Support Vector Machine Classification of Snow Radar Interface Layers

Michael Johnson

December 15, 2011

Abstract

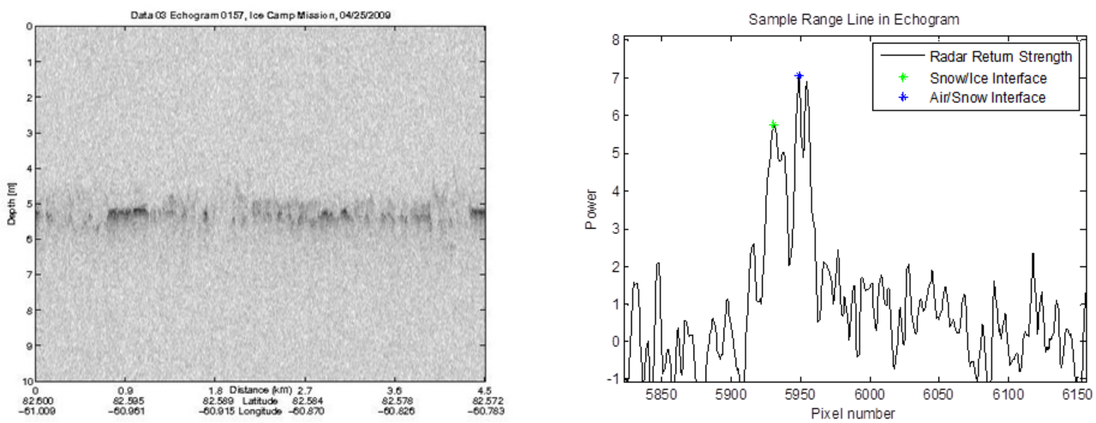
Operation IceBridge is a NASA funded survey of polar sea and land ice consisting of multiple instruments installed on an airborne platform. The Snow Radar [Leu09] instrument is a wide-band Frequency Modulated Continuous-wave (FMCW) radar that has been deployed on IceBridge missions. The properties of this radar (frequency, bandwidth, etc.) allow for the detection of the air/snow, snow/ice, air/ice, and air/water interfaces in the radar backscatter return data. If these interfaces can be extracted from the backscatter data, it will allow one to determine both the distance measured from the aircraft to the surface, as well as the thickness of the snow covering the ice sheet. Retrieving snow thickness will yield important information about the snow pack covering polar ice sheets, allowing for a better understanding of the state of Earth's cryosphere. This paper examines the implementation of a Support Vector Machine (SVM) learning algorithm to extract the various interfaces from the raw radar data.

1 Motivation

The radar backscatter data is presented as measurements of return power as a function of time or range. Although the interfaces do show up as peaks in the radar return data, preliminary analysis of the data has shown that simple peak-finding methods are insufficient in accurately predicting which peaks are the different interfaces. Noise, signal attenuation (especially by thick snow), surface topography, and other geophysical properties all contribute to masking the interface and making detection more difficult. To add to the difficulty, an accurate layer tracker must have the ability to distinguish the following interfaces:

1) Air/Snow (AS)
2) Snow/Ice (SI)
3) Air/Ice (AI)
4) Air/Water (AW)

Each of the above surfaces has different backscatter characteristics that are not easily identified by just the power of the peak in the range line. The power of nearby points, in both range and time, must be used to adequately identify the interface. The relationship between the peak of interest and the nearby points, however, are often very subtle and difficult to explicitly determine. Example echograms for the various interface types are shown below.



(a) Radar echogram showing areas that have no snow (the dark areas near 0.9 km and 4.5 km) and areas that have both snow and ice interfaces (the light areas where two interfaces are visible).

(b) Individual range line of an echogram. In an ideal case, the SI interface will be seen as the highest-powered peak, with the AS interface being of lower power and appearing beforehand.

Figure 1: Example of raw radar data.

The difficulty in determining the explicit factors that describe each interface type is to be handled by implementing an SVM supervised learning algorithm to classify the peaks of a given range line as being one of the identified interface types, or no interface at all (a peak caused by noise, sidelobe interference, or other factors). Other instruments on the airborne platform (LIDAR and digital imagery) can be used as aids in training a given data set.

2 Support Vector Machine

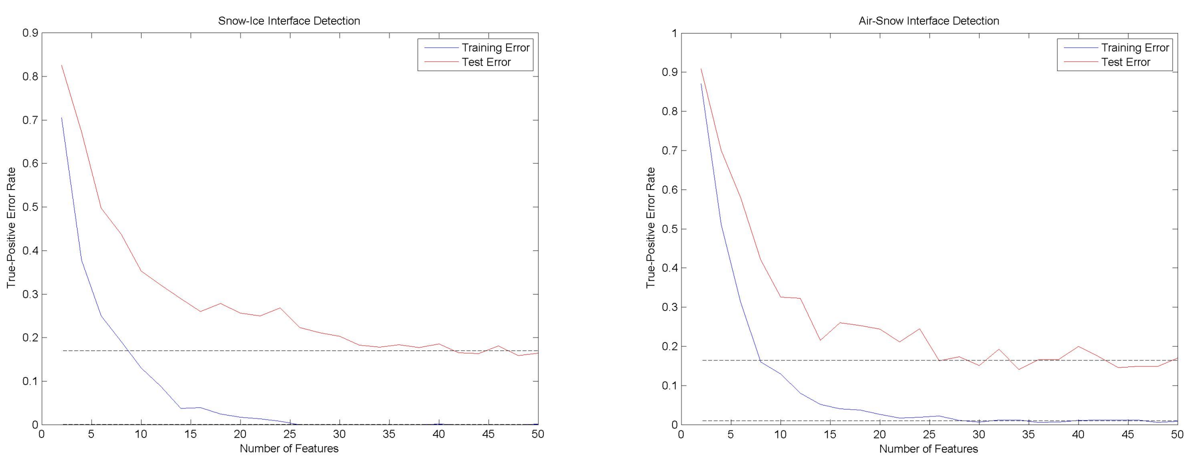
As mentioned, a Support Vector Machine [Ng11] will be applied to this problem. The general method is described here. Classification will be performed on the various peaks in a given range line. In training, the 6 most powerful peaks of a given range line will be extracted. This number was achieved by analysis of several radar files, where it was determined that in general any other peaks beyond this number are either noise or cannot be readily classified. Each of these peak will be analyzed and identifed as being one of the interfaces described, or being no interface at all. Once a peak is identified, a feature vector will be formed from the data and stored. Once a certain number of peaks are classified, the SVM will be trained on the data and used to classify the rest of the radar data.

LIBSVM [CL11], a readily available software package, was employed. LIBSVM provides a robust solution to this problem, with a MATLAB interface that allows the data to be trained, processed, analyzed, and visualized with ease. A model for each type of interface was created, and every peak being analyzed was predicted with each model. Once predicted, the interface classification that returned the highest decision value was used to classify the peak.

3 Feature Selection

As illustrated in figure 1, it is clear that points surrounding each peak are required to properly classify it. Not only are the points in the same range line of interest, but so are the points in adjacent range lines. This is because the radar footprint is sufficiently large so nearby range lines should be well correlated in terms of interfaces they detect. However, noise should not be well correlated between range lines, so this should aid in rejecting peaks that are caused by noise or other undesired sources.

The ideal number of features to use was achieved by performing training and test error analysis using different numbers of features that are processed in different ways. Because of the large amount of data, computational efficiency must be considered. A model could be generated with a large number of features, but this would result in unacceptably large processing times. Some classification errors are tolerable, as additional filtering of the data will be performed that should reduce error rates.



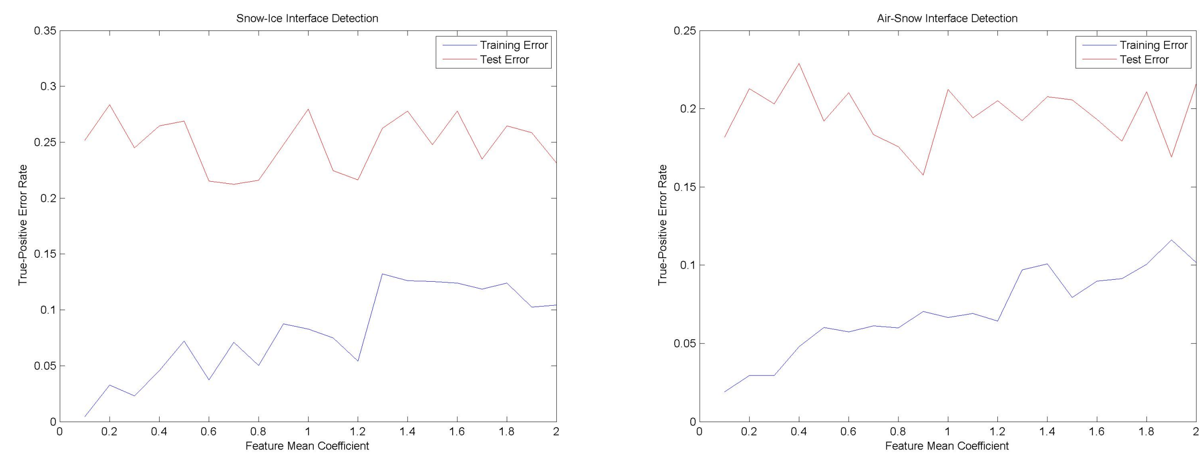
(a) SI Interface

(b) AS Interface

Figure 2: Training and test errors as a function of the number of features (to the left and right of a given peak). A value of 25, for example, means that the 25 points to the left and the 25 points to the right of the peak are used as features for the model.

Figure 2 shows the behavior of the model as the number of features is increased. As expected, if only a small number of features are used, the model is not very good at making predictions. Analysis shows that little improvement is seen beyond using 25 data points to the left and to the right of a peak. For this reason, 25 features will be used. Similar behavior is seen for true-negative, false-positive, and false-negative results, across all interface types.

In order to further improve computation efficiency by reducing the number of features in the model, analysis was performed by averaging some features together. One motivation for this is that because the snow depth is not constant, averaging together a series of features will "bring together" various return profiles, which should both simplify the model as well as possibly giving better performance. Also, averaging points together may help with noise rejection. It was determined that averaging would be done on a larger number of points the further away from the peak the feature lies, also under the assumption that the further away from the peak, the less important a feature becomes because the points start to decorrelate more. This was performed by grouping together points by using a quadratic function $f(x) = (cx)^2$, where c is the "feature mean coefficient" and x is the number of points away from the peak. This gives a pattern that groups together points in increasing numbers the further away from a peak they are.



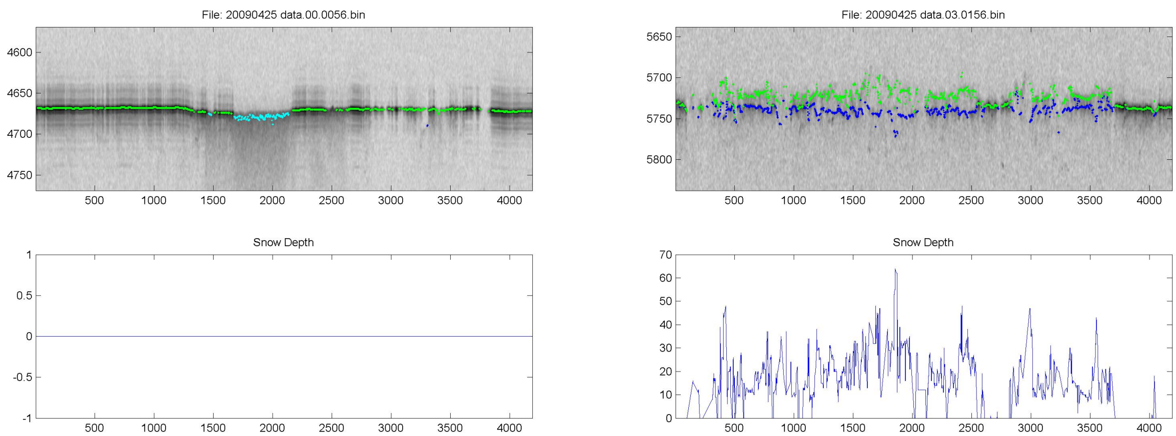
(a) SI Interface **(b)** AS Interface

Figure 3: Training and test errors as a function of the feature mean coefficient.

Figure 3 shows the effect of averaging the features together. Training error gets somewhat worse, but not by much, while test error remains about the same. A somewhat large increase in error is noted after a c value of 1.2, so this is the value that is used. Performance remains good, while computational efficiency is greatly increased by reducing the size of the feature vector from 357 elements (51 points along 7 range lines near the peak being analyzed) down to 49 elements.

4 Results

Multiple radar data files for a specific campaign were processed with the SVM model trained in the manner described. Examples are shown in figure 4. 56.08% of the 430,000 measurements were determined to have been successfully tracked. Most files had better tracking statistics (in the 70%-90% range), but the overall average is brought down by some files having substantial data dropout (for often unknown reasons), rather than poor performance by the SVM.



(a) Example file showing the detection of AI (green) and AW (cyan), resulting in no snow depth.

(b) Example file showing the detection of the AS (green) and SI (blue) interfaces, where snow cover is seen.

Figure 4: Two radar files showing the successful tracking of the various interface types.

References

[CL11] Chih-Chung Chang and Chih-Jen Lin. Libsvm : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:2:27:1–27:27, 2011.

[Leu09] Carl Leuschen. *IceBridge Snow Radar L1B Geolocated Radar Echo Strength Profiles*. Boulder, Colorado USA: National Snow and Ice Data Center. Digital Media, 2009.

[Ng11] Andrew Ng. *Support Vector Machines, CS 229 Lecture Notes, Part V*. Stanford Univeristy, 2011.