ACTIVE LEARNING OF HYPERSPECTRAL DATA WITH SPATIALLY DEPENDENT LABEL ACQUISITION COSTS

Alexander Liu *Goo Jun* *Joydeep Ghosh*

Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin TX 78712, USA
{aliu, gjun, ghosh}@ece.utexas.edu

1. INTRODUCTION

Supervised machine learning techniques can be used to classify various forms of GIS data. However, supervised techniques rely on large amounts of labeled data in order to build accurate models. While the data itself may be comparatively easy to gather, labeling this data is often difficult and costly. One machine learning technique to reduce the amount of labeled data required to build a supervised model is active learning. In active learning, one attempts to reduce the number of labeled training points required for a certain level of classifier performance by allowing the active learning algorithm itself to choose which points should be labeled.

Most active learners assume that: 1) the cost of acquiring the label for a particular point is independent of the costs for all other points and that 2) label acquisition costs are equal. When labeling spatially distributed data, both of these assumptions may be false. For example, for classification of land-cover using hyperspectral data, acquiring labels may involve traveling to a particular location and performing some sort of test such as determining land type or collecting various samples (e.g., soil, water, foliage) that requires physical access. Traveling to this point incurs some type of cost (e.g., gas or time) proportional to distance traveled. The distance traveled also depends on the order in which one labels the points.

In this abstract, we present a novel framework for performing active learning while taking into account spatially sensitive labeling costs. In particular, we pose the problem as a traveling salesman problem with profits. We present example results using hyperspectral data, but the presented approach is also applicable to other GIS problems where supervised learning is used.

Related works: Although many active learning strategies have been proposed during the last 15 years, there exist few algorithms that consider spatial characteristics of unlabeled samples. [1] proposed an active learning algorithm for hyperspectral data that adapts a classifier for spatial variation of spectral signatures. However, it does not take into account any form of varying label acquisition costs based on spatial data. An active learning algorithm to efficiently model spatial phenomena with Gaussian processes has been proposed [2], but the algorithm is used to model spatially varying quantities and is not applicable to classification problems. We are unaware of any active learning studies which take spatially dependent label acquisition costs into account.

2. ACTIVE LEARNING WITH SPATIAL COSTS

Problem setting: In this abstract, we adapt a pool-based active learning technique called uncertainty sampling [3] for handling spatially related label acquisition costs. As in "standard" active learning, active learning on spatial data occurs in an iterative fashion where, on each iteration i, points from some unlabeled set $\mathcal{U}$ are selected by the active learner based on some criteria, labeled by some oracle, and then placed in the labeled set $\mathcal{L}$. The labeler starts and ends each iteration at some "home location" [1]. On each iteration, the labeler labels points in $\mathcal{U}$ until some traveling budget is expended, where the traveling budget is the amount of time available for traveling and labeling per iteration.

We will use the following notation. On the ith iteration, the algorithm selects n_i points for labeling where n_i depends on some traveling budget which we will denote as t_{max}. The actual cost of traveling and labeling points for the ith iteration will be denoted as t_i. t_i depends on the total distance d_i traveled on the ith iteration, the speed s of the labeler's vehicle, the cost of labeling a single point c_l, and the number of points labeled n_i. In particular, $t_i = (d_i/s) + (c_l * n_i)$ and the constraint is that $t_i < t_{max}$. We will measure t_i, t_{max}, and c_l in units of time, d_i in units of length, and s in units of length/time. Finally, we will denote the uncertainty score (as determined by uncertainty sampling) for the jth point in $\mathcal{U}$ as $u(j)$.

Solutions: A simple baseline is to start at home and continue labeling the next closest unlabeled point while $t_i < t_{max}$. We will call this baseline the "closest next" baseline. A second baseline is to pick points via the non-spatial, "traditional" machine learning methods of random sampling and uncertainty sampling. Then, using a solution to the traveling salesman problem, the shortest path through the chosen points is followed. We will refer to this baseline algorithm as "random/TSP" if random sampling is used or "US/TSP" if uncertainty sampling is used to select points.

However, the above techniques are somewhat naive, as they look at either only spatial locations or only the benefit to the classifier. A more sophisticated approach is to pose the problem as a traveling salesman problem with profits (TSPP) [4], allowing for both spatial

[1] This home location may correspond to where the labeler's vehicle is stored/refueled or the labeler's base of operations.

information and benefit to the classifier to be examined simultaneously. Our first proposed method is to pose the problem as a TSPP problem where the profit for visiting the jth point is the uncertainty score $u(j)$ of that point, and the constraint is that the salesman can visit a variable number of cities per iteration as long as the total time required to travel along all cities and reach home is less than the traveling budget t_{max} for that iteration. We refer to this approach as "US/TSPP". Finally, we found a variant of US/TSPP to be empirically useful: instead of supplying all possible unlabeled points in $\mathcal{U}$ to the TSPP algorithm, only the top m points with the highest uncertainty scores (where $m \geq n_i$) are used. We refer to this approach as "US/TSPP (filtered)", and set $m = 100$ in experiments.

3. EXPERIMENTS

In this section, we present example results on the Kennedy Space Center (KSC) and Botswana hyperspectral datasets. The data is pre-processed using both the max-cut algorithm [5] and best-basis feature reduction [6], both of which are useful for classifying hyperspectral data. We use an LDA classifier and average results over five runs of ten-fold cross validation. Experimentally, we tried a variety of values for s and c_l [2], but interestingly, specific values do not seem to affect general trends very much. Here, we present results where $s = 80$ kilometers per hour, $c_l = 10$ minutes, and $t_{max} = 8$ hours.

Example results are plotted in Figure 1. The results should be interpreted by looking at three aspects of each curve: the total amount of effort required to label all of U [3], how quickly the method reduces error rate, and the lowest error rate a method achieves. For both datasets, the initial reduction in error is similar for all but the random/TSP methods, but both US/TSPP and the "closest next" baseline tend to outperform the other techniques. Not surprisingly, the "closest next" baseline requires the least effort to label all points in $\mathcal{U}$, but US/TSPP is very competitive, and is followed by US/TSPP(filtered), US/TSP, and then random/TSP. However, in terms of the minimum error rate achieved, the "closest next" baseline does very poorly, and the US/TSPP method is preferable. In addition, US/TSPP(filtered) and US/TSP tend to achieve the lowest error rates. Thus, US/TSPP(filtered) appears to be the best tradeoff in terms of reducing error rate quickly and achieving a low error rate.

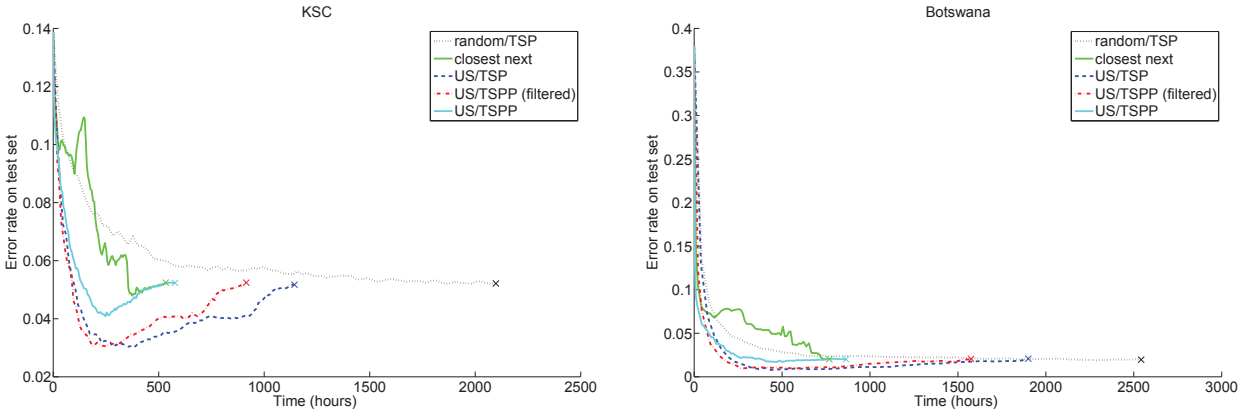

Figure 1: Example results.

4. REFERENCES

[1] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, 2008.

[2] Andreas Krause and Carlos Guestrin, "Nonmyopic active learning of gaussian processes: an exploration-exploitation approach," in *ICML '07: Proceedings of the 24th international conference on Machine learning*, New York, NY, USA, 2007, pp. 449–456, ACM.

[3] David D. Lewis and Jason Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *In Proceedings of the Eleventh International Conference on Machine Learning*. 1994, pp. 148–156, Morgan Kaufmann.

[4] Dominique Feillet, Pierre Dejax, and Michel Gendreau, "Traveling salesman problems with profits.," *Transportation Science*, vol. 39, pp. 188–205, 2005.

[5] Yangchi Chen, M.M. Crawford, and J. Ghosh, "Knowledge based stacking of hyperspectral data for land cover classification," *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, pp. 316–322, 2007.

[6] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. on Geosci. and Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, 2001.

[2] experimental values for s ranged from 15 to 80 kilometers per hour, while values for c_l ranged from 10 to 50 minutes

[3] these points are plotted with an X in the graphs