Modeling a System for Decision Support in Snow Avalanche Warning Using Balanced Random Forest and Weighted Random Forest

Sibylle Möhle[1,2], Michael Bründl[2], and Christoph Beierle[1]

[1] Dept. of Computer Science, University of Hagen, Germany
smoehle@acm.org, beierle@fernuni-hagen.de
[2] WSL Institute for Snow and Avalanche Research SLF, Switzerland
bruendl@slf.ch

Snow avalanches endanger traffic infrastructure

Photograph: M. Bründl (SLF)

Photograph: M. Laternser (SLF archives)

Factors influencing the formation of snow avalanches

- Precipitation
  - New snow
  - Rain
- Wind
  - Wind speed
  - Wind direction
- Solar radiation
- Air temperature

Photograph: D. Bommeli (SLF observer)

Avalanche hazard assessment

- Analyze meteorological variables and snowpack properties
- Compare with similar situations observed in the past

- Experience
- Intuition

Related work

- Forecasting large and infrequent snow avalanches using classification and regression trees
- Forecasting snow avalanches in coastal Alaska using classification trees
- Predicting wet-snow avalanches using classification trees and Random Forests

Contributions of our paper

1. We developed a feasible decision support system for snow avalanche warning.
2. We investigated the suitability of Random Forests and variants thereof.
3. We identified quality measures for assessing the obtained models.

Avalanche hazard assessment in the region considered

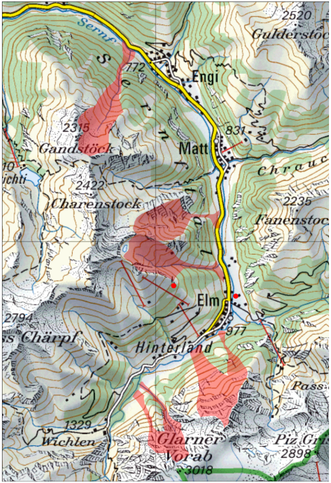
Data

- Meteorological variables
  - Weather data
  - Snow data
- Avalanche information
  - Date
  - Avalanche characteristics

Method

- NXD2000 (nearest neighbours method)
  - Determine 10 most similar situations
  - Consider avalanche activity for a period of 3 days
- Experience and intuition

The region considered



Map: swisstopo,
image editing: A. Stoffel (SLF), S. Moehle

Meteorological variables are measured daily

ELM

- Minimum and maximum temperature in the last 24 hours
- Actual wind speed and direction
- Actual sky cover
- Precipitation in the last 24 hours

Risiboden

- New snow fallen in the last 24 hours
- Snow depth

Derived meteorological variables for ELM

	Abbreviation	Unit	Range
Temperature	e_tmin_0, e_tmin_1	[1/10 °C]	[-251, 157]
	e_tmax_0, e_tmax_1	[1/10 °C]	[-178, 240]
Wind	e_dw_0, e_dw_1		$\{0, 10, \ldots, 350\}$
	e_vw_0, e_vw_1	[kn]	[0, 206]
Sky cover	e_clouds_0		$\{0, 12, \ldots, 96\}$
Precipitation	e_prec_0, e_prec_1	[1/10 mm]	[0, 989]

Derived meteorological variables for Risiboden

	Abbreviation	Unit	Range
New snow	r_hn24_0	[cm]	[0, 550]
	r_hn24_prev	[cm]	[0, 575]
Snow depth	r_hs_0		[0, 432]

Available data

- Period 01.01.1972 – 30.04.2013
- Winter season: 1st November to 30th April
- 6943 data records
  - 6889 non-avalanche days
  - 53 avalanche days
- Positive-negative ratio: $\approx$ 1:130

Avalanche forecasting as a classification problem

		predicted	
		non-avalanche	avalanche
observed	non-avalanche	TN	FP
	avalanche	FN	TP

TN: True negative forecasts
FN: False negative forecasts
FP: False positive forecasts
TP: True positive forecasts

Investigating established quality measures

Sensitivity: $$POD = \frac{TP}{TP + FN}$$

Specificity: $$PON = \frac{TN}{TN + FP}$$

False alarm ratio: $$FAR = \frac{FP}{FP + TP}$$

Results: Additional measures for forecast assessment

Positive predictive value: $$PPV = \frac{TP}{TP + FP} = 1 - FAR$$

Negative predictive value: $$NPV = \frac{TN}{TN + FN}$$

Random Forest

- Ensemble learning method for classification (and regression)

- Deals with unknown variable dependencies and distribution
- Handles discrete and continuous variables

Reviewing the Random Forest algorithm

Creating a forest of size *ntree*

1. Draw a bootstrap sample
2. Construct a decision tree without pruning
3. Add the tree to the forest
4. Repeat 1. to 3. $ntree - 1$ times

Classifying a data record

1. Put the data record down the random forest
2. Assign class with majority decision

Forecasting rare events

Sampling

- Undersampling negative class (non-avalanche days)
- Oversampling positive class (avalanche days)

Cost-sensitive learning

- Consider costs for measures taken
- Different forecast types have different costs
- Assign different weights to positive and negative class

Forecasting rare events with Random Forest

Balanced Random Forest (BRF)

- Equally-sized bootstrap samples for avalanche and non-avalanche days

Weighted Random Forest (WRF)

- Assign a higher weight to the minority class

Defining training and test data sets

Training data set

- 01.01.1972 – 30.04.2002
- 560 non-avalanche days
- 41 avalanche days

Test data set

- 01.11.2002 – 30.04.2013
- 1572 non-avalanche days
- 12 avalanche days

Results: Two feasible types of models

	BRF	BRF	WRF
Identified avalanche days	6	5	5
Missed avalanche days	6	7	7
False alarms	76	55	56
Identified non-avalanche days	1496	1517	1516
Sensitivity	50%	41.7%	41.7%
Specificity	95.2%	96.5%	96.4%
Positive predictive value	7.3%	8.3%	8.2%
Negative predictive value	99.6%	99.5%	99.5%

Discussion

1. The developed models are feasible as a decision support in avalanche forecasting and equivalent from an operational view
2. BRF and WRF are suitable for modeling a system for decision support in avalanche warning
3. PPV and NPV are appropriate measures from an operational point of view

Discussion

The method is suitable for classification problems

- in which rare events or classes are highly unbalanced
- in which dependencies between variables are non-linear and unknown
- in wich the distribution of the variables is unknown

Further work

- Define additional meaningful variables
  - describing the weather situation
  - describing trends
  - containing region-specific expert knowledge
- Apply variable selection
- Discriminate between avalanche paths

Thank you for your attention - questions are welcome

Photograph: M. Bründl (SLF)