year, in the second week of February, March, April, and May by the California Department of Water Resources. It contains forecasts of the volume of seasonal runoff from the state’s major watersheds, and summaries of precipitation, snowpack, reservoir storage, and runoff in various regions of the State.

Our project focused on a sub-section of the Bulletin 120, the American River. We forecasted the April July full natural flow runoff of the American River measured at Folsom (http://cdec.water.ca.gov/cgi-progs/staMeta? station_id=AMF).

II. Approach and Learning Algorithms

We used monthly precipitation and snow data gathered from 10 precipitation monitoring stations and 28 snow monitoring stations located in the American River basin. We also made use of the historical full natural flow data for the American River at Folsom.

We used two learning algorithm(s). The first method consisted of feeding the historical data into a neural-net and using the resulting neural-net to create forecasts. The second method consisted of creating a linear regression equation. The regression equation is of the form:

river_flow = a * station1[oct] + b * station1[nov] + ... + z * station2[oct] + b * station2[nov] + ... +...

Where a,b,... are the coefficients that weight the station inputs by their relevance to the final river_flow.

A. Related Work

In searching for related work we found many articles about using neural networks for water supply and stream flow prediction. Most of the articles focused on short- term changes in stream flow, for example predicting flow after a heavy storm.

Interestingly, we didnt find any articles about stream flow forecasts by neural networks concerning basins in California. Weve listed some of the related papers:

[1] Kuo, Chun-Chao, Thian Yew Gan, and Pao-Shan Yu. ”Seasonal Streamflow Prediction by a Combined Climate-hydrologic System for River Basins of Taiwan.” Journal of Hydrology, 387.3/4 (2010): 292-303. [2] Kentel, Elcin. ”Estimation of River Flow by Artificial Neural Networks and Identification of Input Vectors Susceptible to Producing Unreliable Flow Estimates.” Journal of Hydrology, 375.3/4 (2009): 481-488. [3] Besaw, Lance, Donna Rizzo, Paul Bierman, and William Hackett. ”Advances in

Ungauged Streamflow Prediction Using Artificial Neural Networks.” Journal of Hydrology, 386.1-4 (2010): 27-24. [4] Liu, Fang, Jian-Zhong Zhou, Fang-Peng Qiu, and Jun-Jie Yang. ”Biased Wavelet Neural Network and Its Application to Streamflow Forecast.” Lecture Notes in Computer Science, 3971.2006 (2006): . [5] Makkeasorn, A, N.B Chang, and X Zhou. ”Short-term Streamflow Forecasting with Global Climate Change Implications - a Comparative Study Between Genetic Programming and Neural Network Models.” Journal of Hydrology, 352.3-4 (2008): 336-354.

III. Design of Learning Application

For the design of our learning tool we used WEKA. WEKA enabled us to completely forgo any algorithm coding and focus strictly on the set of algorithms and algorithm parameters we used for our project. Most of the work consisted of formatting the raw data and filtering into a set of attributes and instances that were amenable to use by the neural nets.

IV. Experimental Results

For the neural net we started out with no filtering of the data and over 500 input parameters consisting of precipitation and snow water content entries. The results were extremely disappointing.

We significantly filtered out some of the parameters that had sparse data and trimmed the number of input parameters for the neural network down to 222. Our results were better with the smaller number of parameters, but we still couldnt beet the human ensemble. During drought years the neural network produced particularly bad results.

We made another attempt at the neural network, but this time only using five parameters for input. The input parameters used were the indices produced from the precipitation data and the snow water content data that the forecasters use for input to their regression equations. This neural network showed much more promise. Our results were much closer to matching the human ensemble. Our relative absolute error with this neural network was actually smaller, however we still couldnt match the humans in terms of the other error measurements: mean absolute error, root mean squared error, and root relative squared error. The below training session has the results of training neural net1 with 222 attributes. Please note the error rate is low because we did not run it on a forecast but instead ran it on the full data set.

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.03 -M 0.1 -N 2000 -V 0 -S 0 -E 20 -H ”180, 150, 100” -G -R Relation: AmericanRiv_Train-weka.filters.unsupervised.attribute.Remove-R1 Instances: 110 Attributes: 222 [list of attributes omitted] Test mode: user supplied test set: size unknown (reading incrementally)

=== Predictions on test set === inst#, actual, predicted, error 1 1108933 1099143.411 -9789.589 2 552626 566670.998 14044.998

3 973817 994475.684 20658.684 4 1354434 1364647.994 10213.994 5 632159 649646.758 17487.758 6 2003878 1983893.455 -19984.545 7 2622387 2617264.94 -5122.06 8 522651 511853.949 -10797.051 9 674287 664542.387 -9744.613 10 1068327 1071307.545 2980.545 11 1486780 1475169.222 -11610.778

=== Evaluation on test set === === Summary ===

Correlation coefficient Mean absolute error Root mean squared error Relative absolute error Root relative squared error Total Number of Instances

0.9998 12039.5105 13180.1219 2.2054 % 2.0253 % 11

V. Performance Evaluation and Comparison

To evaluate our performance we compare the two neural nets, two linear regressions to each other and importantly also to the human ensemble forecast. In our case the comparison we are most interested in is the comparison of each machine learning algorithm with the human ensemble.

A. Comparison of Neural Net1 (222 inputs, no filtering of data) with Human Ensemble:

For this case the human ensemble forecast performed significantly better than the neural net. The error rates were 69ensemble and the neural net.

VI. Description of development tools/methodologies used

For our learning algorithms we used WEKA. The data was obtained from the Oracle database that the Department of Water Resources maintains which has the precipitation data for the state of California. We extracted the relevant data using SQL from the database and then used the CSV to ARFF converter to convert it into an ARFF file.

VII. Critique of Learning Algorithms Used

The neural net did not perform well on the outlying years, either the extremely wet years or the extremely dry years. In both cases the humans also did well but they did better than the neural nets. The linear regression performed extremely well on normal years but also performed more poorly on outlying years.

VIII. Conclusion

Our current results with our best neural nets yields error rates that are comparable to the human ensemble forecast. The linear regression results were very promising and more investigation should be completed before we can conclude that it beat the human ensemble.

Our most promising line of future work is two use neural net ensemble instead of using a single neural net for producing forecast. The ensemble would consist of three neural nets that are trained and dry, normal, and wet years respectively.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.