A Discriminative Model Approach to Basin Extreme Precipitation Forecasting using a Rich Feature Set

[1] Chenming Li, 1 Xuezhen Li, [2,]1Jia Zhao and [1] Lizhong Xu [*]

[1] *College of Computer and Information, Hohai University, Nanjing 211100, China*
[2] *School of Information Engineering, Nanchang Institute of Technology, Nanchang, 330099, China*
[*] *Corresponding author, E-mail: zhaojia925@163.com, hhice@126.com*

Abstract

A discriminative learning model with a rich feature set was used for extreme precipitation forecasting of a river basin to effectively use available historical observation data. Discrimination models use decision functions that are suitable for forecasting in cases where factors are complex or even unknown. In this study, we used the neural network technique that belongs to the empirical risk minimization method, and a support vector machine that belongs to the structural risk minimization method to be constructed on rich feature sets of rainfall in the Dongjiang basin. The results of forecast experiments show that our method was more effective compared with four types of traditional time series methods and the Naïve Bayes method that belongs to generative models. The support vector machine yielded the maximum F1 value.

Keywords: Extreme precipitation forecasting, Rich feature set, Neural network, Support vector machine

1. Introduction

Basin-scale management [1-4] requires advanced forecasts of the availability of water. Advanced prediction of precipitation, especially extreme events, is important for managing water delivery and reservoir releases because these may result in catastrophic events such as floods [5-6]. Due to complex factors, however, extreme precipitation forecasting remains a formidable challenge [7].

Precipitation forecast methods usually include synoptic and statistical tools. Because precipitation forecasting using synoptic meteorology relies on the subjective judgment of forecasters, there may be considerable differences in the predictions made by different forecasters, and rainstorms indicated on a map may not clearly distinguish the amount of rainfall between small-scale local areas [8]. The statistical method of precipitation prediction includes mathematical [9], physical [10], and dynamic statistics [11]. The mathematical statistical method can be used to predict the probability of future precipitation events based on existing meteorological data, such as past or present weather and climate. Use of the statistical method for predicting short-term climate began in the early twentieth century. This method has made considerable progress in recent decades thanks to the increase number of global observation networks. In particular, extensive progress has been made in the dynamic statistical method through the development of models and improvements in measured data. This is the mainstream method for weather prediction. However, the origin of precipitation is not considered in mathematical statistical methods, and if these methods are used separately, the prediction of extreme precipitation is poor.

Most of the work conducted in previous studies focused on identifying connections, to count and order the factors affecting precipitation. However, due to the dynamic and

complex interactions between factors, it was difficult to construct a mathematical model based on physics [12]. In addition to aberrations, the traditional physical and existing statistical methods for predicting precipitation are not very accurate, especially with respect to the prediction of extreme events. Statistical learning methods have advanced considerably in the last 20 years. The continuous accumulation of high-quality observation data and the improvement of machine learning methods provide new opportunities for precipitation prediction.

Statistical learning models can be categorized as generative models and discriminative models [13]. The former focuses on the total probability of all variables, e.g., the Naive Bayes and hidden Markov models. The latter models are based on observations of variables and conditional probability of a target variable under certain premises, e.g., neural networks, support vector machines, decision trees, and maximum entropy models. Neural networks and support vector machines have been used in station precipitation forecasting [7, 14] and annual precipitation forecasting [5]. Here, some elements are indirectly obtained from circulation model outputs to improve the forecasts. Recent research has focused on dynamical models, of which the limitations remain unknown. In cases where the driving factors of extreme precipitation are not sufficiently clear, discrimination models, which do not intend to generate data, are advantageous. Therefore, in comparison with generative models, discrimination models are less susceptible to correlations between the potential characteristics of distress and can be more conveniently to make use of rich feature sets.

In this research, statistical learning analysis was performed using a rich feature set in the discriminate SVM and ANN models to estimate daily extreme precipitation with a 1-day lead time for the Dongjiang River Basin. Once the daily extreme precipitation has been estimated, reservoir operation and water distribution can be computed depending on the needs of the user.

This paper is organized as follows: the precipitation data for the climate divisions are presented in Section 2; a description of the discriminative modeling approach for estimating precipitation and the performance measures used to evaluate the model performance are presented in Section 3; the rich feature sets of the statistical properties are described in Section 4; a description of the experiment and an analysis of the discriminative modeling results are provided in Section 5 where the SVM precipitation estimates are compared with those obtained from the ANN, SMA, Linear_WMA, EMA, AR and Naive Bayes models, and the conclusion is presented in Section 6.

2. Data

The data sets used to forecast the basin extreme precipitation include precipitation time series, meteorological data, typhoon data, and meteorological knowledge of the basin.

2.1. Basin Daily Grid Precipitation Dataset

The daily precipitation data for the Dongjiang River basin were analyzed in this study. The daily grid precipitation dataset has a 0.25-longitudinal and latitudinal resolution and covers the period from 2008–2013. This data set was obtained from the National Climatic Centre (NCC) of the China Meteorological Administration (CMA)1. The locations of the Dongjiang River basin and the corresponding geographical grid considered in our study are shown in Figure 1 [6].
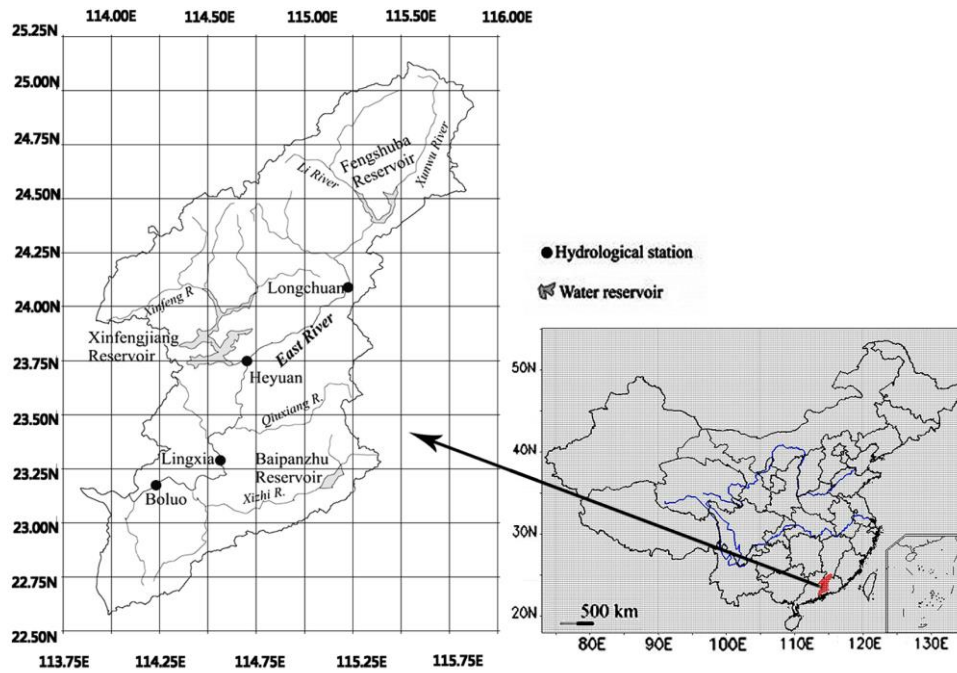
---

[1] http://cdc.cma.gov.cn/home.do

Figure 1. Location of the Study Region and Geographical Grid

In this paper, mean area precipitation, denoted as MAP, was used to analyze the daily precipitation of the River basin. The daily MAP of a given basin is defined as:

$$MAP = \frac{1}{A} \int_A P \cdot dA, \qquad (1)$$

where A is the basin area, and P is the daily precipitation of the finite element dA.

Based on the daily grid precipitation dataset used in our current study, the MAP over the basin can be computed as:

$$MAP = \frac{\sum A_i \cdot P_i}{\sum A_i}, \qquad (2)$$

where grid cell i is the grid cell in the basin or a partial grid cell in the basin. P_i is the precipitation of grid cell i, and A_i is the area proportion of grid cell i that is covered by the basin.

2.2. Basin Meteorological Data

In the Dongjiang river basin area, adiabatic cooling occurs when air rises and expands, which leads to a significant decrease in temperature to the dew point. The basin is subjected to low pressure, and moist air from the Southern Sea may cause extreme precipitation, which occurs in the summer southwest monsoon prevailing period, especially from April to August. We obtained the basin area and associated meteorological data, including surrounding temperature and pressure from the daily surface climate dataset of NCC to generate a feature set for forecasting purposes.

2.3. Typhoon Data

Extreme precipitation in the Dongjiang basin is caused by typhoons from June to October. The latest (1 day before) typhoon forecast data from an area close to the region

may provide reference data for predicting precipitation. Therefore, we obtained daily typhoon forecast data including the path, speed and 10-gale ring area from the Flood Release System of Guangdong Province 2 to generate a feature set for forecasting purposes.

2.4. Basin Meteorological Knowledge

Summer rains constitute approximately 80% of the annual rainfall in the basin. The pre-flood season, which occurs from April to June, is caused mainly by the southwest low-level jet stream and frontal rain, and accounts for 40-50% of annual rainfall. The later flood season, which occurs from July to October, is caused mainly by typhoons, and accounts for 35-40% of annual rainfall. This background knowledge allows us to generate a useful protocol feature set through the use of hierarchical concepts.

3. Discriminative Model

In statistical learning, discriminative models are used for modeling the dependence of an unobserved variable y on an observed variable x by modeling the conditional probability distribution $P(y|x)$, which can be used for predicting y from x. Discriminative models, as opposed to generative models, do not need to generate samples from the joint distribution of x and y. Discriminative models directly model the class posteriors, allowing them to incorporate a rich set of features without consideration of their dependencies on one another.

Based on models of different objectives, learning algorithms of discriminative models can be separated into empirical risk minimization (ERM) and structural risk minimization (SRM) models. ERM models, such as neural networks, compute an approximation, termed the empirical risk, of the training set and minimize the risk. In contrast, SRM models such as support vector machines, the training error (empirical risk) and the regularization term (model complexity) are minimized simultaneously.

3.1 Neural Networks

3.1.1. Overview

Neural networks (NNs) [15] are discriminative models that follow the cognition process of the human brain and have been useful for prediction purposes. In our study, we selected the back-propagation (BP) neural network, which is a feed-forward multi-layer network based on the back-propagation algorithm and has become one of the most widely used neural networks in practice. The activation transfer function (ATF) of a BP network, is typically a differentiable Sigmoid (S-shape) function, which helps to apply non-linear mapping from inputs to outputs. A two-layer BP network is shown in Figure 2, which is a Tansig/Purelin network [16]. This network can be used as a general function approximator that can approximate any function with a finite number of discontinuities arbitrarily well, given sufficient neurons in the hidden layer.
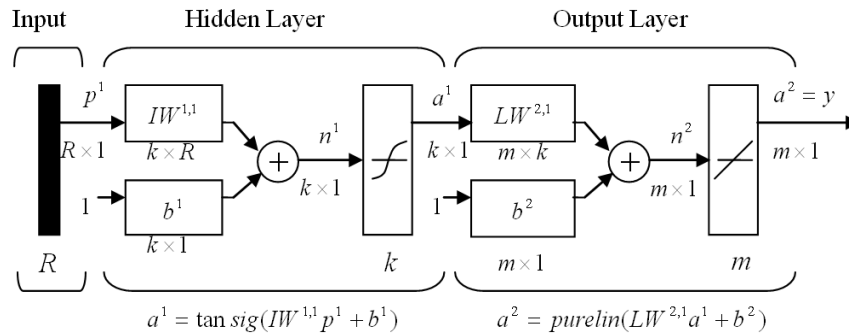
---

2 http://www.gdwater.gov.cn:9001

Figure 2. Two-Layer BP Network Architecture

3.1.2. Operation Principle

First, the hidden layers receive the data from the input. Each connection in the hidden layer has a weight, or strength, of the connection associated with it. Each neuron in the input layer is connected to each neuron in the hidden layer, and each neuron in the hidden layer is connected to each neuron in the next layer. The next layer may be another hidden layer or the output layer. In a BP neural network, the feed-forward pass is followed by a backward pass during which the interconnection weights between neurons are modified based on error values. After training is completed, the "knowledge" gained by the network is stored in the interconnection weights.

The relationships learned by the neural network must be tested to determine the quality of the knowledge. For other portion of samples, *i.e.,* the testing set, the real observed output is compared to the output produced by the trained neural network to determine whether the model output is acceptable. If the output produced by the trained neural network is correct within accepted error ranges, the trained neural network can be used for future prediction.

3.1.3. Selection of Crucial Parameters

The number of hidden layers and the number of neurons in hidden layers are crucial parameters of neural networks [17].

Using additional hidden layers enables greater processing power and system flexibility. However, this may greatly increase the amount of time required for training and testing and may cause over-fitting. It has been shown that a single hidden layer in neural networks is sufficient to approximate any continuous function with arbitrary accuracy. Therefore, one hidden layer is sufficient for most applications.

Determining the number of neurons in the hidden layers is another very important issue in neural network modeling. Too few hidden neurons can prevent the system from properly fitting the input data, and reduces the robustness of the system. In contrast, having too many hidden neurons is analogous to a system of equations with more equations than there are free variables; the system is over specified, and is incapable of generalization. Hecht-Nielsen [18] suggested that for neural networks with a single hidden layer, the number of hidden layer neurons k, can be determined by

$$k = 2R + 1, \qquad (3)$$

where R is the number of input neurons, and k is the number of neurons in the hidden layer, as shown in Figure 2.

3.2. Support Vector Machines

3.2.1. SVM Modeling

Support vector machines (SVMs) [19] are alternative discriminative models that possess a high level of generalization and can therefore be used for real-world forecasting.

Consider a set of training data $T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, where $x_i \in R^m$ is an m-dimensional input vector and y_i is the corresponding output. The concept of an SVM is to nonlinearly map the input data x into a higher dimensional feature space such that

$$f = \sum_{j=1}^{m} w_j \cdot \phi_j(x) + b \,, \tag{4}$$

where w is the weight vector, b is the bias, and ϕ is the mapping function. By performing the mapping, one anticipates that the nonlinear regression problem in the x-space will be converted into a linear regression problem in the ϕ-space.

Both w and b are coefficients that are estimated by minimizing the regularized risk function

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \,, \tag{5}$$

subject to

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \cdots, N \,, \tag{6}$$

$$\xi_i \geq 0, \quad i = 1, 2, \cdots, N \,, \tag{7}$$

where x_i is mapped to a higher dimensional feature space by function ϕ; ξ_i is referred to as a slack variable. The regression quality is controlled by an appropriate selection of three parameters, namely, the cost of error C and the mapping function ϕ.

SVMs avoid under- and over-fitting the training data by minimizing the training error $C \sum_{i=1}^{N} \xi_i$ as well as the regularization term $\frac{1}{2} \|w\|^2$. This aligns with the principle of structural risk minimization (SRM), where the training error (empirical risk) and the regularization term (VC dimension) are minimized simultaneously.

3.2.2. Random Search-Based Parameter Selection

The SVM parameters that are not directly learnt within estimators, e.g., C, kernel and gamma, can be set by searching a parameter space for the best prediction. Such parameters are often referred to as hyper-parameters. Several strategies, e.g., manual search, grid search, and random search have been used for hyper-parameter optimization. For given values, a grid search exhaustively considers all parameter combinations, while a random search implements a randomized search over the parameters where each setting is sampled from a distribution over possible parameter values. Bergstra and Bengio [20] showed empirically and theoretically that randomly selected trials are more efficient for hyper-parameter optimization than trials on a grid. The current study used random search for the parameter selection of SVM. In a random search, a computation budget, *i.e.,* the number of sampled candidates or sampling iterations, is specified. For each parameter, either a distribution over possible values or a list of discrete choices (that are uniformly sampled) can be specified.

3.2.3. K-fold Cross Validation

To ensure the robustness of the SVM model, a cross validation approach is needed to determine the performance of the predictive model. Different cross validation approaches *i.e., k*-fold, leave-one-out, and bootstrapping have certain advantages and disadvantages

but they all help assessing how the results of a statistical analysis will generalize beyond an independent data set.

The current study used a k-fold cross validation technique to test the stability of the SVM model results. The k-fold cross validation technique [21] is used in the field of machine learning to determine how accurately a learning algorithm will be able to predict data that it was not trained on. When using the k-fold method, the training dataset is randomly partitioned into k groups. The learning algorithm is then trained k times, using all of the training set data points except those in the kth group. In general, if the correct value of k is used, k-fold cross validation provides the best estimate of cross validation error. Unfortunately, there is no theoretically 'perfect' way of determining the appropriate k value. The true best value of k may differ for each algorithm and dataset. In our study, k was set to 5.

4. Rich Feature Set

In this section, we describe the features that are central to the performance of discriminative models. These features were generated from the four data sets mentioned in Section 2. We used function-based feature generation methods, such as aggregate, synthesis, statistical, boolean operators, and concept hierarchies by data reduction. Aggregation, which is characterized by the implementation of converged computing, refers to a set of data and returns the new features generated by a single value. Aggregate functions typically include the mean, total, maximum, minimum, count, first value, last value, median, and standard deviation. Synthesis function is a function that uses original features or aggregate features as input to calculate a new meaningful feature. The basic synthetic functions, such as add, subtract, multiply, and divide, were used in our framework. Four statistical features, *i.e.,* level number, range, fairness index, and volatility index, were applied from the serialization precipitation data. In addition, concept hierarchies by data reduction were used. Concept hierarchies, which are combined with background knowledge and redact the original data to a limited number of labels, are usually operated by discretization. The rich feature set of our approach is shown in Table 1.

Table 1. Basin Precipitation Feature Set

Data	Features	Generation function
Basin precipitation data	30 days mean precipitation	Aggregate
	Weighted 30 days mean precipitation	Aggregate
	First day precipitation of 30 days	Aggregate
	Last day precipitation	Aggregate
	30 days precipitation level number	Statistical
	30 days precipitation range	Statistical
	30 days volatility index	Statistical
	30 days fairness index	Statistical
	30 days de-noising fairness index	Statistical
	7 days mean precipitation	Aggregate
	Weighted 7 days mean precipitation	Aggregate

	First day precipitation of 7 days	Aggregate
	7 days precipitation level number	Statistical
	7 days precipitation range	Statistical
	7 days volatility index	Statistical
	7 days fairness index	Statistical
	7 days de-noising fairness index	Statistical
Basin meteorological knowledge	wet/dry periods	reduction
	Pre flood season	reduction
	Later flood season	reduction
Basin meteorological data	Meteorological factor	Aggregate, Boolean, reduction
Data of typhoon	Typhoon factor	Boolean, reduction

5. Experiment

5.1. Experiment Description

In this experiment, data covering the period from April 1st of 2008 to March 31st of 2013 was divided into training, validation, and test subsets to ensure good generalization ability of the prediction models [6]. The first 4 years were used for training, while the 5th year was used for validation. The NN model used the data from the 5th year to select the model with minimum MSE, while the SVM used these data to adjust the parameters based on a random search. Data from the last year were used for testing purposes.

We took the multi-year daily 90th percentile largest value at each station in the basin as an index of station extreme precipitation [22] and 15 mm of MAP as an index of basin extreme precipitation because at this MAP, 50% of the stations experienced extreme precipitation events.

5.2. Performance Measures

In the current study, model performance was measured using Precision (P), Recall (R) and F-Measure (F) [23].

Precision denotes the proportion of predicted positive cases (extreme precipitation) that are correct real positives:

$$P = \frac{\#correctly\,predicted\,positive}{\#predicted\,positive}. \tag{8}$$

Conversely, Recall is defined as the proportion of real positive cases that are correctly predicted as positive:

$$P = \frac{\#correctly\,predicted\,positive}{\#real\,predicted\,cases}, \tag{9}$$

where P and R represent the false alarm rate and missing report rate, respectively, in precipitation forecasting. The larger the P value, the lower the false alarm rate; while the larger the R value, the lower the missing report rate.

The F-Measure is defined as a weighted average of the precision and recall:

$$F = \frac{(\alpha^2 + 1)PR}{\alpha^2 (P + R)}, \tag{10}$$

where an F score reaches its best value at 1 and worst score at 0; α represents the weight. Typically $\alpha = 1$ is used to obtain the $F1$-measure:

$$F1 = \frac{2PR}{P+R}. \tag{11}$$

5.3. Comparative Approach

The discriminative models in this paper were evaluated by comparison with 4 time series prediction methods and the Naïve Bayes algorithm, which is a generative model. The comparative feature sets are as follows:

● Simple moving average method (SMA): the mean precipitation value of the evidence window serves as the prediction for the future precipitation.

● Linear weighted moving average method (Linear_WMA): the linear weighted mean precipitation is considered as the mean precipitation prediction for the future, that is

$$F_{wml}(e) = \frac{\sum_{i=1}^{d} w_i e_i}{\sum_{i=1}^{d} w_i}, \tag{12}$$

where e_i is the precipitation level over time i, and w_i is the corresponding weight.

● Exponential Moving Average method (EMA): this predicted value (denoted $S(t)$ at time t) is calculated based on the following formula:

$$S(t) = \alpha \cdot e_1 + (1-\alpha) \cdot S(t-1), \tag{13}$$

where e_1 is the last precipitation level and α between 0 and 1 is tuned empirically to optimize accuracy.

● Auto-regression method (AR): the classic AR method is performed according to the following formula:

$$X(t) = \sum_{i=1}^{p} \phi_i e_i + \varepsilon_t, \tag{14}$$

where $X(t)$, p and ε_t refer to the predicted value, the order and the white noise, respectively, at time point t.

● Naïve Bayesian algorithm (NB): This algorithm is based on the Bayesian theorem. It is provided with an input space set $A \subseteq R^n$ for n dimensional vector, and the output space for the class tag set $C = \{c_1, c_2, \cdots, c_k\}$. The input is the feature vector $X \subseteq A$, and the output is $Y \in C$. X is the random vector defined on A, and Y is the random variable defined on C. $P(X,Y)$ is the joint distribution probability distribution of X and Y.

The prior probability and conditional distribution probability is as follows:

$$P(Y = c_k), \quad k = 1,2,3,\cdots,K, \tag{15}$$

$$P(X = x | Y = c_k) = p(X^{(1)} = x^{(1)}, X^{(2)} = x^{(2)}, \cdots, X^{(n)} = x^{(n)} | Y = c_k), \tag{16}$$

Due to the exponential number of parameters in $P = (X = x | Y = c_k)$, the Naïve Bayes algorithm makes conditional independence assumptions for each feature. Formula (16) can be simplified as:

$$P(X = x | Y = c_k) = \prod_{j}^{n} p(X^{(j)} = x^{(j)} | Y = c_k), \tag{17}$$

□□

Based on the Bayesian theorem, the posterior probability $P(Y = c_k | X = x)$ can be calculated as follows:

$$P(Y = c_k | X = x) = \frac{P(X = x | Y = c_k) P(Y = c_k)}{\sum_k P(X = x | Y = c_k) P(Y = c_k)}, \tag{18}$$

A maximum a posteriori (MAP) decision rule is used to select the hypothesis that is most probable:

$$\hat{y} = \arg\max \quad P(Y = c_k | X = x). \tag{19}$$

For a given input x, *i.e.,* evidence vector, we can calculate $\hat{y}$ as the estimated precipitation level.

5.4. Results and Analysis

5.4.1. Comparison of Different Feature Sets

To demonstrate the utility of rich feature sets for extreme precipitation forecasting, we applied different feature sets to two discriminative models, *i.e.,* SVM and NN, and one generative model, *i.e.,* NB. Five feature sets were used for comparison. Feature set 1 only contained the aggregate features of basin precipitation data, while feature set 2 added the statistical features from the same data set based on feature set 1. New features derived from basin meteorological knowledge by reduction were then added to feature set 2 to generate feature set 3. Feature set 4 incorporated a meteorological factor, while feature set 5 further added the typhoon factor. The P, R and F1 measures of different models with different feature sets are shown in Figure 3, Figure 4, and Figure 5, respectively.
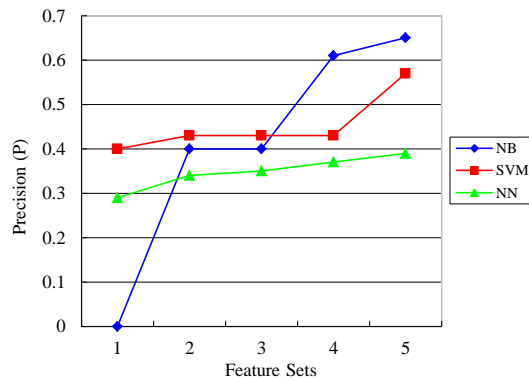

Figure 3. Comparison of Precision (P) Between Three Statistical Learning Models with Different Feature Sets
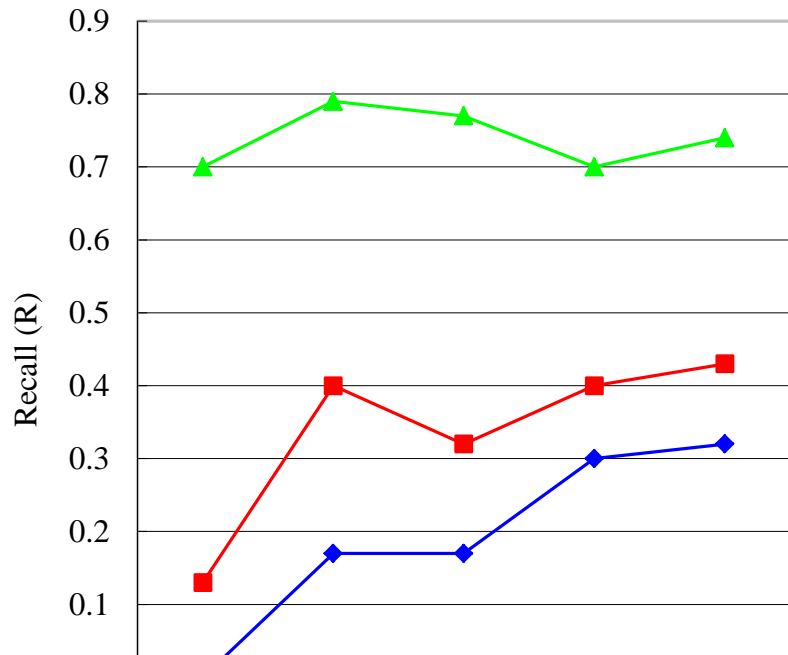
Figure 4. Comparison of Recall (R) Between Three Statistical Learning Models with Different Feature Sets
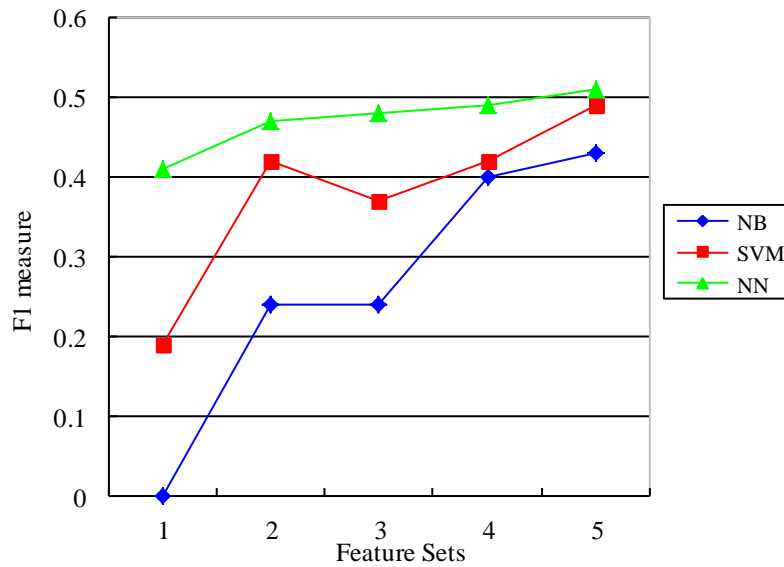

Figure 5. Comparison of F1 Measure between Three Statistical Learning Models with Different Feature Sets

5.4.2. Comparison of Different Prediction Models

Then, we compared proposed discriminative models, *i.e.,* SVM and NN with rich feature sets (full features in our approach) with types of traditional time series methods, *i.e.,* the simple moving average method (SMA), linear weighted moving average method (Linear_WMA), exponential moving average method (EMA), and auto-regression method (AR), and the Naïve Bayes (NB) generative model, which also contained rich feature sets.

The results are shown in Table 2. The highest scores of each measure are highlighted in bold.

Table 2. Comparison of Basin Extreme Precipitation Forecasting Among Different Prediction Models

Models	Measures		
	Precision(P)	Recall (R)	F1- score
SVM	0.39	**0.74**	**0.51**
NN	0.57	0.43	0.49
NB	**0.65**	0.32	0.43
SMA	0.24	0.11	0.15
Linear_W MA	0.18	0.09	0.12
EMA	0.29	0.26	0.27
AR	0.35	0.15	0.21

As we can see from Table 2, SVM obtains the best results of Recall and F1-score, while NB obtains the best result of precision. The improvement of F1 score of SVM compared to NN, NB, SMA, Linear_WMA, EMA and AR are 4.1%, 18.7% , 240%, 325%, 88.8% and 143% respectively.

6. Conclusion

This paper proposes the use of discriminative learning models with rich feature sets for extreme precipitation forecasting. Two discriminative models, *i.e.,* the neural network method that belongs to empirical risk minimization (ERM), and the support vector machine that belongs to the structural risk minimization (SRM) method, are adopted to predict basin extreme precipitation events whose causes are not yet fully understood. The features derived from four different data sets were generated by function-based methods and were incorporated into discriminative modeling. The experimental results show that the discriminative models are superior to the generative model and the time series methods based on the F1 score. The support vector machine model with a full feature set achieved the best prediction performance.

Experiments using more features that are dependent on other data sets are ongoing, especially those that can reflect the causes of extreme precipitation in the pre-flood season or the dry season. Moreover, although the Recall of the Naïve Bayes generative model was not that good, it did yield the highest Precision. Therefore, we will further research combination forecasting technology that incorporates different statistical learning models.

Acknowledgements

References

[1]   H. Wang and J.J. You, "Advancements and Development Course of Research on Water Resources Deployment", Journal of Hydraulic Engineering, vol.39, no.10, **(2008)**, pp.1168-1175.
[2]   H.M. Wang and Y.J. Xu, "Prioritized Aggregation in Non-homogeneous Group Decision Making and Its Application in Water Resource Management", Economic Computation and Economic Cybernetics Studies and Research, vol.48, no.1, **(2014)**, pp.247-258.
[3]   H.M. Wang and J.P. Tong, "An Adaptive Method for Water Resource Allocation and Its Application", Science Press, Beijing, China, **(2011)**.
[4]   X.W. Wu, "A Study Based on Chaos Optimization for Multi-Objective Reservoir Hydropower Station Considering Ecological Environmental Effects", China Water & Power Press, Beijing, China, **(2011)**.
[5]   A. Kalra and S. Ahmad, "Estimating Annual Precipitation for the Colorado River Basin Using Oceanic-atmospheric Oscillations", Water Resources Research, vol.48, W06527, **(2012)**, pp.1-24.
[6]   X.Z. Li, L.Z. Xu, Z.L. Ma, Y.G. Chen and P.J. Huang, "Quotient Space Based Flood Risk Analysis", International Review on Computers and Software, vol.7, no.1, **(2012)**, pp.344-352.
[7]   J.J. Min, J.R. Sun, H.Z. Liu, S.G. Wang and X.Z. Cao, "An Improved BP Algorithm and Its Application to Precipitation Forecast", Journal of Applied Meteorological science, vol.21, no.1, **(2010)**, pp.55-62. (in Chinese)
[8]   D.R. Stauffer and N.L. Seaman, "Use of Four-dimensional Data Assimilation in A Limited-area Mesoscale Model. Part I: Experiments With Synoptic-scale Data", Monthly Weather Review, vol.118, no.6, **(1990)**, pp.1250-1277.
[9]   Z.P. Yu, P. Chu and T. Schroeder, "Predictive Skills of Seasonal to Annual Rainfall Variations in the U.S. Affiliated Pacific Islands: Canonical Correlation Analysis and Multivariate Principal Component Regression Approaches", Climate, no.12, **(1997)**, pp. 2586-2599.
[10]   E.J. Becker, H. Van Den Dool and M. Peña, "Short-term Climate Extremes: Prediction Skill and Predictability", Journal of Climate, vol.26, no.2, **(2013)**, pp.512-531.
[11]   J. Yang, "Predictive Reliability of Summer Precipitation in China Based on Error Distribution of Numerical Model", Acta Phys. Sin., vol.63, no.14, 149202, **(2014)**, pp.1-12. (in Chinese)
[12]   G.F. Lin, G.R. Chen, M.C. Wu, Y.C. Chou, 'Effective Forecasting of Hourly Typhoon Rainfall Using Support Vector Machines', Water Resources Research **(2009)**, Vol.45, W08440, pp.1-11.
[13]   H. Li, "Statistical Learning Methods", Tsinghua University Press, Beijing, China, **(2012)**. (in Chinese)
[14]   Q.F. Xiong and X.Q. Zeng, "Applicat ion and Improvement of SVM Method in Precipitation Forecast", Meteorological Monthly, vol.34, no.12, **(2008)**, pp.90-95. (in Chinese)
[15]   S. Haykin, "Neural networks and learning machines", 3rd edn, Prentice Hall, **(2008)**.
[16]   M.T. Hagan, H.B. Demuth and M.H. Beale, "Neural network design", PWS Published Company, **(1996)**.
[17]   G.J. Bowden, H.R. Maier and G.C. Dandy, "Real-time Deployment of Artificial Neural Network Forecasting Models: Understanding the Range of Applicability", Water Resources Research, vol.48, W10549, **(2012)**, pp.1-26.
[18]   V. Vapnik, "Statistical Learning Theory",Wiley, New York, **(1998)**
[19]   R. Hecht-Nielsen, "Counter Propagation Networks", Applied Optics, vol.26, no.23, **(1987)**, pp.4979-4984.
[20]   J. Bergstra and Y. Bengio, "Randomsearch for Hyper-parameter Optimization", Journal of Machine Learning Research, no.13, **(2012)**, pp.281–305.
[21]   S. Geisser, "The Predictive Sample Reuse Method with Applications", J. Am. Stat. Assoc., vol.70, **(1975)**, pp.320–328.
[22]   S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Milller, "Contribution of Working Group Ⅰ to the Fourth Assessment Report of the intergovernmental Panel on Climate Change", Cambridge University Press, Cambridge, United Kingdom and New York, USA, **(2007)**.
[23]   D.M.W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", Journal of Machine Learning Technologies, vol.2, no.1, **(2011)**, pp.37-63.

Authors

Chenming Li, he is an associate professor in College of Computer and Information Engineering, and Institute of Communication and Information System Engineering, Hohai University. He received his B.S., M.S. and Ph.D. degree in computer applications technology from Hohai University, Nanjing, P.R. China, in 1993, 2003 and 2010, respectively. He is a senior member of China Computer Federation, and Chinese Institute of Electronic, his current research interests include

information processing system and its applications, system modeling and system simulation.

Xuezhen Li, she is an associate professor in Guangdong Technical College of Water Resources and Electric Engineering. She received her Ph.D. degree in hydroinformatics from HoHai University, Nanjing, P. R. China in 2015. Her research interests include intelligent computing and hydro-information technology.

Jia Zhao, he is currently pursuing the Ph.D. degree with the College of Computer and Information Engineering, Hohai University, Nanjing, China, received the M.A. degree in technology of computer application form Nanchang Hongkong University , Nanchang ,China. He has been an assistant professor. His research interests include evolutionary computation, swarm intelligence, wireless sensor networks, Big Date processing..

Lizhong Xu, he received his Ph.D. degree from China University of Mining and Technology in 1997. Currently, he is a professor and Ph.D. supervisor in College of Computer and Information Engineering, Hohai University. His current research areas include signal processing in remote sensing and remote control, multi-sensor system and information fusion, information processing system and its applications, system modelling and system simulation.