

Improving Onboard Analysis of Hyperion Images by Filtering Mislabeled Training Data Examples

Umaa Rebbapragada¹, Lukas Mandrake², Kiri L. Wagstaff², Damhnait Gleeson²,
Rebecca Castaño², Steve Chien², Carla E. Brodley¹

¹Dept. of Computer Science
161 College Ave.
Medford, MA 02155
{urebbapr,brodley}@cs.tufts.edu

²Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
FirstName.LastName@jpl.nasa.gov

Abstract—This paper presents PWEM, a technique for detecting class label noise in training data. PWEM detects mislabeled examples by assigning to each training example a probability that its label is correct. PWEM calculates this probability by clustering examples from pairs of classes together and analyzing the distribution of labels within each cluster to derive the probability of each label’s correctness. We discuss how one can use the probabilities output by PWEM to filter, mitigate, or correct mislabeled training examples. We then provide an in-depth discussion of how we applied PWEM to a sulfur detector that labels pixels from Hyperion images of the Borup-Fiord pass in Northern Canada. PWEM assigned a large number of the sulfur training examples low probabilities, indicating severe mislabeling within the sulfur class. The filtering of those low confidence examples resulted in a cleaner training set and improved the median false positive rate of the classifier by at least 29%.¹²

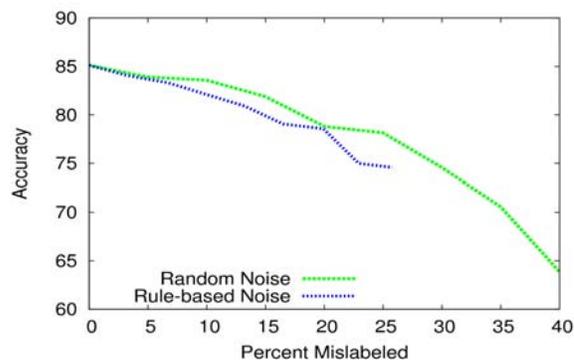


Figure 1. Effect of random and rule-based noise on a landcover data set containing 3398 items. The y-axis measures accuracy on an independent (clean) test set. The x-axis measures the percentage of mislabeled examples in the training set.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. PWEM	2
3. DATA CLEANING OPTIONS.....	3
4. SULFUR DETECTION ON EO-1	3
5. DATA	5
6. RESULTS	5
7. CONCLUSION	7
ACKNOWLEDGMENTS	7
REFERENCES	8
BIOGRAPHY	8

1. INTRODUCTION

Supervised learning algorithms are popular for building automated classifiers from a set of labeled training examples. Achieving maximal classifier accuracy depends upon the selection of an appropriate learning algorithm, as well as high quality training data.

A human expert usually labels training data. Unfortunately, the labeling of the training data can be an error prone process due to data entry error, human subjectivity (different experts may disagree on which label is correct [7]), or the use of an external data source to label data. For example, when labeling pixels in image data, the expert may use visual data rather than numeric values.

Figure 1 shows the effect of mislabeled training data on classifier accuracy. We introduced two types of noise into an eleven-class landcover data set: random and rule-based noise. Under random noise, a training example’s label may be flipped to any of the other labels. Under rule-based noise, we only flip labels if the example’s assigned class is easily confused with any of the others according to rules specified by a domain expert. We introduce up to 40% random noise and 25% rule-based noise in the landcover data set. In both cases, we observe a significant decrease in classifier accuracy as class noise levels increase.

This paper presents PWEM [5], a technique for detecting mislabeled training examples. PWEM assigns to each training example a probability of cleanliness. These probabilities are calculated by clustering instances (without their class labels) from pairs of classes using Expectation

¹ 978-1-4244-2622-5/09/\$25.00 © 2009 IEEE

² IEEEAC Paper #1607, Version 3, Updated January 8, 2009

Maximization (EM) [8] to perform the clustering. Once the EM algorithm estimates the data distribution, we apply the assigned labels to the clusters and evaluate how well the labels correspond to the natural groupings of the data. Intuitively, if an example that was assigned to class A clusters with examples that are mostly assigned to class B, there is a good chance that the "A" example is mislabeled. From the set of clusterings and the assigned labels, we calculate for each example the probability that the example was assigned the correct label.

We present three options for using the probabilities calculated by PWEM. One can filter out examples with low probability, use the probabilities as weights during training, or iteratively query the domain expert for new labels on examples with low probability. We also present an in-depth discussion of how the filtering option was used to improve the performance of a sulfur detector designed for use onboard the Earth-orbiting satellite EO-1. Members of our team hypothesized that mislabelings in the training data caused the false positive rate of the classifier to be unacceptably high. Because a positive classification can be programmed to trigger follow-up imaging, it is crucial to minimize the classifier's false positive rate. We applied PWEM to this problem, and discovered mislabeled examples in the training set's sulfur class. Filtering out the low-confidence sulfur examples improved classification accuracy and reduced the false positive rate of the classifier by approximately one half.

We organize our paper by presenting the PWEM method in Section 2 and possible ways to use its probability outputs in Section 3. In Section 4, we discuss the sulfur detection task to which we apply PWEM and the steps that led the members of our team to conclude that the sulfur training set contained mislabeled examples. In Section 5, we describe the sulfur training data, and the contributions of PWEM to this problem in Section 6.

2. PWEM

PWEM works by calculating a probability of correctness for each example's assigned label. Label correctness is defined in relation to the features of the training example. PWEM also assumes that a mixture of Gaussians generates each class's data.

For each example x , PWEM outputs the probability $P(l/x)$ that the label of x is l from set L , where L is the set of class labels. We use clustering to find $P(l/x)$ since, intuitively, one might expect instances from the same class to cluster together. We can use clustering to generate a set of class probabilities by having each instance inherit the distribution of classes within its assigned cluster. The drawback of this method is that there is no guarantee that a multi-class data set will cluster perfectly along class lines. Feature selection may improve class separability, but it is possible that two or

more classes may not separate under any circumstances because their data distributions overlap.

We improve class separability by clustering pairs of classes. For each pair of classes, we cluster only those examples assigned a label from one of the classes in that pair. Thus, each example belongs to only $|L|-1$ clusterings. If a clean example has a lower confidence in one clustering due to class inseparability, it may still receive a high confidence in its other clusterings if its assigned class separates well from others.

Given a set of $|L|-1$ clusterings for instance x , we calculate the probability that x belongs to class l as follows:

$$\begin{aligned} P(l/x) &= \sum_{\theta} P(\theta) P(l/x, \theta) \\ &= \sum_{\theta} P(\theta) \sum_{c=1}^{l-1} \sum_{k=l}^{|L|} P(l/c, \theta) P(c/x, \theta) \quad (1) \end{aligned}$$

where l is a class label, x is an instance, c is a cluster, k is the number of clusters, and θ is a given clustering. $P(l/x)$ represents the probability that instance x should have class label l . $P(l/x, \theta)$ represents the probability that x should have label l given clustering θ . This probability is calculated by summing the probability that x belongs to cluster c (as calculated by our clustering method EM) times the probability that c should be labeled as l . Summing over all clusters results in the probability that x should be labeled l . If $P(l/c, \theta)$ and $P(c/x, \theta)$ form probability distributions, it is trivial to show that $P(l/x)$ also forms a probability distribution over the class labels. If we assume each clustering θ is equally likely, $P(\theta)$ is $1/(|L|-1)$. Each $P(l/x)$ acts as a confidence on the class label l for instance x .

Although $P(l/x)$ forms a probability distribution over the class labels, only the probability for the assigned label is useful given this scheme. This probability distribution cannot be used to correct labels. The reason is that example x only participates in clusterings where its assigned label is represented. We leave to future work the calculation of a meaningful probability distribution that may be useful in correcting labels.

We use the Expectation Maximization (EM) algorithm to perform the pairwise clusterings [8]. EM is an optimization method that finds the maximum likelihood estimates of parameters in probabilistic models that depend on some hidden variables. Because we assume our training data conforms to a Gaussian mixture model, the hidden variables estimated by EM are the means and covariance matrices of the k Gaussians in the mixture, where k is a user-supplied input parameter to EM. We perform clusterings with multiple values of k , typically between two and four, in case a particular class is multi-modal. We select the clustering with the k value that minimizes the Bayesian Information Criterion [6], a model selection criterion from statistics.

DATA	M/C	RA10	RA20	RA30	RU10	RU20	RU30
Segm	M	0.45 (0.22)	0.46 (0.19)	0.47 (0.15)	0.39 (0.18)	0.45 (0.16)	0.53 (0.17)
Segm	C	0.82 (0.11)	0.77 (0.11)	0.73 (0.11)	0.83 (0.12)	0.78 (0.11)	0.74 (0.11)
Road	M	0.47 (0.15)	0.46 (0.13)	0.47 (0.12)	0.62 (0.16)	0.63 (0.15)	0.65 (0.13)
Road	C	0.85 (0.16)	0.83 (0.16)	0.81 (0.15)	0.86 (0.15)	0.84 (0.15)	0.82 (0.15)
Land	M	0.46 (0.19)	0.47 (0.16)	0.47 (0.13)	0.70 (0.17)	0.72 (0.16)	0.75 (0.15)
Land	C	0.86 (0.13)	0.83 (0.13)	0.79 (0.14)	0.90 (0.11)	0.89 (0.11)	0.88 (0.11)

Table 1. Mean (standard deviations are in parentheses) probabilities of mislabeled (M) and clean (C) examples on three data sets with synthetic class noise. RA denotes random noise, and RU denotes rule-based noise. We artificially mislabeled 10, 20, and 30% of each data set.

Our implementation of PWEM in which we estimate mean and covariance matrices via EM assumes that the training data is well conditioned prior to using our methods. If not, the user will experience computational issues associated with matrix singularity. In the case of the sulfur training data, feature selection was applied to the original data and resulted in twelve bands that were well conditioned. Thus, we were able to use an implementation of EM that estimates a full covariance matrix. For data that is not well conditioned, a possible workaround is to use a version of EM that only assumes a diagonal covariance matrix for the mixture models. We also implemented this version of PWEM, using an open source machine learning software package called WEKA [9]. We found that the assumption of a diagonal covariance matrix does not significantly impact the results of PWEM.

3. DATA CLEANING OPTIONS

PWEM is successful when it assigns low probabilities to mislabeled examples and high probabilities on clean examples. Table 1 shows probability results from two scene segmentation data sets (segmentation, road), and the landcover data set. We introduce up to 30% random and ruled-based noise on each. Table 1 demonstrates that PWEM assigns lower probabilities to mislabeled instances versus clean instances on average. Having calculated probabilities for each example in the training set, we present three options for using PWEM’s probabilities to clean the training set. The first is to choose a threshold and discard examples whose probabilities are below this value. The advantage of this method is that it is the simplest of the three options to implement. The drawbacks are the selection of the threshold value and the potential loss of valuable training examples.

The second option is to use PWEM’s probabilities as instance weights during training. Rebbapragada et al. compared weighting to discarding and found that instance weighting outperformed discarding on the data sets tested [5]. However, the instance weighting option requires a classifier that can incorporate instance weights during training.

The third option is to use the PWEM probabilities to engage the domain expert in the process of iteratively cleaning the training data. In this option, the domain expert is presented with the m lowest probability examples from PWEM. The expert examines the labels on this set and relabels anything that is incorrectly labeled. The training set is updated with the new labels. The process repeats with PWEM only presenting examples the domain expert has not already seen. The process terminates when the domain expert is satisfied with the cleanliness of the sets. This option is clearly the best in terms of ensuring the quality of the training data. The drawback is that it requires a significant time commitment from the domain expert.

For the sulfur detection problem, we chose to discard low probability examples. We could not take advantage of instance weighting because our classifier implementation could not accommodate instance weights. In future work, we plan to present low probability examples to our domain expert for iterative relabeling.

4. SULFUR DETECTION ON EO-1

In this section, we describe the sulfur detection problem and how we used PWEM to identify mislabeled examples in the training set of ice, rock and sulfur examples. The detection of sulfur in remote-sensing imaging supports NASA’s mission to discover evidence of living systems within our solar system. The importance of sulfur detection to both science and NASA is described thoroughly in [1]. We summarize the key points in this paper, but refer the reader to [1] for more detailed information.

Sulfurous glacial springs are of scientific interest because the presence of elemental sulfur (among other compounds) may be an indicator of microbial activity. The Borup-Fiord Pass on Ellsmere Island in northern Canada is known to have sulfurous glacial springs. Indeed, tests of these springs have yielded evidence of yet unknown strains of bacteria [10].

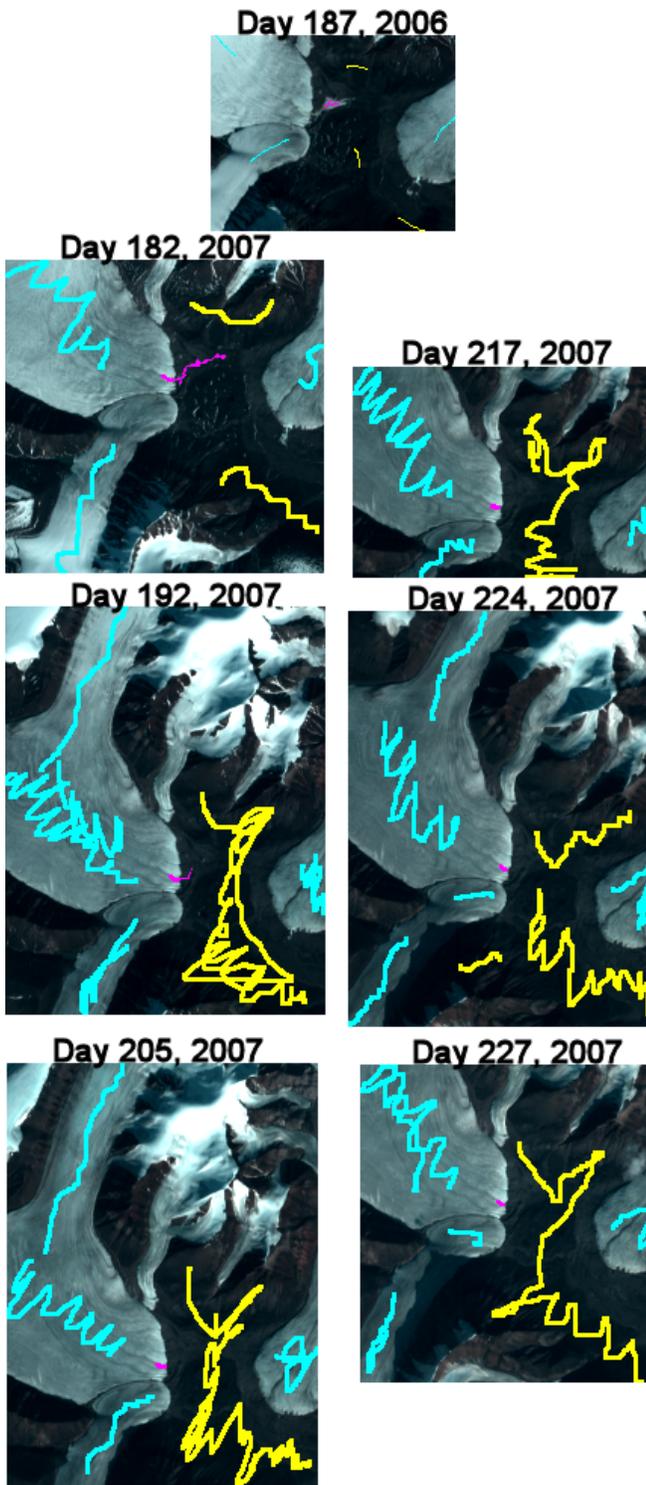


Figure 2. Expert-provided labels for source images of Borup Fiord. Springs are at the edge of the glacier. (yellow = rock, cyan = ice, magenta = sulfur)

Jupiter's moon Europa is also thought to have an analogous landscape of ice, rock, and sulfur [4]. Images of Europa's icy landscape show reddish-tinged upwellings that may be an indicator of sulfurous compounds on solid ice [3]. Thus,



Figure 3. Example of a 4-class training image. The bottom frame shows labeled pixels of ice (cyan), rock (yellow), bright sulfur (magenta) and dark sulfur (green).

the study of sulfurous ice in the Borup Fiord pass may inform the study of Europa's biological potential. Given the inaccessibility of the Borup Fiord Pass and Europa, remotely-sensed images are the only source of regular data for those regions.

EO-1, managed by NASA's Goddard Space Flight Center, is an Autonomous ScienceCraft that has onboard decision-making abilities. This means that EO-1 may autonomously modify its observation plan and schedule repeated observations of a scientific event or point of interest. Onboard processing can speed such modifications by the order of hours. A sulfur detector onboard EO-1 can trigger follow-up imaging of the Borup-Fiord pass on the satellite's next pass over the region. It is crucial that imaging is not triggered on areas that do not contain sulfur. Thus, minimizing the classifier's false positive rate is an important goal.

Mandrake et al. [1] discuss the challenges inherent in detecting sulfur from Hyperion images of the Borup Fiord pass. The first set of challenges is related to the computational constraints of onboard processing. For example, only twelve out of 220 bands may be analyzed at once. One must employ feature selection to search for the best feature subset to use during training and classification. One must also select a classifier that yields good performance under integer arithmetic and tune its associated parameters.

The second set of challenges concerns the quality of the training data. Figure 2 shows the source of the training data. These are seven flyover images of the Borup Fiord pass that were labeled by a geologist on our team. Detailed information about the training data and its labeling process can be found in [1]. The amount of training data available to Mandrake et al. was significantly larger compared to previous work [2]. However, the quality of the training data was questioned when preliminary results did not yield an adequate false positive rate.

Mandrake et al. [1] identified two problems with the training data. The first problem was that a single sulfur

class did not adequately represent the sulfur examples. A clustering of the training data revealed that the sulfur examples formed a bi-modal distribution. Thus, the decision was made to partition the sulfur examples into two classes: bright sulfur and dark sulfur. Thus, the training set was now relabeled to have four classes. However, the re-labeling of the training data did not result in a major reduction to the false positive rate of the classifier. At that time, the researchers from Mandrake et al. suspected that the problem was due to mislabelings in the training data. We used the PWEM method to determine whether the training data was mislabeled and present our analysis in Section 6.

5. DATA

Figure 2 shows the source images that were used to generate labels for the training data. The original labels were rock, ice, and sulfur, depicted by the yellow, cyan, and magenta streaks. Detailed information on how the images were labeled is provided by [1]. The training set contained 12029 pixels labeled as rock, 9485 labeled as ice, and 235 labeled as sulfur. As described in [1], the sulfur class was eventually partitioned into two sub-classes: bright sulfur and dark sulfur. There were 144 bright sulfur pixels and 91 dark sulfur pixels. Tables 2 and 3 summarize the breakdown of examples by class in both training sets. In both training sets, the sulfur examples are in the extreme minority.

ROCK	12029
ICE	9485
SULFUR	235

Table 2. Class breakdown of the 3-class training set.

ROCK	12029
ICE	9485
BRIGHT SULFUR	144
DARK SULFUR	91

Table 3. Class breakdown of the 4-class training set.

Figure 3 shows examples of light and dark sulfur in one of the training images. The top and bottom panels are the same image. The bottom panel shows the location of the labeled pixels of ice (cyan), rock (yellow), bright sulfur (magenta) and dark sulfur (green). Using the bottom panel, one can locate bright and dark sulfur pixels in the (unlabeled) top panel to see how those classes manifest in the images.

6. RESULTS

We performed one run of PWEM on both the 3- and 4-class training data sets, and analyzed both sets of

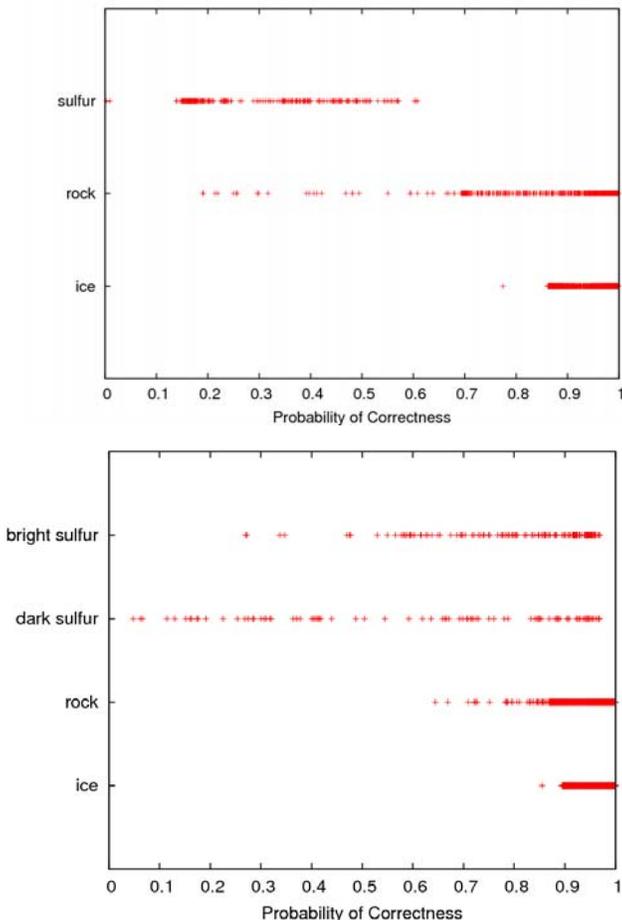


Figure 4. Distribution of PWEM probabilities for the 3-class (top) and 4-class (bottom) training sets.



Figure 5. The left panel shows labeled bright sulfur (light green) and dark sulfur (dark green) pixels. The right panel shows the probability ranges on those sulfur examples. Green pixels have probabilities greater than 0.9, yellow pixels have probabilities greater than 0.75, and red pixels have probabilities less than 0.5.

probabilities. Figure 4 shows the distribution of probabilities in each training set. The top and bottom panels show that in both training sets, the majority of rock and ice examples are assigned probabilities above 0.7. This is not surprising since these classes dominate the training data set and therefore have the most influence on the cluster models. If the rock and ice examples did not separate well in the rock/ice clustering of PWEM, then the probabilities would have been lower. This is good evidence that the rock and ice examples were cleanly labeled.

The top panel of Figure 4 shows that the 3-class training set has very few sulfur examples with a probability of correctness greater than 0.6. Given the skewed nature of the 3-class training set, it is expected that the sulfur-rock and sulfur-ice clusterings would result in low probabilities for sulfur examples given the overwhelming amount of rock and ice present. Using PWEM under extremely skewed class distributions could result in low probabilities that are indicative of the imbalance in the class distributions, not the correctness of the training set.

The 4-class training set rectifies this problem because probabilities on sulfur examples are evaluated against three clusterings each, rather than two. A bright sulfur example is evaluated against the bright-sulfur/ice, bright-sulfur/rock and bright-sulfur/dark-sulfur clusterings. With an extra (approximately) class-balanced clustering to evaluate against, it is possible for the sulfur examples to achieve a higher probability. Indeed, the bottom panel of Figure 4 shows a greater spread in the sulfur probabilities, with many examples achieving a probability above 0.75. With a larger spread in the probabilities, it is easier to determine which examples are likely mislabeled, and which are likely clean.

The bottom panel of Figure 4 shows that PWEM calculates higher probabilities for the bright sulfur examples pixels compared to the dark sulfur. Indeed, the mean probability for bright sulfur is 0.81. For dark sulfur, it is 0.62. PWEM indicates that the dark sulfur class is more severely mislabeled. The left panel of Figure 5 shows the original bright and dark sulfur labels in light green and dark green. Note that a few dark sulfur pixels are mislabeled as bright sulfur. The right panel shows the PWEM probabilities for those pixels with probabilities over 0.9 in green, probabilities over 0.75 in yellow, and probabilities below 0.5 in red. Note that the area containing light sulfur pixels is mostly yellow. The area labeled as predominantly dark sulfur, has splotches of red because they were assigned low probabilities. Our hypothesis is that the dark sulfur and rock examples are easily confusable. Because correctly labeling and classifying dark sulfur pixels are the harder problem, we focus our evaluation efforts on classifying bright sulfur.

Due to our use of the SVM learning algorithm, we chose to clean our training set via filtering. We selected an arbitrary confidence threshold of 0.75 and discarded any bright or dark sulfur examples below this threshold. We retained 104 out of 144 bright sulfur examples, and 39 of 91 dark sulfur examples. This was approximately 72% of the bright sulfur examples, and 43% of the dark sulfur examples. As discussed in [1] and above, the loss of more dark sulfur examples is understandable given that dark sulfur is difficult to distinguish from rock. In contrast, bright-sulfur is relatively easier to distinguish from dark sulfur, rock and ice.

Having discarded examples from the training set, we trained a Support Vector Machine (SVM) classifier to test the effects of the filtered training set on the false positive rate of the classifier. We report results from an SVM classifier using a linear kernel. We performed multiple runs of the classifier using different values for the regularization parameter C . For each value of C , we performed seven-fold cross validation. Specifically, we trained the classifier on six of training images, and test on the seventh, rotating each image as the test image. We calculated an F-measure score for each run, where F-measure was calculated as $F=2PR/(P+R)$, and P and R are precision and recall respectively. We modified the definitions of P and R to penalize less for mistakes on dark sulfur. If the SVM identified an ice or rock example as dark sulfur, it did not count as a false positive. Similarly, if the SVM failed to positively identify a dark sulfur example, the mistake was not counted as a false negative. We only penalized for precision and recall mistakes on bright sulfur examples.

A second evaluation metric was the mean number of false positives per image. This is simply the number of false positives averaged across each test image. For this metric, we used two additional sets derived from different evaluation images. The first, referred to as “Likely”, consisted of all unlabeled pixels from the original training images (see Figure 2). Because these images were unlabeled, they were most likely from regions that did not contain sulfur. Therefore, positive classifications in these regions are likely to be false positives. Of course, without explicit labels we cannot be absolutely certain that no sulfur exists in these areas. The second set of evaluation images, called “Sulfur Free”, consists of other Hyperion images (not of the Borup Fiord Pass) in which we are certain there are no known sulfur sources. Positive classifications on these regions are definitely mistakes.

We present the F-measure results in Table 4, and the mean false positive results in Table 5. EXPERT denotes results on the unfiltered training data using features (image bands) that were selected by the domain expert. PWEM FILT is the same training set minus the any example whose probability of correctness was less than 0.75. We report the median and best result of all runs.

	MEDIAN	BEST
EXPERT	0.92	0.98
PWEM-FILT	0.91	0.98

Table 4. Median and best F-measure results on the original (EXPERT) and filtered (PWEM-FILT) training sets.

	EXPERT		PWEM FILT	
	MEDIAN	BEST	MEDIAN	BEST
Likely	981.59	422.49	697.16	459.73
Sulfur-Free	160.12	10.51	27.45	0.78

Table 5. Median and best mean false positives per image results on both the “Likely” and “Sulfur-Free” evaluation sets.

According to Table 4, the filtered training set provides no improvement over the original training set in terms of F-measure. PWEM’s contribution is in the reduction of false positives. On the Likely evaluation set, PWEM reduced the median false positives per image by approximately 29%. On the easier Sulfur Free evaluation set, PWEM reduced the median false positives per image by approximately 83% and also recorded a best false positive rate of one false positive per image.

7. CONCLUSION

This paper presents the PWEM method for detecting mislabeled training data. We describe PWEM’s contributions to NASA’s sulfur detector that is scheduled for use onboard the EO-1 spacecraft. By filtering the labeled training set based on PWEM’s estimated label probabilities, the false positive rate of the classifier is reduced, ensuring that fewer resources would be wasted if EO-1 is programmed to do follow-up imaging upon positive sulfur classifications.

Coping with erroneous labeled data is critical for many scientific domains. This paper provided an in-depth study of how PWEM improved classifier performance in one particular domain through the filtering of low probability examples. No doubt, PWEM can be useful in mitigating label noise in other domains. Furthermore, using the probabilities to filter training sets is only one of three options available. We also discussed the use of the probabilities as instance weights in learning algorithms that can take instance weights as inputs. We also discuss a third, and possibly better option, to iteratively query the domain expert on labels that have a low probability of correctness. Because the probabilities calculated by PWEM are robust, we expect classifier performance to improve under any of these three options.

ACKNOWLEDGMENTS

We gratefully acknowledge support, assistance, and data from the Earth Observing-1 mission. Expert data labeling was done by Damhnait Gleeson with the support of NASA Planetary Geology and Geophysics Program grant #NNX07AR28. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2009. All rights reserved.

REFERENCES

- [1] L. Mandrake et al, "Onboard Detection of Natural Sulfur on a Glacier via an SVM and Hyperion", In Proc. of the 2009 IEEE Aerospace Conference, March 7-14, 2009.
- [2] R. Castaño, "Onboard Detection of Active Canadian Sulfur Springs: A Europa Analogue," International Symposium on Artificial Intelligence Robotics and Automation in Space (i-SAIRAS), Los Angeles, Feb. 25-29th, 2008.
- [3] C. Sotin et al, "Europa: Tidal Heating of Upwelling Thermal Plumes and the Origin of Lenticulae and Chaos Melting," Geophysical Research Letters, vol. 29, No. 8, doi: 10.1029/2001GL013844, 2002.
- [4] R. W. Carlson et al, "Sulfuric Acid on Europa and the Radiolytic Sulfur Cycle", Science, Vol. 1, No. 5437, pp. 97-99, doi: 10.1126/science.286.5437.97, Oct 1999.
- [5] U. Rebbapragada, "Class Noise Mitigation through Instance Weighting", In Proc. of the European Conference on Machine Learning, pp. 708-715, 2007.
- [6] D. Pelleg and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters", In Proc. of the 17th. International Conference on Machine Learning, pp. 727-734, 2000.
- [7] P. Smyth, "Bounds on the mean classification rate of multiple experts." Pattern Recognition Letters, Vol. 17, pp. 1253-1327, 1995.
- [8] A. Dempster, N. Laird and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 1-38, 1977.
- [9] I. H. Witten, and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, 2005.
- [10] S. E. Grasby et al, "Supraglacial Sulfur Springs and Associated Biological Activity in the Canadian High Arctic – Signs of Life Beneath the Ice," Astrobiology, 3(3): 583-596, doi: 10.1016/S0375-6742(03)00026-8, Sep 1st 2003.

BIOGRAPHY



Umaa Rebbapragada is a Ph.D. candidate in the Department of Computer Science at Tufts University. She graduated from the University of California, Berkeley with a B. A. in Mathematics, and has six years of industry experience as a web developer and software engineer at the Internet media company CNET Networks. At Tufts,

Umaa's field of research is machine learning, with specific focus on supervised learning, anomaly detection, and time series data mining methods. Umaa's advisor is Professor Carla Brodley. She also collaborates with astrophysicists at the Harvard-Smithsonian Center for Astrophysics, and has completed a summer internship at the Jet Propulsion Laboratory.



Dr. Lukas Mandrake is a member of the technical staff in the Machine Learning and Instrument Autonomy group at the Jet Propulsion Laboratory in Pasadena, CA. He is involved in the application of machine learning techniques to Earth-sensing satellite missions, autonomous spectroscopy, and autonomous sensor networks and is

a key member of the VCAM development / data analysis team. Lukas particularly enjoys studying computational models of natural systems. Lukas received his Ph.D. and M.S. in computational plasma physics from UCLA in 2004 and his B.A. in engineering physics from the University of Arizona in 1995.



Dr. Kiri Wagstaff is a senior researcher at the Jet Propulsion Laboratory in Pasadena, CA. She is a member of the Machine Learning and Instrument Autonomy group, and her focus is on developing new machine learning methods that can be used for data analysis onboard spacecraft. She has applied these techniques to data

being collected by the EO-1 Earth-orbiting spacecraft, Mars Odyssey, and Mars Pathfinder. She has also worked on crop yield prediction from orbital remote sensing observations, the fault protection system for the MESSENGER mission to Mercury, and automatic code generation for the Electra radios used by the Mars Reconnaissance Orbiter and the Mars Science Laboratory. She holds a Ph.D. in Computer Science from Cornell

University and an M.S. in Geological Sciences from the University of Southern California.



Damhnait Gleeson is a Ph.D. candidate in the Department of Geological Sciences at the University of Colorado at Boulder, currently carrying out research at the Jet Propulsion Laboratory. Her research explores the potential for life in icy, sulfur-rich environments such as may exist at Jupiter's moon Europa, and investigates the

detectability of such life by the study of analogous terrestrial field sites. She is co-advised by Alexis Templeton at the University of Colorado and Bob Pappalardo at the Jet Propulsion Lab.



Dr. Rebecca Castaño is an Assistant Manager of the Instruments Software and Science Data Systems section at JPL. She received her Ph.D. in Electrical Engineering from the University of Illinois with her dissertation in the area of computer vision. Dr. Castaño has been advancing the state of the art in onboard science

methods for the past five years and has been lead author on numerous publications in this field. From 1999 – 2001 Dr. Castaño served as the application lead for phenomenological computational field geology efforts in the MLS Group. She is the technology lead for science data processing for the Autonomous Sciencecraft Experiment on the New Millennium Program's ST6 project. Dr. Castaño is also the Team Lead of the Onboard Autonomous Science Investigation System (OASIS) project and the subtask lead for Data Analysis on the automated Multirover Integrated Science Understanding System (MISUS). Her research interests include machine learning, computer vision, and pattern recognition.



Dr. Steve Chien is a Principal Computer Scientist in the Mission Planning and Execution Section at the Jet Propulsion Laboratory, California Institute of Technology and Adjunct Associate Professor with the Department of Computer Science of the University of Southern California. He holds a B.S. with Highest Honors in

Computer Science, with minors in Mathematics and Economics, M.S., and Ph.D. degrees in Computer Science, all from the University of Illinois. Dr. Chien is a recipient of the JPL Lew Allen Award for Excellence and two NASA Exceptional Achievement Medals for his work in research and development of planning and scheduling systems for

NASA. He is the Principal Investigator for the Autonomous Sciencecraft Experiment, a co-winner of the 2005 NASA Software of the Year Award.



Dr. Carla E. Brodley is a professor in the Department of Computer Science at Tufts University. She received her Ph.D. in computer science from the University of Massachusetts, at Amherst in 1994. From 1994-2004, she was on the faculty of the School of Electrical Engineering at Purdue University, West Lafayette, Indiana. Professor

Brodley's research interests include machine learning, knowledge discovery in databases and computer security. In 2004-2005 she was a member of the Defense Science Study Group. In 2001 she served as program co-chair for the International Conference on Machine Learning (ICML) and in 2004, she served as the general chair for ICML. Currently she is on the editorial board of the Journal for Machine Learning Research and an associate editor on the Machine Learning Journal. She is a member of the Computing Research Association's Committee on the Status of Women in Computing Research (CRA-W).