International Research Journal of Earth Sciences ISSN 2321 – 2527
Vol. **5(9),** 18-23, October **(2017)** *Int. Res. J. Earth Sci.*

Review Paper

Nearest neighbour classification model in avalanche prediction- a review

Neha Kushe* and Ganesh Magar
P.G. Department of Computer Science, S.N.D.T. Women's University, Mumbai-400049, India
neha.kushe@hotmail.com

Available online at: www.isca.in
Received 29th July 2017, revised 2nd October 2017, accepted 23rd October 2017

Abstract

Prediction of snow avalanche has come a long way since its inception. Nearest neighbour algorithm which mainly incorporates the use of snow-meteorological variables as input for finding the probability of avalanche occurrence, forms an integral part in these prediction algorithms. These methods search the historical data to find days with similar conditions and events that are associated with them to formulate a forecast for the current day. The early methods that began by using simple implementation of nearest neighbour have now been modified so as to find a more accurate forecasting model. In this paper, we present the study of the nearest neighbour classification models that have been used for the prediction of snow avalanche.

Keywords: Avalanche, Classification, Forecasting, Nearest Neighbour, Snow.

Introduction

Avalanches are one of the most lethal hazards in the mountainous terrain that endangers the human life and infrastructure. Mainly consisting of rocks, soil, vegetation and ice, these snow masses rapidly descend the steep slopes[1]. Avalanches are typically triggered due to mechanical failure in the snowpack when the force on the snow exceed the strength of snowpack. Because of this weight, the snow breaks and it starts sliding down the mountain slope. Prediction of these avalanches is a necessity not only for the people living in that area but also for the adventure enthusiasts like skiers and mountaineers. This danger to lives and property is not limited to any particular region as these avalanche sites are present throughout the world. As a result, there are many places where avalanche information centres have been established to help and guide people in case of an avalanche. The geographical locations where Nearest Neighbour Classification Models are being used for Avalanche Prediction are shown in Figure-1.

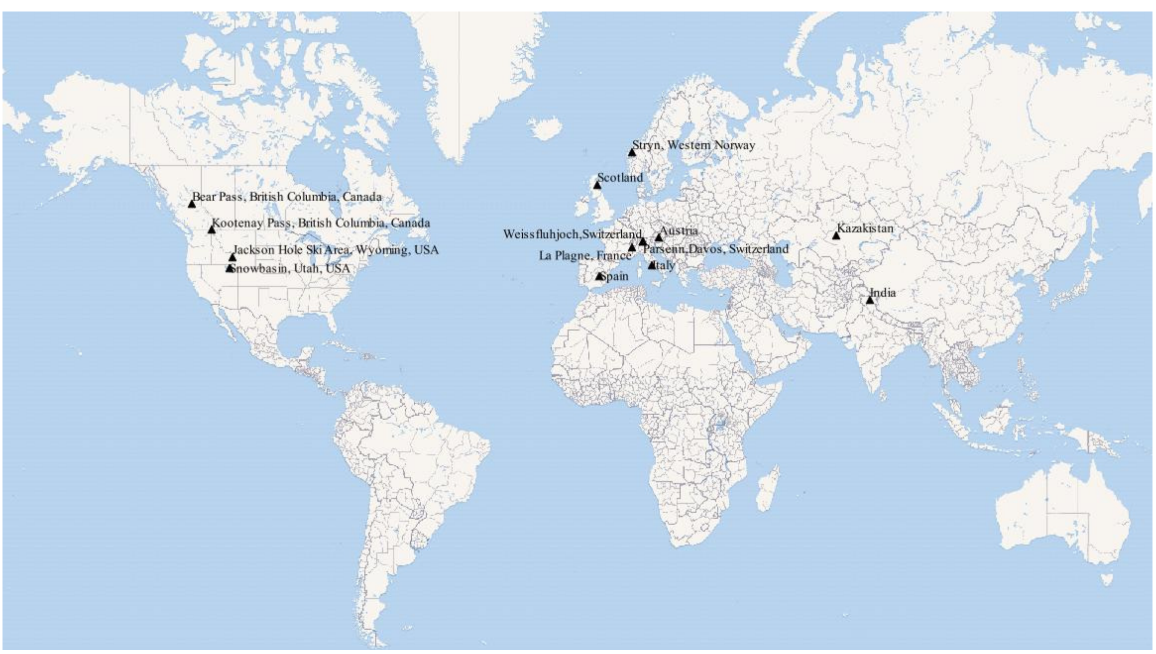

Figure-1: Locations having Nearest Neighbour Classification Models for Avalanche Prediction.

Though the main factor contributing to an avalanche is the snow cover stratigraphy[1], other factors like terrain, meteorological conditions and human intervention also play a major role in possibility of an avalanche. Out of all these factors, the methods relating the local weather condition to the avalanche occurrence data have been used predominantly for forecasting avalanche[2]. These methods are generally based on the idea that if particular weather conditions had led to an avalanche, then similar weather conditions in the future can also give rise to an avalanche. Many methods like linear regression analysis, nearest neighbours, pattern recognition techniques have been used for this purpose[3-5]. However, for this paper, we will be focusing on the methods that incorporate the concept of the nearest neighbor as a classification model on the meteorological and snow parameters for finding the possibility of an avalanche[6].

Methodologies Review

An avalanche day is considered as a day when at least one avalanche has occurred. The Non-Parametric approach named "Nearest Neighbours", was one of the three methods put forth, that consisted of finding the nearest neighbour of a particular point from the total days (with and without avalanche) considered for finding the avalanche[6]. The conditional probability for a particular point was found out by

$$\Pr\{\text{avalanche on day } i\} = \frac{nA}{N}$$

Where: 40 nearest neighbours were considered. The high dimensionality of the parameters was considered to be a problem as the error was directly proportional to the dimensions. Further it was also stated that the avalanche situation depends on the parameters of the days leading to a particular point.

Nearest neighbour model named NXDAYS for avalanche prediction was created in which the distance between the two days was calculated by

$$r^2=\sum_i \Delta x_i^2$$

and 10 nearest neighbours were considered[4]. The priori probability is given by n/N, where n avalanche days are present out of N nearest neighbours. Weighting factor was introduced to provide additional preference to important parameter which could be observed by this model at runtime. Some of the parameters were further normalized either by mean value or maximum value. It was also put forth that if the space spanned by variables is considered as a topological space, then there is flexibility to define distance as per our convenience.

The performance of the NXDAYS was reviewed after being in operation for three years at test area of Weissfluhjoch[7]. The probability of a day to be an avalanche day was found twice, once by only considering the certified avalanche days and other by considering certified days along with the days when records were not available due to natural problems but the possibility of avalanche was present. Based on the model, it was decided that a particular day will be considered as an avalanche day if 3 or more avalanche days are present in its set of 10 nearest neighbours. Consistency was proposed in the various phases of this model like collecting and handling of data, use of the model and its interpretation. Further a need for elaborate variables or computation of additional parameters from the original data was found.

NXD3, an extension of the previous model was developed at Weissfluhjoch by taking into account the earlier weather conditions along with the evolution of the snow cover[8]. For avoiding the seasonal influence, some variables were calculated taking into account their behaviour for a particular period or previous day. Further this model was also successful in finding similar periods along with similar days. A comparative study was also done of the base model, NXD3 and conventional model on the basis of the 6 levels of avalanche dangers. On its basis, it was put forth that the model provides better forecast than the conventional method. This model could be used as an aid in avalanche forecasting as a complementary method.

NXLOG 2.0 was implemented by Swiss Federal Institute as a result of combining NXD and AVALOG that led to the usage of machine learning functions for improving the reliability in avalanche prediction[9]. NXLOG2.0 consisted of a Data Manager and Diagnosis Model that takes date and time of the day to be checked for as an input and provides the probability of avalanche for each gully for the chosen area. The weighting parameters used in the distance calculation are determined by the rules which are in keeping with the context. Future work with NXLOG 2.0 was to optimize the machine learning techniques so that the performance of the machine will be improved by itself.

A variation to the base model with regards to selection and usage of parameters, NN3 was developed for Avalanche Forecasting in Norway to assess the situation up to 24 hrs from present condition by using the data acquired by an automatic weather station[10]. The problem with the base model was that it used to take into consideration the values of the parameters at a period of 24 hrs which did not reflect the course of weather properly. Hence, here the period for data was taken to be 3 hrs between the values of the parameters. The 22 parameters that were obtained from the basic 4 parameters like Wind Speed, Wind direction, Temperature and Precipitation, were used in the program if they fitted the selection criteria constructed for each parameter for distance calculation. The distance calculated was further normalized so that the value of the theoretical maximum distance will always be 100 regardless of the weighting factor used.

NXD2000, an improved version of the base version was implemented in the form of a database program which dealt with the introduction of explanatory variables and functions that gave importance to certain values of the variables[11]. Elaborate variables were calculated by using the values of the variables from the previous days. Use of elaborate variables along with applications of weights by using the experience and information provided by local avalanche forecaster led to improvement at the test sites at Snowbasin and Parsenn where the program has been implemented.

Regional avalanche forecast was introduced for providing avalanche information for an area of about 5000 square kilometre[12]. This model along with finding the nearest neighbours for a particular day(NEX-BEO), was also responsible for forecasting the hazard level according to the European avalanche hazard scale(NEX-MOD). This hazard level was calculated by taking a mean of the nearest neighbour days. The number of nearest neighbour k was selected to be between 1 to 30 depending on its performance. The output of the model depicted its performance to be directly proportional to the increase in the value of k and hence two models having k=30 were selected.

Astral, a local avalanche forecasting tool incorporated the principle component analysis to find the independent variables for all the days which were then used to find the 10 nearest neighbours of the analysed day having a correlation coefficient greater than 0.7[13]. Another method of evaluation was creating a contingency table which transformed the avalanche code (J 13h and J+1 8h) into an index using a weight table on avalanche code. If the composite indexes consisting of the sum of the indexes that were weighted by distance for two (J 13h and J+1 8h) avalanche codes was greater than 0.8, then it could lead to an occurrence of avalanche. However, the absence of knowledge about the internal structure of the snowpack was considered as a limitation that was tried to overcome by the development of Anis that integrated variables describing internal structure of snowpack and spatial variability of weather.

Cornice, which was built upon the base version, consisted of batch testing and using genetic algorithms for automating the process of weighting the parameters[14]. Developed in Java, it consisted of outputs in three formats mainly textual, symbolic and map in the form of an ‘Open’ data format that would facilitate the import and export of data from Software packages. Cornice applied the concept of scaling and weighting on the variables though it did not use the concept of elaborate variables. Symbolic and map output allowed the forecaster to check if some similar factors were prominently affecting the nearest neighbour days and also allowed the forecaster to get information about the geographic clustering of avalanches on the map. The performance of Cornice was found similar to NXD2000 which showed that genetic algorithm provided similar results when used for weighting purpose even in the absence of elaborate variables. It was further implemented in Scotland and it was shown through its performance how the use of genetic algorithms can result in an enhanced performance of the model[15]. In cornice, the distance is measured for the forecast day as well as three previous days from their neighbours so as to find pattern of similar days along with similar days. The working of the algorithm was compared by running it with three fitness measures namely bias, unweighted average accuracy and the Hanssen and Kuipers. Usage of bias decreased the plausibility of the model to over forecast whereas unweighted average accuracy and the Hanssen and Kuipers provided qualitatively better results. Future work was suggested in checking whether over-forecasting of avalanche events indicated the high hazard-low level avalanche conditions or led to a consistent over-forecasting model.

A probabilistic method consisting of modified meteorological nearest neighbour approach by integrating Geographic Information System was put forth to present a technique for analysing avalanche and weather data and implement it as Geo WAX- an interactive program, to find the effect of new snowfall, wind speed and wind direction on avalanche probability[16,17]. Before applying the nearest neighbour, individual operational filters were used to limit the days from the historical data. Here the probability of avalanching for each slide path in a given area was calculated by averaging the avalanche activity for a particular slide path on the most similar days that were provided by nearest neighbours and was visualized using GIS. Nearest neighbour avalanche probability profile was defined by combining the results of the output along with the target day, which were further created for each set of search parameter by changing one weather parameter at a time to finally create a three-Dimensional series space called as Series signature. The analysis showed that the series signatures having similar slide path often were a part of the same geographic area. Further with regards to the effect on avalanche probabilities, new snow fall showed a directly proportional relation to avalanche probability whereas wind speed and wind direction had a differentiating effect on the avalanche probability.

A supplement to the avalanche forecasting method was tested at the Chowkibal Tangdhar axis[18]. The main addition proposed was to only consider those parameters which fall in between a particular range with regards to the particular day and the neighbours were searched only in the boundary of $\pm$15 days from the forecasted day. In order to reduce the effect of critical transitions of parameters on avalanching, a criteria was setup with regards to base range and the range constants which incorporated these uneven effects of the parameters. Further the flexibility to change the range of parameters was also provided so as to extend the range if no similar days is found and vice versa. The days where avalanche information was not obtained due to bad weather conditions were marked distinctively so that it won’t be considered as a non-avalanche day during analysis of the model.

A Forecasting model predicting the Avalanche occurrence for up to four days at a spatial resolution of 5 km was put forth by integrating the Mesoscale Weather forecast model(MM5) with the current NN based method model in the Chowkibal Tangdhar Road axis[19]. By using the weather parameters predicted by MM5 for 4 days, the values for the parameters like temperature, precipitation, snow pack equivalent, wind speed and ram penetration were derived. Snow pack depth was carried as it is without any change. Preliminary results of this model were encouraging even though the prediction accuracies for 2nd, 3rd and 4th day were low and revised assumptions for derived parameters were needed to achieve a better performance.

Artificial Neural Network(ANN) was used in accordance with the nearest neighbour method to find avalanche occurrence in the Chowkibal Tangdhar axis[20]. Instead of using the traditional approach of a threshold value to find avalanche probability, P_{av} and P_{nav} values were calculated by taking into consideration the relative distance of the neighbours with regards to the forecasted point for avalanche as well as non-avalanche days present in the neighbours. These were further used to train the ANN in supervised learning mode so that avalanche day is 1 and non-avalanche day is 0. Probability of Detection (POD) and False Alarm Rate(FAR) was used as accuracy measures for analysing the performance of the model. It was observed for the same amount of FAR, Better POD was obtained by the ANN model as compared to the traditional model, however scope for improvement was still present in fixing the decision boundaries.

An Automatic Forecast system(AFS) for Avalanche Prediction for up to 12hr was applied operationally at two sites in British Columbia by accepting electronic weather sensors data on hourly basis for the parameters like precipitation, snow depth, air temperature and wind data[21]. Filtering mechanism for the data was developed for unavailability of the data due to sensor issues or missing values. AFS worked on the principle of nearest neighbours, however since parameters on the same day had a high probability to being similar to the forecasted point, all the observation for the same day were ignored in the calculation of the neighbours. Jack-Knife cross correlation was used to generate the statistics and Peirce skill score, unweighted average accuracy, proportion correct and bias were the fitness metrics that were used to assess and correlate the performance and accuracy of AFS. AFS achieved average accuracies of 72% and 76% for Bear Pass and Kootenay Pass respectively.

Identification of the features that play a significant role in the formation of Avalanche is an important aspect for prediction method development. Hence for this, two feature ranking methods-Sequential Forward Generation and Relief-F were used to find the various snow and meteorological features that are believed to be connected with the Avalanche Occurrences[22]. Both these methods were applied on the snow and meteorological data and avalanche activity data for the Chowkibal Tangdhar (CT) and Drass Kargil (DK) Sector. The weights after normalizing were applied to the most important features that were identified by the Feature Ranking Method. After the application of weights, the set of parameters were used as input in Nearest Neighbour Method operational at SASE and the respective output was observed. Both the methods had remarkable consistency with regards to the output for both the study areas and hence can be considered as a very important step which could help in removing the subjectivity involved in the avalanche forecasting process.

The use of Genetic Algorithms for optimizing the nearest neighbour model was extended further to find an automated objective method to determine the proper values for weights (w). This led to the calibration of the nearest neighbour model (eNN10) by using the Artificial Bee Colony Algorithm to maximise the forecast accuracy[23]. The Dataset was divided into two parts-Reference Dataset and Verification Dataset. The objective function Heidke Skill Score was adopted as a fitness matrix and it was evaluated the same number of times as the values of w. The comparison of the HSS value after model calibration by ABC Algorithm and without calibration by keeping all the decision variables to 1, showed a substantial gain of 20.1% in CT sector and 25.8% in the DK sector under the calibration phase (Reference Dataset) respectively. Further for verification phase, gain of 112% and 46.3% was obtained for CT sector and DK sector respectively. However, in spite of this accuracy, the mean value for HSS is 0.55 for CT and 0.42 for DK sector respectively, which was far from the ideal score of 1.0. Further the determination of K and $P_{threshold}$ was touted as a subject for future study. To perform the calibration of eNN10, around 400 min were required by the associated sequential MATLAB program. As a result, General Purpose Graphical Processing Unit(GPU) was used to speed up the process of calibration of eNN10 by using ABC Algorithm by 10 times[24]. The NVIDIA Tesla C2050 GPU governed by MATLAB-CUDA programming framework was used to accelerate the process by using parallel processing for the implementation of the ABC algorithm and k-NN algorithm. However, the proposed methodology could be used to process a dataset for up to 1792 data points only and hence though it was enough for operational purpose, this could be considered for future research work.

Discussion of Significant Parameters in NN model for Avalanche Prediction

The snow-meteorological variables considered for avalanche prediction play a very important role in the nearest neighbor model. These variables are decided on the basis of their variation with regards to avalanche activity and also as suggested by experienced people who have thorough information about that region. Further these variables also depend on the environmental conditions and the equipment present to capture the values of these variables. However, there are a particular set of variables that are considered while prediction for most of the regions like Temperature, Wind, Precipitation, etc. The most significant parameters used in the nearest neighbor classification model for snow avalanche prediction as per the reviewed papers are given in Table-1.

Table-1: Significant Parameters in Avalanche Forecasting.

Parameter	Unit
Wind Speed	Km/hr, mph, knots
Temperature	°C
Snowpack Depth/Height	cm, inches
Penetration Depth	cm
Fresh snow	cm
Precipitation	mm, inches
Snow Temperature	°C
Wind Direction	°
Cloud Condition	%
Snowdrift	index
Snow Pack Water Equivalent	mm
Sunshine	hr:mm

There are also some set of variables that have been unique to a particular region which have been used for prediction purposes. Further some variables were used by considering their combined value after a period of 2 days or more like fresh snowfall over 2-3 days, air temperature change over 24 hours, sum of snow index over full winter, 3-day sum of new snow depth, sum of fresh snow and water equivalent over 2-3 days, etc[12-15,18-20,22-24]. The usage of these variables helped in understanding the behavior of a particular parameter over a period of days before leading to avalanche and thus helped in forecasting.

Conclusion

We have reviewed the various avalanche prediction models that have used the nearest neighbour model for classification. This review portrayed the importance and the success of Nearest Neighbour Classification model for avalanche forecasting purpose. The results and problems reported in the literature are noted. The improvements that are done to basic Nearest neighbour model have led to an increase in the prediction probability of Snow avalanche.

References

1. Schweizer J., Bruce Jamieson J. and Schneebeli M. (2003). Snow avalanche formation. *Reviews of Geophysics*, 41(4). doi:10.1029/2002RG000123

2. Schweizer J., Bartelt P. and van Herwijnen A. (2015). Snow Avalanches. In *Snow and Ice-Related Hazards, Risks and Disasters*, 395-436.

3. Jurg S. and Paul M.B.F. (1996). Avalanche Forecasting-An Expert System Approach. *Journal Of Glaciology,* 42(141), 318-332.

4. Buser O. (1983). Avalanche forecast with the method of nearest neighbours: an interactive approach. *Cold Regions Science and Technology*, 8(2), 155-163.

5. Pozdnoukhov A., Matasci G., Kanevski M. and Purves R.S. (2011). Spatio-temporal avalanche forecasting with Support Vector Machines. *Natural Hazards and Earth System Science,* 11(2), 367-382. doi:10.5194/nhess-11-367-2011

6. Obled C. and Good W. (1980). Recent Developments of Avalanche Forecasting by Discriminant Analysis Techniques: A Methodological Review and Some Applications to the Parsenn Area (Davos, Switzerland). *Journal Of Glaciology*, 25(92), 315-346.

7. Buser O. and Good W. (1984). Avalanche forecast: experience using nearest neighbours. *Int. Snow Science Workshop*, Aspen, Col, 24-27.

8. Othmar B., Monika B. and Walter G. (1987). Avalanche forecast by the nearest neighbour method. *International Association of Hydrological Sciences Publication,* 162, 557-569.

9. Bolognesi R., Buser O. and Good W. (1994). Local avalanche forecasting in Switzerland: strategy and tools, a new approach. *International Snow Science Workshop*, 463-472.

10. Kristensen K. and Larsson C. (1994). An avalanche forecasting program based on a modified nearest neighbour method. *International Snow Science Workshop*, 22-30.

11. Gassner M., Etter H.J., Birkeland K. and Leonard T. (2000). NXD2000: An improved avalanche forecasting program based on the nearest neighbor method. *International Snow Science Workshop*, 52-59.

12. Bernhard B. and Roland M. (2001). A Nearest Neighbor Model For Regional Avalanche Forecasting. *Annals Of Glaciology,* 32, 130-134.

13. Mérindol L., Guyomarc'h G. and Giraud G. (2002). A French local tool for avalanche hazard forecasting: Astral, current state and new developments. *In Proceedings of the International Snow Science Workshop, Pentiction, Canada*, 105-108.

14. Purves R., Morrison K., Moss G. and Wright B. (2002). Cornice—development of a nearest neighbors model applied in backcountry avalanche forecasting in Scotland. *In Proceedings of International Snow Science Workshop*, 117-122.

15. Purves R.S., Morrison K.W., Moss G. and Wright D.S.B. (2003). Nearest neighbours for avalanche forecasting in Scotland—development, verification and optimisation of a model. *Cold Regions Science and Technology,* 37(3), 343-355. doi:10.1016/S0165-232X(03)00075-2

16. McCollister C., Birkeland K., Hansen K., Aspinall R. and Comey R. (2002). A probabilistic technique for exploring multi-scale spatial patterns in historical avalanche data by combining GIS and meteorological nearest neighbors with an example from the Jackson Hole Ski Area, Wyoming. Proceedings of the 2002 International Snow Science Workshop.

17. McCollister C., Birkeland K., Hansen K., Aspinall R. and Comey R. (2003). Exploring multi-scale spatial patterns in historical avalanche data, Jackson Hole Mountain Resort, Wyoming. *Cold Regions Science and Technology,* 37(3), 299-313. doi:10.1016/S0165-232X(03)00072-7

18. Singh A. and Ganju A. (2004). A supplement to nearest-neighbour method for avalanche forecasting. *Cold Regions Science and Technology,* 39(2-3), 105-113. doi:10.1016/j.coldregions.2004.03.005

19. Singh A., Srinivasan K. and Ganju A. (2005). Avalanche Forecast Using Numerical Weather Prediction In Indian Himalaya. *Cold Regions Science and Technology,* 43(1-2), 83-92. doi:10.1016/j.coldregions.2005.05.009

20. Singh A. and Ganju A. (2008). Artificial Neural Networks for Snow Avalanche Forecasting in Indian Himalaya. *In Proceedings of 12th International Conference of International Association for Computer Methods and Advances in Geomechanics*, *IACMAG*, 16.

21. Cordy P., McClung D.M., Hawkins C.J., Tweedy J. and Weick T. (2009). Computer assisted avalanche prediction using electronic weather sensor data. *Cold Regions Science and Technology,* 59(2-3), 227-233. doi:10.1016/j.coldregions.2009.07.006

22. Chandra B., Singh A., Singh D. and Ganju A. (2010). Features ranking for avalanche forecasting: method and results for north-western Himalaya. *International Snow Science workshop*, 196-203.

23. Singh A., Damir B., Deep K. and Ganju A. (2015). Calibration of nearest neighbors model for avalanche forecasting. *Cold Regions Science and Technology,* 109, 33-42. doi:10.1016/j.coldregions.2014.09.009

24. Singh A., Deep K. and Grover P. (2017). A novel approach to accelerate calibration process of a k -nearest neighbours classifier using GPU. *Journal of Parallel and Distributed Computing,* 104, 114-129. doi:10.1016/j.jpdc.2017.01.003