Forecasting Avalanches in the Pacific Northwest

Wes Dyer
wesdyer@gmail.com

1. INTRODUCTION

From 1950-2010, avalanches killed 97 people in the Washington state [Cen10]. Over one million dollars is spent annually to manage avalanche hazards that threaten highways in the state. In addition to the loss of life and the public cost, ski resorts also invest a great deal to mitigate the risks of avalanches within their boundaries. Accurate forecasts assist both professionals and recreationalists by reducing costs and increasing safety.

Despite the benefits of avalanche forecasts, it is often difficult to accurately forecast avalanches for specific locations and times. This is because there are simply not enough forecasters to provide information for the vast number of possible avalanche sites. Using machine learning, avalanche forecasts can be given for a much larger region at significantly greater detail than a handful of human forecasters can hope to provide.

This paper investigates applying machine learning to avalanche forecasting in the maritime climate of the Pacific Northwest. It begins with a discussion of the dataset and the measures used to gauge success. Next, the paper discusses feature and model parameter selection. It then compares the forecasting performance of three statistical methods: support vector machines, logistic regression, and nearest neighbors. Finally, it discusses future research possibilites and concludes with recommendations.

2. DATA

Suprisingly, avalanche data is rather hard to come by. Despite more than twenty years of research into machine assisted avalanche forecasting, there are no public datasets. Furthermore, even if there were avalanche datasets, avalanche occurrence is greatly influenced by climate [MS06] and the Pacific Northwest has a decidely maritime climate which differs dramatically from that of the Rockies or the Alps.

2.1 Preparation

Fortunately, several entities within the state maintain pertinent datasets. First, the Northwest Weather and Avalanche Center (NWAC) collects weather and snowpack data from a number of sensor sites scattered throughout the Cascade Mountains with a bias towards locating sensors near ski resorts and highways. Although other organizations operate these sensor sites, they report the readings to NWAC. Second, the Washington State Department of Transportation (WSDOT) collects and records data used to help winter maintenance and avalanche hazard management crews. This data includes weather and snowpack data, which incidentally is reported to NWAC, but more importantly it includes avalanche observations.

The avalanche observations are for the Snoqualmie Pass area in Washington state. The data extends from 2000/2001 winter season to the 2009/2010 winter season. Each season begins October 1st and extends into April of the next year. Three nearby weather and snowpack observation datasets are selected to build avalanche forecasting models: hourly weather observations from the 3800 foot level, hourly weather and snow pack observations from 3000 foot level, and daily observations that include both sensor data and human readings.

The three weather and snowpack observation datasets are combined with the avalanche observations to produce a daily conditions summary. The dataset contains a number of weather related features: high temperature, average temperature, low temperature, new snow, new snow water equivalent (how much water is contained in the new snow), rain, precipitation (total new water), barometric pressure, relative humidity, fog, sky conditions, average solar input, wind speed, and wind direction. It also contains some snowpack based features: snowpack height, snow crust, snow wetness, ram drop penetration (how far a probe drops into the snow from a fixed height), hand penetration, and surface hoar.

The dataset is combined with itself to produce summary daily conditions which include the current day's observations, yesterday's observations, and the observations from two days ago (Table 1). Note that the inclusion of previous days' conditions also allows us to add features indicating observed avalanche activity on the previous two days because they are in the relative past.

2.2 Problems

There are few interesting charactertistics of the dataset. First, the dataset contains a number of missing values which represent times when some sensor was either broken, inoperable, or not yet installed. These missing values are left in the dataset to be removed after feature selection since it isn't yet clear which features are needed. Similarly, there are some wildly inaccurate values like negative wind speeds which are resolved by simply interpolating the data if it is straightforward or changing them to missing values.

Second, the avalanche observations are somewhat limited. These observations only cover a fixed number of important avalanche paths that threaten Interstate 90. Furthermore, about half of the avalanches are artificially triggered by some

Label	Conditions		
Avalanche	Today	Yesterday	Two Days Previous

Table 1: combined feature vector

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Table 2: confusion matrix

POD (probability of detection)	Probability that the event is forecasted when it occurred: $POD = TP/(TP + FP)$
SR (success rate)	Probability that the event occurs when it is forecasted: $SR = TP/(TP + FN)$
HR (hit rate)	Probability that forecast is correct: $HR = (TP + TN)/(TP + TN + FP + FN)$
TPR (true positive rate)	Probability that the event is forecasted when it occurred: $TPR = TP/(TP + FN)$
FPR (false positive rate)	Probability that the event is forecasted when it did not occur: $FPR = FP/(FP + TN)$

Table 3: forecast-accuracy measures

means like explosives. However, note that any artificially triggered avalanche is still an avalanche. It simply means that the avalanche might not be triggered by a less severe stimulus such as a skier crossing the slope. Our goal is to reduce the risk of exposure to avalanches and any artificially triggered avalanche means that avalanche conditions exist, but perhaps will not actually be triggered. In this light, naturally triggered avalanches are actually worse since they could possibly have been triggered previously with less powerful stimuli.

Third, WSDOT and NWAC both state that the avalanche observations might not be complete. This means that there may be other avalanches that are not listed because either conditions such as thick fog make it hard to observe them or it is too dangerous to make the observations. This is another reason to favor predicting avalanches.

Fourth, earlier seasons seem to have spottier sensor data and avalanche observations. Because of this, only the 2004/ 2005 through the 2009/2010 seasons are kept.

The final dataset contains 1472 training examples with 72 features. Of the 1472 days, only 89 are avalanche days.

3. MEASURING SUCCESS

After constructing the dataset, initial experiments are conducted using all of the features and removing training examples with any missing values. The resulting training set contains 232 training examples with 31 avalanche days. Immediately a model is created with 86% accuracy. Unfortunately, this coincides exactly with the percent of non-avalanche days. The model simply predicts no avalanche for every day and gets it right most of the time. Clearly, accuracy alone is not a good measure of success in avalanche forecasting.

Like many other fields, avalanche forecasting instead uses the confusion matrix (Table 2) and the associated forecasting measures POD, HR, and SR (Table 3). In the interest of obtaining a single number to represent the quality of a model, the F-score is computed as the harmonic mean of the POD and SR (Equation 1).

$$F = \frac{2 * POD * SR}{POD + SR} \tag{1}$$

However, the F-score is sometimes also unsatisfactory. Note that if we build a statistical model to forecast the probability of avalanches, we can vary the F-score by increasing or decreasing the threshold probability for which to forecast an avalanche, but the model is still the same. The F-score does not capture this. Instead, we can use another measure called the receiver operator characteristic (ROC) curve which has been used for years in medicine and biology [HM82] and has recently been shown to excel at comparing classification models in machine learning [Bra97]. This curve represents the tradeoff between correctly forecasting avalanches and raising false alarms. It plots the TPR against the FPR. To compute a single number, we take the area under the curve using the trapezoid rule.

To choose good enough features and parameters for a model, a search is performed which maximizes the F-score. The F-score is chosen as an expedient to reduce the time needed to perform the search. It would be much better to use the ROC area under the curve and we aim to use that statistic in the future. Note that the ROC area under the curve is still used, but only to compare models and not for searching purposes.

4. FEATURE SELECTION

If you open any book on avalanches, you will find a detailed list of features that spell certain doom on the snowy slopes. It is therefore somewhat surprising that the feature list can be improved. To pick the set of features to use, we select an initial set of features commonly known to be good indicators of avalanche activity. These features include: day of season, new snow (0,-1,-2), snowpack height, rain water (0, -1, -2), wind direction (0, -1, -2), wind speed (0, -1, -2), sky conditions, ram drop (0, -1, -2), high temperature (0, -1, -2), and avalanche activity (-1, -2). Note that 0, -1, and -2 represent the current day, yesterday, and two days previous respectively. If no number is included then it means only the current day is used.

Next, a backward feature search algorithm eliminates unnecessary features followed by a forward feature search algorithm to add unincluded features to the list. Note that during the search, the same folds must be used throughout in order to have stable statistics to compare. Otherwise instead of comparing features, the search algorithm may end up comparing the quality of partitions. During feature selection, 100 or more folds are used in order to have reliable measures of the quality. Furthermore, since missing values still persist in the data, only the training examples without missing values for the tested feature set are used and since the same set must be used throughout for a stable comparison then only the intersection of the training examples that correspond to each subset of the possible feature selections is used during the search.

After the feature search, the final set of features are: new

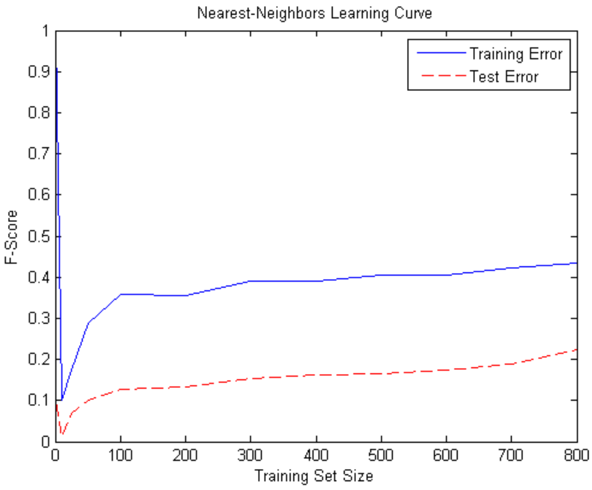

Figure 1: Nearest neighbors learning curve

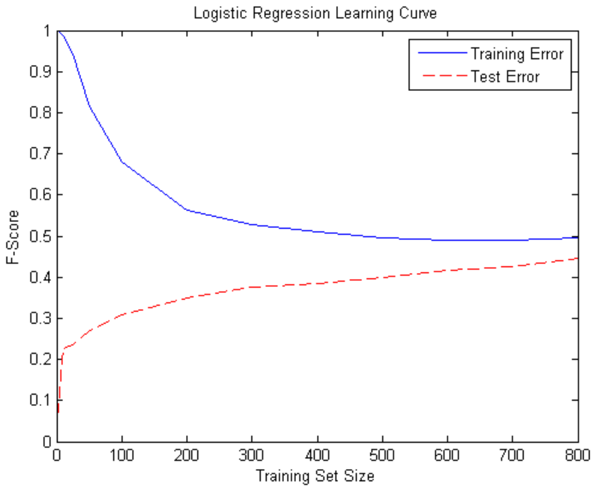

Figure 2: Logistic regression learning curve

snow (0, -1), snowpack height, rain water, wind direction (0, -1), wind speed (0, -1), sky conditions, ram drop, high temperature (0, -1, -2), average temperature, low temperature (-1), precipitation (-1), and avalanche activity (-1, -2). Note that nine features are removed and three features added. Now that the final feature set had been selected, all training examples with missing values are removed from the original dataset of 1472 days. This leaves 862 days with 73 avalanche days.

5. PARAMETER SELECTION

Another problem is selecting the right parameters for various statistical models. Remember that initially the models simply forecast every day as a no avalanche day. For many models, one way to improve the behavior of the system is to penalize misclassification differently for different classes. It seems intuitive to penalize misclassifying avalanche days more than misclassifying non-avalanche days. This is done by changing the weights of various classes. Furthermore, many models also have other important parameters which means that many parameters must be selected.

At first, grid search was attempted to select the optimal parameters, but this took much to long and requires human attention to be tractable. Next, a coordinate ascent method was tried, but this was eventually replaced with a much more efficient simulated annealing method that quickly finds good parameter values even when many parameters are optimized at once.

6. STATISTICAL MODELS

Three statistical machine learning models are compared using the prepared dataset, selected features, and parameter selection methods: nearest neighbors (NN), logistic regression (LR), and support vector machines (SVM).

6.1 Nearest Neighbors

The nearest neighbors algorithm has long been used in avalanche forecasting [Bus83] despite repeated attempts to replace it with other more sophisticated models such as artificial neural networks. The reason that nearest neighbors is so popular is that it is simple, easy to understand, and it gives human forecasters the ability to interpret results. Because of these reasons, professional forecasters often employ nearest neighbors to assist in forecasting. Nevertheless, it has been shown that nearest neighbors requires more training data and may suffer from overfitting problems [MBAC03]. The model was included as a baseline to compare with other models.

The nearest neighbors algorithm takes a training set and converts each training example into a unit vector. Then when testing, nearest neighbors takes the given vector and converts it to a unit vector and computes the distance between it and all of the training vectors. The k closest vectors are used to provide a prediction where the probability of an avalanche is equal to the percentage of the k closest vectors which were avalanche days. Testing shows the that nearest neighbors algorithm performs well when $k \geq 10$. Note that the resolution of the estimated probability is a function of k. Therefore, to provide higher resolution estimates for later comparisions, k is chosen to be 20. The learning curve, the F-score plotted against the training set size, shows an interesting characteristic of the algorithm that the training error is the same as the test error only shifted higher due to the inclusion of the test vector in the training set (Figure 1).

6.2 Logistic Regression

The next algorithm is logistic regression. Previous studies have on occasion used logistic regression to forecast avalanches although this method does not enjoy the widespread use that nearest neighbors has achieved.

LIBLINEAR is used to train and test logistic regression models [FCH$^+$08]. As previously mentioned, while training classifiers for avalanche forecasting, equal weights should not be given to misclassification of avalanche days and non-avalanche days. There are two reasons for this. First, avalanche days are much less common and mistakes that misclassify avalanches are also less common. Second, while repeatedly forecasting avalanches on a non-avalanche days is an annoyance to recreationalists and causes maintenance crews to spend more money then they otherwise would need to, it is not injurious to life or limb. It is difficult to say that there would not be an avalanche give a more powerful stimulus and as long as not too many false positives occur there is some tolerance for misclassification. On the other hand, not forecasting avalanches on avalanche days is a fatal mistake.

To compensate for these facts, the best logistic regression

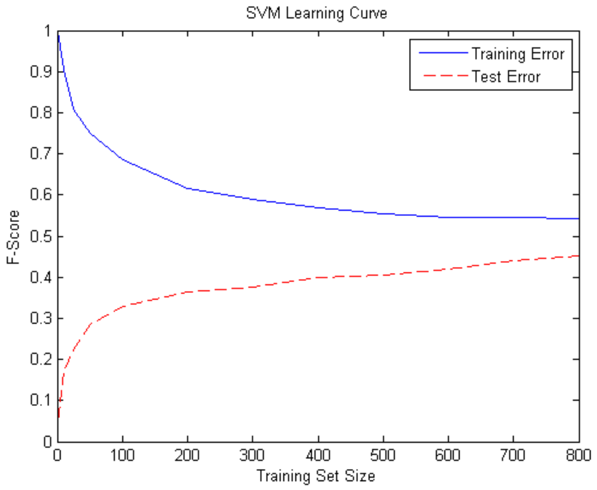

Figure 3: SVM learning curve

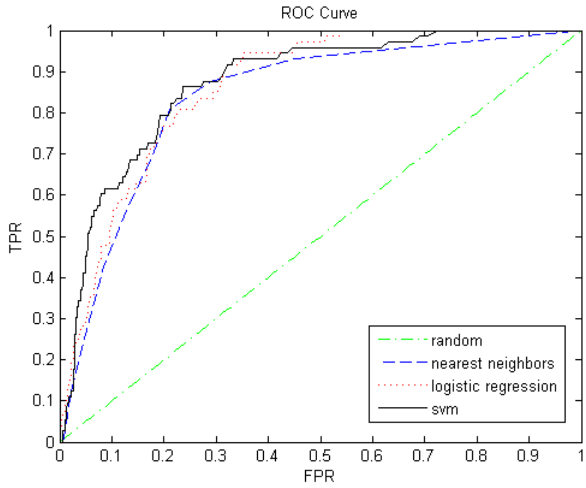

Figure 4: ROC Curve

models weights misclassifying avalanches approximately eight times worse than misclassifying non-avalanche days. The resulting learning curve shows that the algorithm converges nicely (Figure 2), but there is some room for improvement due to high variance possibly by further eliminating features or increasing the dataset size.

6.3 Support Vector Machines

The last model is support vector machines. Despite their relatively recent development, support vector machines have been used to forecast avalanches with success [PPK08] although it remains to be seen if the model can overcome the advantages and broad acceptance of the nearest neighbors algorithm.

LIBSVM is used to train and test the support vector machines [CL01]. Similar to logistic regression, the class weights need to be adjusted to produce good results with the optimal weights penalizing avalanche day misclassification about four times the amount as non-avalanche day misclassification. A gaussian kernel is used with $\gamma = .78726$. Finally, in order to produce probabilitic measures of the likelihood of avalanches, the sigmoid function is used both in training and in test [Pla99].

The resulting model behaves very similar too, although slightly better than, the logistic regression model (Figure 3). Like logistic regression, it appears that there is high variance that can probably be addressed with more data or less features.

6.4 Comparison

Interestingly, when the three models are plotted on a ROC curve against each other it is clear that they all perform fairly well (Figure 4). Each of the methods makes good to excellent forecasts. Support vector machines perform the best, especially when the tolerance for misclassifying non-avalanche days is lower. The area under the ROC curve is similarly instructive (Table 4). It shows that support vector machines perform best, then logistic regression, and finally nearest neighbors.

Nearest Neighbors	.8385
Logistic Regression	.8637
Support Vector Machines	.8726
Random	.5
Perfect	1

Table 4: ROC area under the curve

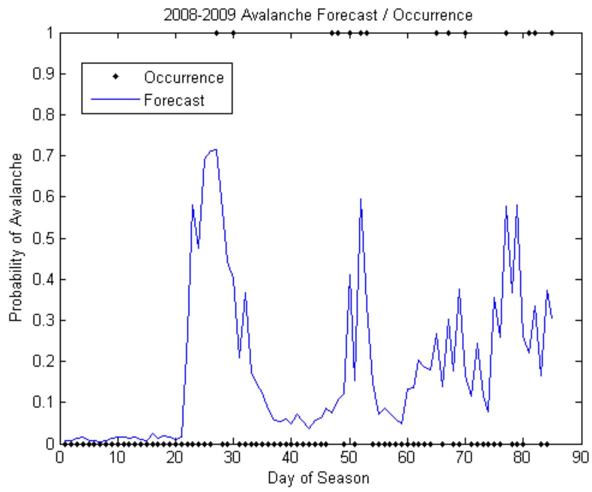

Figure 5: 2008/2009 Avalanche Forecast

I suspect that with further effort in eliminating the higher variance and improving parameter selection especially for maximizing the area under the ROC curve instead of maximizing the F-score, these numbers could be improved although they are already quite impressive.

7. DISCUSSION

The investigations into the possibility of using machine learning models to forecast avalanches in the Pacific Northwest show that they are quite promising. For example, the support vector machine model can be applied to the 2008/2009 season (Figure 5) and the 2009/2010 season (Figure 6). For each season, all other seasons are used as training data and the season in question is tested. The forecasted probability of an avalanche is plotted for each day in the season. Furthermore, actual avalanche days are indicated with dots on the top and non-avalanche days are indicated with dots on the bottom.

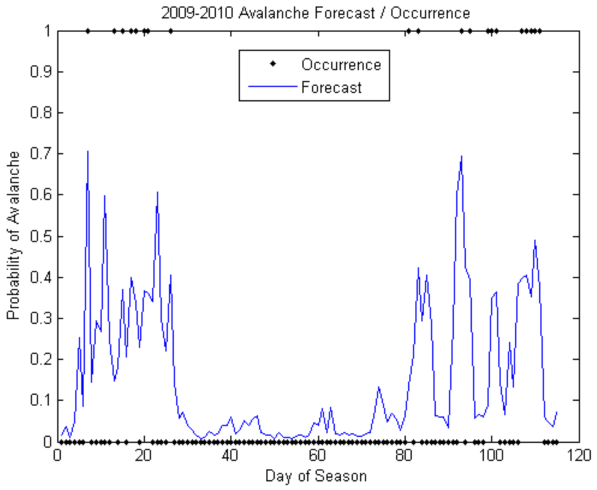

Figure 6: 2009/2010 Avalanche Forecast

The graphs show that the estimated probability closely follows the occurrence of actual avalanche days. Furthermore, days near avalanche days also have a high probability of avalanches. This is good because avalanche probability follows a curve that increases sharply as weather events occur such as snow storms or abnormal heat and then drop as the stimulus is removed and the snow stabilizes over time.

8. FUTURE WORK

There is quite a bit of work still needed to achieve the goal of providing detailed location and time specific forecasts for the Northwest.

First, spatial data needs to be incorporated into the model. The data that WSDOT and NWAC provide contains coordinates for the sensors and the coordinates for avalanche occurrences. From the avalanche occurrence coordinates, important spatial features such as latitude, longitude, elevation, slope angle, and slope aspect can be computed. This information can be used together to provide location specific forecasts.

Second, the current work only uses daily observations; however, the data actually contains observations on an hourly basis. This can be used to provide hourly-based forecasts and allow people like recreationalists to choose time-dependent safe routes through otherwise questionable terrain.

Third, it would be interesting to take the human forecasters predictions for the last decade and use them as a model and compare them against the other models. It would be fascinating to see the ROC curve of human forecasters.

9. CONCLUSIONS

The results are very encouraging. Initially, the modest amount of avalanche occurrence data and its subjectivity seemed to perhaps thwart efforts to apply machine learning to avalanche forecasting in the area. For example, the current data only includes avalanche occurrences near I-90 over approximately twenty predefined avalanche paths that threaten I-90's passage over Snoqualmie Pass in Washington state. Perhaps, additional avalanche data can be found for the area or maybe avalanche sensors could be installed that give more reliable and safe accounts of avalanche activity. Despite the modest amount of data, the models perform very well and confirm intuition about important avalanche features, typical avalanche conditions, and patterns of occurrence. Using machine learning to assist in avalanche forecasting is a promising endeavor to pursue on a broader scale.

10. ACKNOWLEDGEMENTS

John Stimberis and the Washington State Department of Transportation are thanked for providing avalanche occurrence, weather, and snowpack data. Mark Moore, Garth Ferber, and the Northwest Weather and Avalanche Center are also thanked for providing detailed weather and snowpack data. Without their help, this work could not have been done.

11. REFERENCES

[Bra97] A. P. Bradley. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30:1145–1159, 1997.

[Bus83] O. Buser. Avalanche forecast with the method of nearest neighbours: an interactive approach. *Cold Regions Science and Technology*, 8:155–163, 1983.

[Cen10] Colorado Avalanche Information Center. Avalanche accident statistics. `http://avalanche.state.co.us/acc/accidents_stats.php`, December 2010.

[CL01] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[FCH+08] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[HM82] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143:29–36, 1982.

[MBAC03] C. McCollister, K. W. Birkeland, R. Aspinall, and R. Comey. Exploring multi-scale spatial patterns in historical avalanche data. *Cold Regions Science and Technology*, 37:299–313, 2003.

[MS06] D. McClung and P. Shaerer. *The Avalanche Handbook*. The Mountaineers Books, third edition, 2006.

[Pla99] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. pages 61–74. MIT Press, 1999.

[PPK08] A. Pozdnoukhov, R. Purves, and M. Kanevski. Applying machine learning methods to avalanche forecasting. *Annals of Glaciology*, 49:107–113, 2008.