Cold Regions Science and Technology 109 (2015) 33–42

Contents lists available at ScienceDirect

Cold Regions Science and Technology

journal homepage: www.elsevier.com/locate/coldregions

Calibration of nearest neighbors model for avalanche forecasting

Amreek Singh [a,*], Bhanu Damir [b], Kusum Deep [b], Ashwagosha Ganju [a]

[a] Snow & Avalanche Study Establishment, Chandigarh, India
[b] Indian Institute of Technology, Roorkee, India

ARTICLE INFO

Article history:
Received 27 February 2014
Accepted 21 September 2014
Available online 28 September 2014

Keywords:
Nearest neighbors method
Model calibration
Uniform random sampling
Artificial bee colony algorithm
Heidke skill score
Avalanche forecasting

ABSTRACT

Nearest neighbors (NN) method is very popular among avalanche forecaster world-over for data exploration and forecast guidance. The NN method based models employ snow-meteorological variables as input to search for past days with similar conditions and then analyze the events associated with past similar days to generate forecast. In order to achieve high forecast accuracy, a NN model needs to be calibrated by way of assigning distinct weights to various input variables. Thus model calibration may be treated as an optimization problem with objective to maximize the forecast accuracy. To investigate the structural characteristics of the problem, uniform random sampling method was applied on *eNN10* (a NN model developed in India for avalanche forecasting). The problem has an issue of multiple optima. Population based metaheristics are suggested to handle such optimization problem better in comparison to classical analytical methods. Thus artificial bee colony algorithm, a metaheuristic inspired by the foraging behavior of honey bees, was explored to calibrate *eNN10*. The study was conducted for two climatologically diverse avalanche prone regions of Indian Himalaya. The results have been verified for operational forecast with significant gains in forecast accuracy in terms of Heidke skill score.

1. Introduction

Avalanche forecasting is primarily a decision making process taking into account various contributory factors. McClung (2002a,b) describes the process of avalanche forecasting in detail. Avalanche forecasters often rely upon NN method based models for assistance in decision making since operation of these models is intuitive and similar to the processes of conventional forecasting (Purves et al., 2003). NN models search for past days with similar snow-meteorological conditions and the information about the events associated with these past days (called nearest neighbors) then helps in generating forecast for the current day. However, in order to achieve high forecast accuracy from NN models, certain model parameters are required to be pre-determined as emphasized by many avalanche workers (Buser et al., 1987; Gassner et al., 2000; Purves et al., 2003). Mathematically, it is an optimization issue. In the field of hydrology, Duan et al. (1992) carried out a somewhat similar exercise of pre-determining proper values of parameters for conceptual rainfall-runoff (CRR) model. The procedure to determine optimal values of model parameters is termed as model calibration (Duan et al., 1992).

To our knowledge, study by Purves et al. (2003) is the only previous published work on automatic calibration of a NN method based model for avalanche forecasting. Purves et al. (2003) applied genetic algorithm (Beasley et al., 1993a,b) for calibration of the *Cornice* model and discussed in detail the effect of various statistical measures of forecast accuracy on model calibration. However, the subject of NN model calibration needs to be investigated in detail in order to ensure globally optimal solution and the operational effectiveness of calibrated model.

The present work investigates the structural characteristics of the optimization problem of NN model calibration, suggests an appropriate optimization method and verifies the operational effectiveness of the calibrated model. In this regard a popular metaheuristic (Blum and Roli, 2003) known as artificial bee colony (ABC) algorithm (Karaboga, 2005) was deployed to calibrate *eNN10*, a NN model for avalanche forecasting developed at Snow & Avalanche Study Establishment (SASE) of India. The study was conducted for two climatologically diverse avalanche prone regions of Indian Himalaya.

The paper consists of 7 main sections. Section 2 describes the *eNN10* model in detail. Section 3 explains the scope of study. Section 4 describes the study area and data characteristics. Section 5 discusses the model calibration as an optimization problem, highlights the issue of multiple optima and introduces the ABC algorithm for global optimization. Results and discussion are presented in Section 6 illustrating that despite some challenges NN models can be efficiently calibrated and implemented for operational forecast with higher accuracy. Conclusions are summarized in Section 7.

2. *eNN10*: a nearest-neighbors model for avalanche forecasting

NN method for avalanche forecasting was introduced by Obled and Good (1980). Buser (1983) packaged it in the form of an interactive model named NXD. Later many other feature-rich variants of NN models were proposed for avalanche forecasting (Brabec and Meister,

* Corresponding author. Tel.: +91 172 269 9804; fax: +91 172 269 9970.
E-mail address: amreek@sase.drdo.in (A. Singh).

http://dx.doi.org/10.1016/j.coldregions.2014.09.009
0165-232X/

2001; Buser et al., 1987; Gassner and Brabec, 2002; Gassner et al., 2000; Heierli et al., 2004; Kristensen and Larsson, 1994; McClung and Tweedy, 1994; McCollister et al., 2002; Mérindol et al., 2002; Purves et al., 2003; Singh and Ganju, 2008; Singh et al., 2005). NN models may employ direct and derived snow-meteorological variables as input. Table 1 describes the 10 snow-meteorological variables of the *eNN10* model. While other researchers, including *eNN10* developers, chose Euclidean distance metric to determine nearest neighbors, McClung and Tweedy (1994) used Mahalanobis distance metric for avalanche prediction at Kootenay Pass, British Columbia, Canada. The use of Euclidean distance metric assumes no correlation between variables whereas Mahalanobis distance metric has implicit features to deal with correlated variables. Invariably, NN models with Euclidean distance metric apply distinct weights to input variables. In *eNN10*, for *D*-dimensional vector-space, the weighted Euclidean distance $d(i,j)$ between two data vectors $\boldsymbol{x^i}$ ($x^i{}_1, x^i{}_2,\ldots, x^i{}_D$) and $\boldsymbol{x^j}$ ($x^j{}_1, x^j{}_2,\ldots, x^j{}_D$) is calculated as —

$$d(i,j) = \sqrt{\sum_{k=1}^{D} w_k\left(norm\left(x_k^i\right) - norm\left(x_k^j\right)\right)^2} \quad (1)$$

where,

w_k: weight corresponding to variable x_k, $k \in \{1, \ldots, D\}$
$norm(x_k)$: value of variable x_k, $k \in \{1, \ldots, D\}$ normalized by its range in the database.

The following steps are implemented in *eNN10* to generate forecast for any day *i* —

1. Calculate distance $d(i,j)$ between current day *i* (under consideration for forecast) and past referenced record $j \in \{1, \ldots, n\}$ using Eq. (1), where *n* is the number of past referenced records in the database,
2. Sort all $d\ (i,j)$ (calculated in step 1) in ascending order,
3. Corresponding to *K* (generally 10) shortest distances $d\ (i,j)$, segregate records from database (call '*K*-nearest neighbors'),
4. Note events (yes/no) corresponding to *K*-nearest neighbors,
5. Calculate relative frequency $P_a(i)\left(=\frac{K_a}{K}\right)$ of avalanche occurrence with respect to current day *i*, where K_a is no. of *yes* (*avalanche*) *days* out of *K*-nearest neighbors.
6. If $P_a\ (i) \geq P_{threshold}$ then forecast the current day as an *avalanche day*, else forecast the current day as *no-avalanche day*.

In the above procedure, a commonly adopted value of $P_{threshold}$ is 0.3 (Buser et al., 1987). It is also obvious that with $K = 10$, $P_a \in \{0, 0.1, 0.2, \ldots, 1.0\}$.

3. Scope of study

In Eq. (1), the introduction of weight parameters w_k, $k \in \{1, \ldots, D\}$ is supposed to improve the model accuracy if proper values are assigned to them. However, the values of w_k, $k \in \{1, \ldots, D\}$ are generally assigned by the experts subjectively on the basis of their experiences of using the model. The values of *K* (the number of nearest-neighbors considered for calculating P_a) as well as of $P_{threshold}$ are also decided by experts out of their experience of using the model. Hence, the optimal values of these parameters are not assured leading to uncertainty about the model forecast accuracy. These facts call for application of an automated objective method to calibrate the model (i.e. to determine the appropriate values of $\boldsymbol{w}$ ($w_1, w_2,\ldots, w_D$), *K* and $P_{threshold}$) for maximized forecast accuracy. Since, there are three parameters i.e. $\boldsymbol{w}$, *K* and $P_{threshold}$ which play their roles at different stages of the process, treating all of them simultaneously is a complex and computationally intensive task. In order to avoid these circumstances the scope of present study is limited to the determination of globally optimal values of $\boldsymbol{w}$ only, while *K* and $P_{threshold}$ continue to carry fixed values as determined by the experts (i.e. 10 and 0.3 respectively).

Table 1
Input snow-meteorological variables of *eNN10* model.

No. (k)	Variable (x_k)	Unit
1	Snow surface temperature	°C
2	Air temperature	°C
3	Air temperature change from previous day	°C
4	Wind speed (averaged over last 24 h)	m/s
5	New snow in 24 h	m
6	New snow in 48 h	m
7	New snow in 72 h	m
8	Snowpack depth	m
9	Snowpack water equivalent	m
10	Free penetration from surface	m

4. Study area and data characteristics

In order to validate the versatility of our study and develop a wider perspective on the subject of NN model calibration, we conducted our study for two climatologically diverse areas of Indian western Himalaya — (1) Chowkibal–Tangdhar (CT) sector in Pir Panjal range and (2) Drass–Kargil (DK) sector in Great Himalayan range. The rough geographical extents of these regions are shown in Fig. 1. Snow-meteorological data from Stage-II (2650 m a.s.l.) and Drass (3230 m a.s.l.) observatories have been used to represent the CT sector and DK sector respectively. While the CT sector may be classified to have maritime climate, the DK sector has characteristics of continental climate (McClung and Schaerer, 2006). A total of 17 avalanche sites affect the part of CT sector (approx. 25 km^2 area coverage) considered for the study, while the DK sector (approx. 400 km^2 area coverage) is affected by 109 major avalanche sites along about 126 km of roads and treks.

The past records corresponding to 10 snow-meteorological variables (Table 1) for a period from November-1999 to April-2013 (covering the months from November to April each winter) were archived in a database with respective dates of observation for each study area. Thus there are 1762 and 1714 past records corresponding to CT sector and DK sector respectively available in the database (only complete records with values for all the variables were considered). The reference time of these observations is 8:30 a.m. IST (GMT + 05:30) on the date of observation. Corresponding to each date/record, the associated avalanche information (Occurrence — Yes/No, size, distance traveled, aspect, natural/artificial etc.) are also stored in database. According to database, all the avalanche occurrences considered under this study (with respect to both study areas) are naturally triggered. Artificial avalanche release is not a common practice in India.

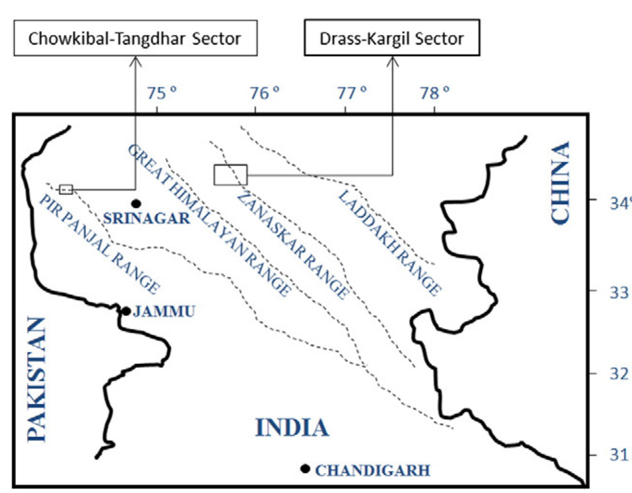

Fig. 1. Study areas — CT sector and DK sector.